In-flight photometry extraction of PLATO targets : optimal apertures for detecting extra-solar planets

Victor Marchiori

PSL UNIVERSITÉ PARIS

USP

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Observatoire de Paris
Dans le cadre d'une cotutelle avec l'Université de São Paulo (USP)

Extraction photométrique en vol des étoiles de la mission PLATO: Masques optimaux pour la détection de planètes extra-solaires

In-flight photometry extraction of PLATO targets: Optimal apertures for detecting extra-solar planets

Soutenue par
Victor Atilio MARCHIORI
Le 16 September 2019

École doctorale n°127
Astronomie et Astrophysique d'Île-de-France (AAIF)

Spécialité
Astronomie et Astrophysique

Composition du jury :

Marie-Christine ANGONIN Professeur, Observatoire de Paris	*Présidente*
Magali DELEUIL Professeur, Université Aix Marseille	*Rapporteure*
Suzanne AIGRAIN Professeur, University of Oxford	*Rapporteure*
Eduardo JANOT PACHECO Professeur, Universidade de São Paulo	*Examinateur*
Frédéric BAUDIN Astronome, Université Paris-Sud	*Examinateur*
Alexandre SANTERNE Astronome adjoint, Univ. Aix Marseille	*Examinateur*
Réza SAMADI Astronome, Observatoire de Paris	*Directeur de thèse*
Fábio de Oliveira FIALHO Professeur, Universidade de São Paulo	*Directeur de thèse*

l'Observatoire de Paris | PSL

Victor Atilio Marchiori

In-flight photometry extraction of PLATO targets: Optimal apertures for detecting extra-solar planets

Thesis presented to the Polytechnic School of the University of São Paulo for obtaining the degree of Doctor of Engineering. Thesis presented to the Paris Observatory for obtaining the degree of PhD in Astronomy and Astrophysics.

Universidade de São Paulo / Observatoire de Paris

Departamento de Telecomunicações e Controle da Escola Politécnica (PTC/EPUSP) / Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique (LESIA)

Engenharia de Sistemas / Astronomie et Astrophysique

Supervisor: Fábio de Oliveira Fialho (EPUSP)

Supervisor: Réza Samadi (LESIA)

Meudon

2019

To my beloved and brave wife, Gunila, and our son, Lucas.

To my mom, Margarete, and my dad, Atilio.

To my grandma, Maria.

To my friends.

In honour of the unwavering Laws that permeates the whole of Creation.

Acknowledgements

I am deeply grateful to Marie-Christine, Suzanne, Magali, Eduardo, Frédéric, and Alexandre for having accepted being members of my thesis jury and for having dedicated their valuable time reading this manuscript. I give special thanks to the reviewers, Suzanne and Magali, for having also devoted their time to write the thesis report.

I would like to thank the French National Centre for Space Studies (CNES), the Paris Observatory – PSL and the Brazilian National Council for Scientific and Technological Development (CNPq) for having provided me financial support to develop this thesis.

I acknowledge the directorate and the members of the Doctoral School of Astronomy and Astrophysics of Ile-de-France (ED127) for their readiness to support PhD students whenever needed. I express my gratitude to the administration of LESIA and to the members of *Pôle Étoile* for their kindness and for having provided me excellent working conditions throughout my whole PhD. In especial, I thank Benoît, Bram, Caroline, Charlotte, Charly, Coralie, Daniel, Eric, Jordan, Marie-Jo, Morgan, Kévin, Rhita, Richard, Steven and Yveline for the great moments we shared together in the everyday working life *au Labo*.

I would like to express the deepest appreciation to my advisors Réza and Fábio. They were remarkably committed and responsive throughout the whole duration of this thesis, and their advices and support were key for me to obtain the results presented in this manuscript.

Finally, I am deeply grateful to my beloved wife, Gunila, for her invaluable company, support and encouragement to keep moving forward all the time. She and our one year old little Lucas were certainly one of my greatest motivators to take this thesis until its end.

"We recognize and discern the artist in his work,
and he becomes familiar to us without
our needing to see him personally."
(In the Light of Truth, Grail Message by Abdrushin)

List of Figures

List of Tables

Contents

1 Introduction

The search for life beyond Earth is certainly one of the most captivating and inspiring questions in modern science. When contemplating the boundlessness of our universe, with its hundred billions galaxies composed of countless stars and planets, we might fell impelled to interrogate ourselves with a simple yet meaningful question: "Are we alone?". It turns out that scientifically answering to this question has been proven to be one of the toughest challenges in the cohesive scientific fields of astronomy, astrophysics and astrobiology (see Caporael (2018), Galante *et al.* (2016)). Indeed, even with the most powerful telescopes available to date, which are sensitive to electromagnetic radiation in a wide range of wavelengths and thereby capable to probe the universe at mind-blowing distances far away from our Solar System, no clear sign or evidence of existing life beyond Earth was ever found yet.

1.1 Searching for potentially habitable planetary systems

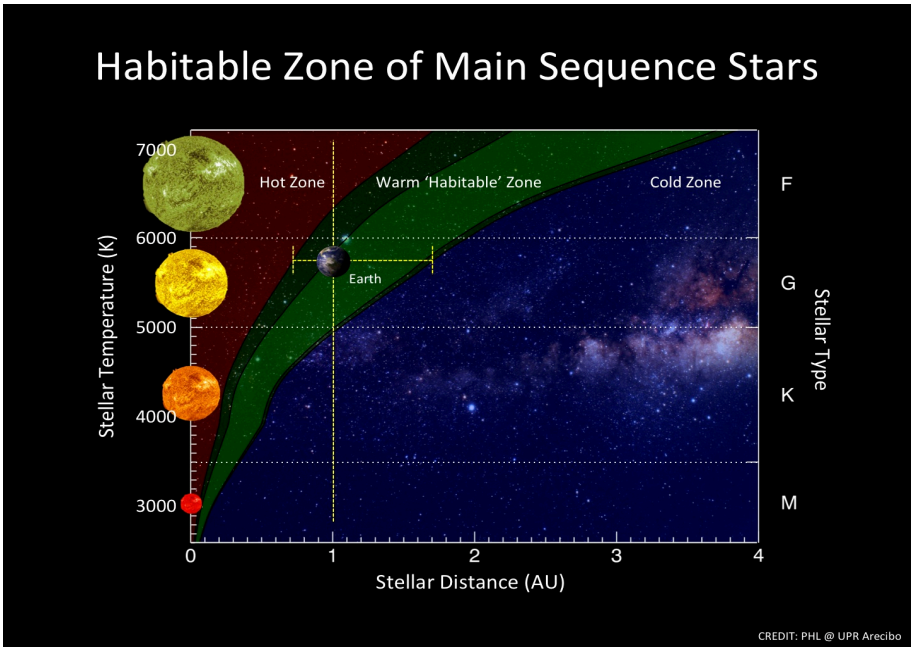
From a scientific point of view, unveiling the secrets of the origin of life requires carefully examining the environment conditions and the physicochemical processes that favoured and sustained life. A valuable step towards that comprehension resides in searching for potentially habitable planetary systems beyond our Solar System, so that we can properly understand how these systems typically form and evolve. In this context, the concept of habitable zone gives, based on the only known life-sustaining sample (the Earth), a clue on where life as we know is (more likely) expected to be encountered.

Habitable zone (HZ), also called Goldilocks zones, may be understood as the zone around a star where the temperature is just right (i.e. not too hot and not too cold), so that liquid water can exist on the surface of a planet orbiting that star. Accordingly, the hotter a star is, the farther away and the larger will be its HZ, and vice-versa. In the literature, habitable zones are estimated based on planetary climate models such as those presented in Kasting, Whitmire & Reynolds (1993), Kopparapu *et al.* (2013), Shields, Ballard & Johnson (2016), Bin, Tian & Liu (2018). As represented in Figure 1, HZ sizes may be more optimistic (wider) or more conservative (narrower), depending on the considered model. According to the conservative definition (Kopparapu *et al.* (2013)), the Earth is located at the inner boundary of Sun's habitable zone.

In a judicious way, habitable zones shall not be however strictly thought in terms of temperature only. There are other fundamental physical aspects that significantly contribute to make a planet capable to harbour life. An atmosphere (air), for example, is essential for the existence of life in a planet, because without it, there is not sufficient

pressure to keep water in liquid state, even if the planet has the right temperature. To some extent, an atmosphere also act as a protective skin against spatial objects on collision courses with a planet, by incinerating them before they can hit the planet surface. Persistent atmospheres require, in turn, the existence of sufficiently strong magnetic shields, otherwise the radiation winds emitted by a star may easily vaporize the atmosphere of a planet orbiting within its habitable zone. Hence, the conditions for the emergence and the maintenance of life are in practice much more complex than the formal definition of habitable zone might suggest. Yet, this concept still represents today our best guess on where to search for potentially habitable planetary systems.

Figure 1 – Habitable zone around main sequence stars of spectral type F, G, K, and M, as a function of the stellar temperature and the planetary orbit distance to its host star. Boundaries of HZ may be divided into optimistic (light green) and conservative (dark green) versions.

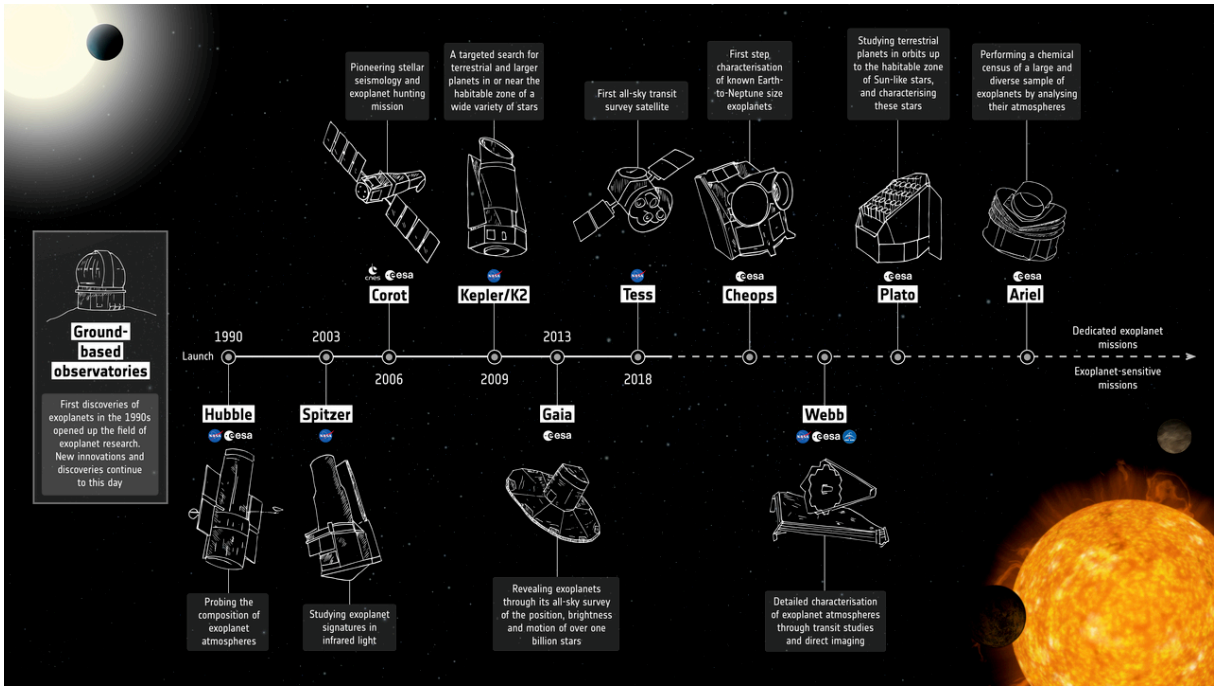


Source: Planetary Habitability Laboratory (PHL), University of Puerto Rico, Arecibo.

The first confirmed exoplanet discoveries dates from the 90's (Wolszczan & Frail (1992), Mayor & Queloz (1995)). Yet, spatial missions dedicated to widely search for new exoplanets only started operating by the second half of the 2000's (see Figure 2); first, the pioneer mission Convection, Rotation and planetary Transits CoRoT (Baglin *et al.* (2006), Auvergne *et al.* (2009), Deleuil & Fridlund (2018)) from the French Space Agency (CNES), in 2006; then the breaking though mission *Kepler* (Borucki *et al.* (2010)) from NASA, in 2009. At the present date, both CoRoT and *Kepler* spacecrafts are no longer operational. In contrast, the NASA exoplanet hunter spacecraft TESS (Ricker *et al.* (2014)), launched in April of 2018, is fully operational and has already detected quite a few exoplanets.

Additionally, the ESA spacecraft CHEOPS (Cessa *et al.* (2017)), expected to be launched by the end of 2019, will observe bright targets which are already know to host super-Earth to Neptune mass planets. The aim of this mission is to accurately determine the radii of these planets, for which the respective masses have already been obtained by ground-based spectroscopic surveys.

Figure 2 – Timeline of exoplanet missions.



Source: ESA.

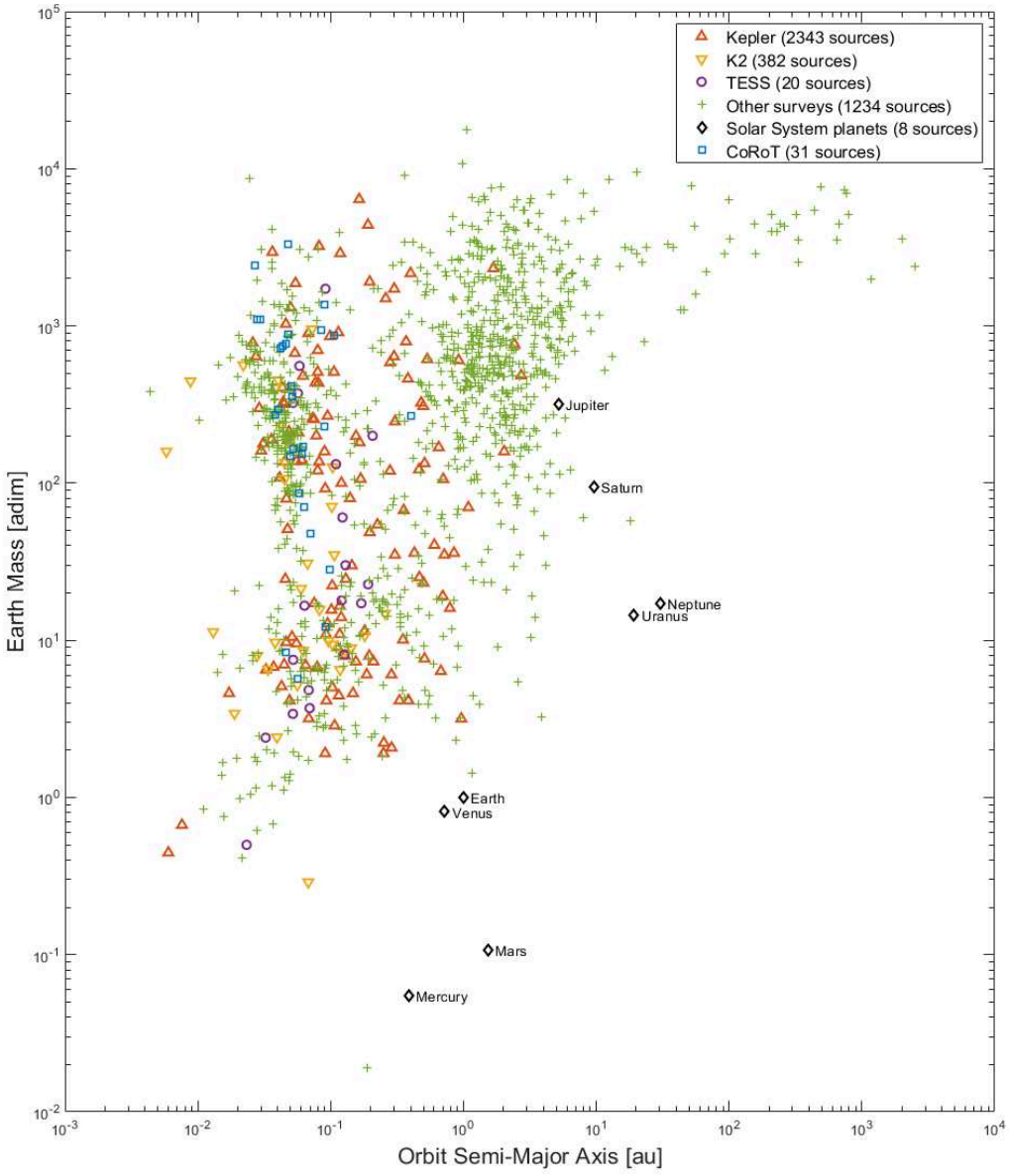
Since the first detections, a substantial number of exoplanets were discovered and confirmed. To date, this number adds up to 4,009 exoplanets according to the NASA Exoplanet Archive[1]. However, most of these planets (see Figure 3) are either much closer to their host star or much more massive when compared to the Earth (or both). Hence, even considering the important advances occurred in the matter of the exoplanet search science over the last two decades, which allowed us to discover thousands of new exoplanets, we are still incapable to assert if there are effectively other planets like our Earth, let alone how many of them and what type of star they orbit.

Besides, we cannot state either that Earth-like planets are simply rare, so that we can consider that we live in an "outlier" planet. The reason is that most of the stars around which the planets of Figure 3 were detected are, typically, relatively faint (see Figure 4). In other words, the photometry of these stars do not have sufficient signal-to-noise ratio (SNR) allowing Earth-sized planets – orbiting within their habitable zone – to be detected

[1] <https://exoplanetarchive.ipac.caltech.edu>

and accurately characterized in terms of mass and radii. Therefore, the fact that we have not found yet a planet like our Earth could be owed to the fact that we have still not observed a sufficiently large sample of bright stars.

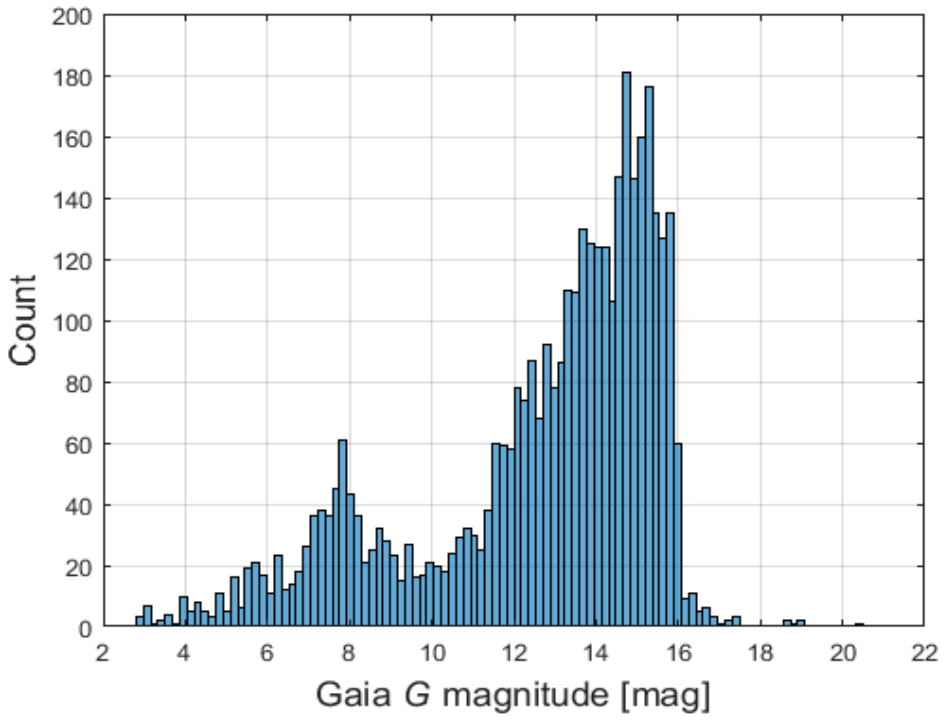
Figure 3 – Confirmed exoplanets to date (4,009 in total).



Source: Nasa Exoplanet Archive.

The only spatial mission – within the next few years – that is qualified to find new exoplanets, TESS, will not be capable of making significant impact on the detection and characterization of Earth-sized planets, except for those orbiting coolest M-type stars, for which the mission is mostly designed for. Therefore, a clear gap exists when it comes to find Earth-like planets orbiting the habitable zone of Sun-like stars.

Figure 4 – Magnitude of the stars hosting the planets of Figure 3.



Source: Nasa Exoplanet Archive.

In this respect, the ESA planet-hunter spatial mission PLATO appears as a promising solution to cover this gap. Expected to start operating in 2026, the PLATO mission takes place in the context of ESA's long-term planning for space science missions called Cosmic Vision. This Program was conceived to address fundamental scientific questions related to the origin of our Universe and the physical laws that drives it, the formation of planetary systems and the origin of life. In that thematic, PLATO aims at investigating three major questions (ESA (2017)):

- How do planets and planetary systems form and evolve?
- Is our solar system special or are there other systems like ours?
- Are there potentially habitable planets?

To do it, PLATO builds upon the well-proven techniques that allowed the success of its predecessors CoRoT and *Kepler* missions. That is the transit method for detecting exoplanets, along with radial velocity spectroscopy follow-up from the ground, and the analysis of stellar oscillations via asteroseismology for characterizing their host stars. However, PLATO stands out from any other mission of its category in the sense that it is being specially designed to not only discover but also characterize (in terms of mass, radii, and density) Earth-like planets orbiting within the habitable zone of bright main-sequence Sun-like stars. Moreover, by also focusing on constraining stellar ages PLATO might be capable of determining changes in planetary systems architecture over the time, such as the dependency of exoplanet frequency with main-sequence stellar age (Veras *et al.* (2015)). Ultimately, PLATO is expected to provide enough elements allowing us to finally determine with accuracy whether Earth-like planets exist or not, how many of them and which type of star they orbit (Baudin & Damiani (2019)).

The work presented in this thesis takes place in the preparation phases of PLATO space mission, more precisely in the development of data processing algorithms.

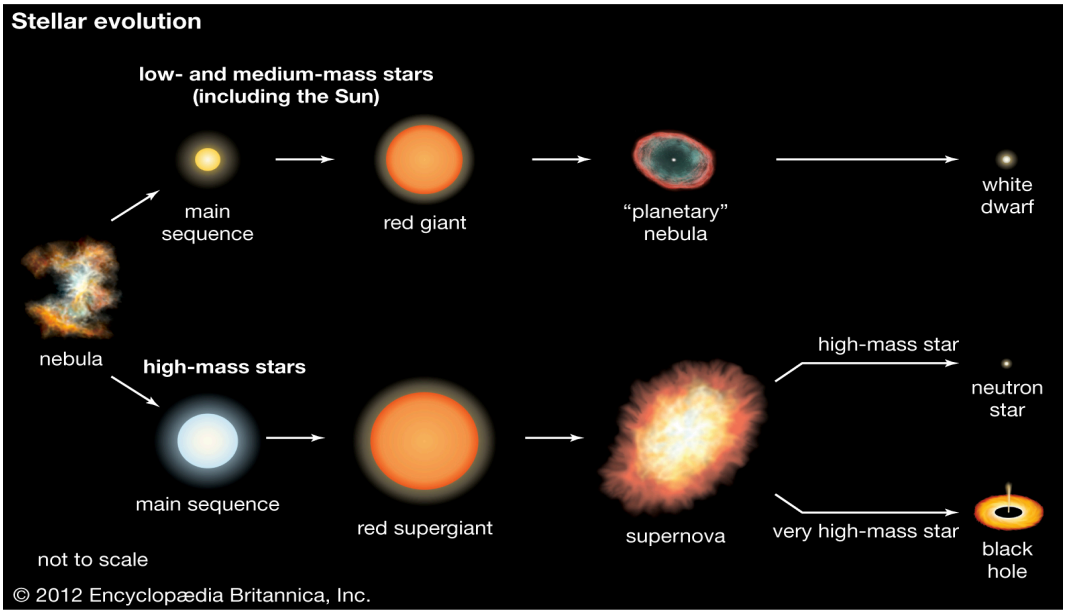
1.2 Stellar classification

Simply stated, a star is a spherical celestial body of plasma (i.e. super hot gas) whose formation is resulted from the gravitational collapse of molecular clouds of cold gas – mostly composed of hydrogen and helium – that are present in the interstellar medium. As the clouds accumulate, they form a central region – called protostar (see Campante, Santos & Monteiro (2018)) – that becomes denser and hotter than the outer regions. At a certain point owing to the increasing pressure, the protostar temperature becomes high enough so that fusion reactions starts to take place in its core, thereby causing hydrogen to be converted in helium. Such splendid moment characterizes the birth of a star. From that point on, the star's life expectancy is determined by the nuclear time-scale associated to its total fusing hydrogen mass (fuel). Hence, the more massive the star, the quicker it consumes hydrogen, thereby resulting in shorter lifespan, and vice-versa. The amount of gas and dust material orbiting a young star, known as the circumstellar disc, may also accumulate away from that star to form planets (Fortier *et al.* (2012)).

The different evolutionary states of a star throughout its lifetime (see schematic in Figure 5) are commonly represented using a Hertzsprung–Russell diagram (HRD), which displays the correlation between stellar surface luminosity and effective temperature. An example of HRD produced with stars observed by *Gaia* (Gaia Collaboration *et al.*, 2016) is shown in Figure 6, with stellar luminosities normalized by that of the Sun. The diagram clear evidences that most of the stars found in our sky belongs to the main sequence branch, which is consistent with the fact this corresponds to the longest phase in stellar evolution. Effective temperature and abundances of heavier elements than hydrogen or

helium (i.e. metallicity) defines the spectral classification of a star, which is designated with the letters O, B, A, F, G, K, and M (in order of decreasing temperature). Besides, stars are also subdivided according to their luminosity class (Cox (2002)): I (supergiants), II (bright giants), III (giants), IV (subgiants), and V (dwarfs or main sequence). Moreover, the effective temperature of a star has a direct link with the colours form the electromagnetic spectrum. Accordingly, coolest stars look redder and hotter stars look bluer.

Figure 5 – Stellar evolution scheme.



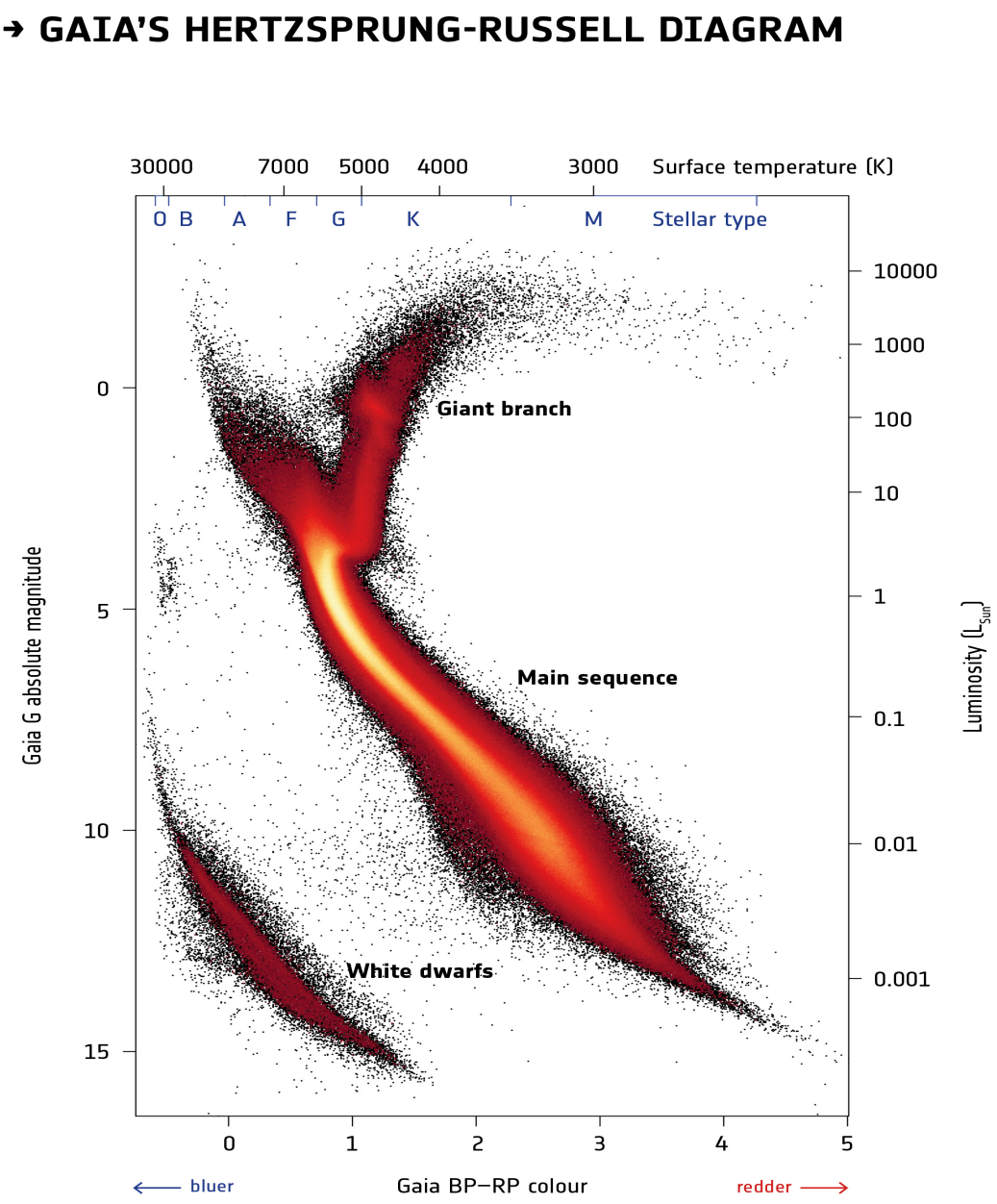
Source: Encyclopedia Britannica, Inc.

Our Sun is a main sequence (dwarf) yellow G2V-type star with effective temperature of about 5,800K. Its total life expectancy is about 10 billion years (Campante, Santos & Monteiro (2018)). M-type stars of the main sequence branch have a fraction of the Solar mass, so they are relatively faint and cool ($\lesssim 3,500$K) stars, and have an estimated lifespan of the order of trillion years (Adams, Graves & Laughlin (2004)), thereby extremely longer than the age of the Universe (about 13.8 billion years). In other words, such stars can be considered relatively young in their evolution. Also, this class of star is the most commonly encountered in the Milky Way (Henry, Kirkpatrick & Simons (1994)). Lastly, the earlier mentioned red giants are relatively cool (K or M) spectral type stars, but much brighter than our Sun as previously explained.

In terms of planetary formation, latest observations evidenced the existence of a direct correlation between stellar metallicity and the occurrence of gas giants (e.g. Neptune- and Jupiter-like) planets (Fischer & Valenti (2005), Johnson *et al.* (2010)), which however do extend to small ones, more specifically for those whose radius is smaller than four times that of the Earth ($\mathcal{R}_p < 4\mathcal{R}_\oplus$). Furthermore, analysis based on *Kepler* data indicate that

terrestrial or smaller planets are not constrained by environment metallicity, suggesting therefore that such category of planets might be relatively common across our Galaxy (Buchhave *et al.* (2012)).

Figure 6 – Hertzsprung–Russell diagram of stars observed with *Gaia*.



Source: ESA/*Gaia*/DPAC.

1.3 Probing stellar interior with asteroseismology

In astrophysics, stellar physics is the science of studying the internal structure of stars and how these evolve throughout time, which involves understanding several physical and chemical notions such as hydrodynamics, thermonuclear reactions, radiative transfer, quantum mechanics and general relativity. Although there exist no actual technology capable to directly explore the interior of stars (i.e. with physical instruments inside them), their composition, functioning and evolution can be fairly well understood thanks to the great advances – achieved in the last few decades – in the research field of asteroseismology (Baglin *et al.* (2006), Aerts, Christensen-Dalsgaard & Kurtz (2010), Mosser & Miglio (2016), Campante, Santos & Monteiro (2018)).

Asteroseismology is a branch of stellar physics that allows one to probe the internal structure of stars by studying the seismic waves that propagate inside them, as analogous to the approach used in terrestrial seismology. In other words, in the same way as analysing the seismic waves generated by earthquakes provides information about Earth's interior (temperature, pressure, rocky composition etc.), analysing the oscillation modes of a star provides us key information about its internal physical properties and dynamics (e.g. mean density, chemical composition, rotation etc.).

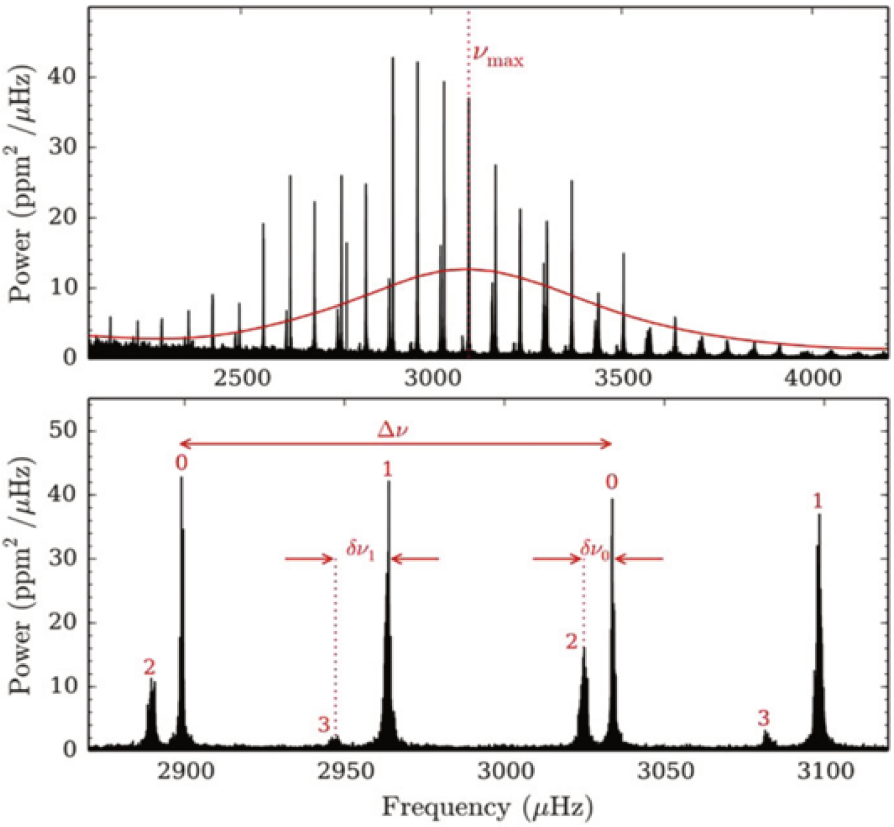
Stellar oscillations cause periodic variations in star brightness that exhibit regular patterns in the frequency domain (see Figure 7). These patterns, which carry important information on specific characteristics of the stellar structure, can be measured by a sufficiently sensitive (high signal-to-noise ratio) photometer. For example, the average large separation $\Delta\nu$ between larger peaks contains information about stellar mean density, whereas the small separations $\delta\nu$ carry finger prints of the chemical composition in the stellar core, which can be used to infer the amount of hydrogen in it and ultimately the corresponding stellar age. Besides, inversion techniques (Reese *et al.* (2012), Buldgen *et al.* (2015)) can be applied to several observed and measured oscillation mode frequencies to determine e.g. the internal rotation profile of the star. In particular, stellar rotation may impact on both internal structure and evolution of stars (Gehan (2018)), including their ages (Lebreton & Goupil (2014)).

Asteroseismic studies can be performed through Doppler spectroscopy from ground-based observations. However, since measuring stellar oscillations requires sufficiently high duty-cycle and low-noise observations, better results in asteroseismology are obtained from spectral analysis of photometry signals extracted from stars with space-based surveys.

1.4 Detecting exoplanets

The exoplanet search science counts on a variety of detection methods, two of which are widely employed: the transit photometry and the radial velocity. Indeed, among

Figure 7 – Power density spectrum of the Sun.



Source: Campante, Santos & Monteiro (2018).

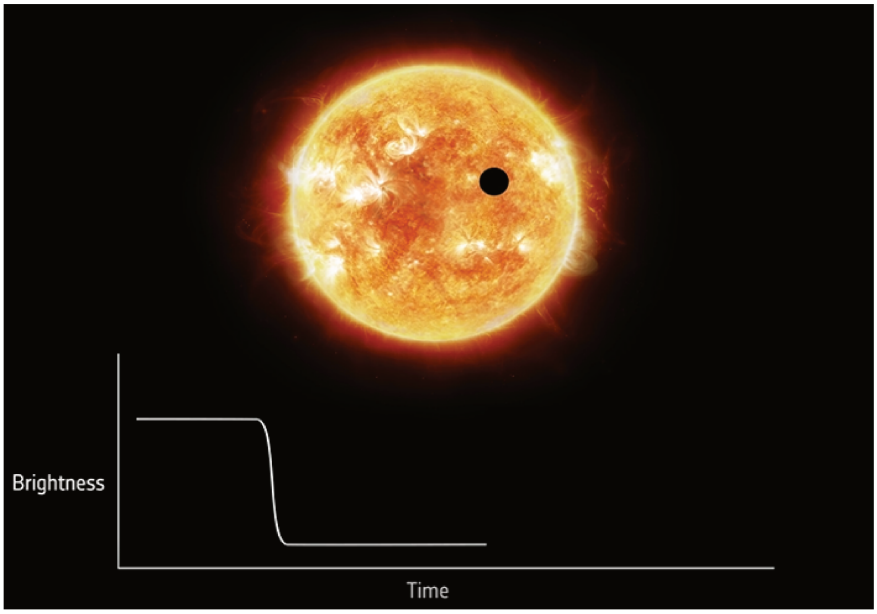
all the confirmed planets of Figure 3, which were detected using ten different methods, transit photometry and radial velocity account respectively for ∼ 77% and ∼ 19%, that is ∼ 96% altogether, of these discoveries. The following sections provide brief descriptions of both methods. This thesis is focused in the science and usage of the transit method, on which the PLATO mission relies for detecting exoplanets. A complete list with detailed descriptions of planet detection techniques is provided in Perryman (2018).

1.4.1 The transit photometry method

The principle of the transit photometry method is simple: a planet passing in front of (eclipsing) its parent star (see representation in Figure 8) causes an apparent dip in its brightness whose amplitude is proportional to the second power of the planetary-to-stellar radius ratio.

Accordingly, dips produced by larger planets such as gas giants are much easier to detect than those produced by terrestrial planets. For example, a Jupiter-like planet orbiting a Sun-like star causes a stellar brightness dip (transit depth) which is about two orders of magnitude larger than that produced by an eclipse of an Earth-like planet

Figure 8 – Representation of transit photometry method.



Source: ESA.

orbiting that same type of star. In contrast, another important aspect that significantly influences in the detection capabilities of transit method is the angle between the observer's line of sight and the planetary orbital plane. Assuming planets with randomly oriented orbits, their corresponding visibility angles in the sky represent a small fraction of the celestial sphere, decaying exponentially with respect to the planetary-to-stellar distance. For example, terrestrial planets orbiting at 1 AU distant from Sun-like stars produce transits that are visible from only $\sim 0.46\%$ of the celestial sphere. For a Jupiter-like planet orbiting at 5.2 AU distant from Sun-like stars, that visibility drops to only $\sim 0.09\%$ (Bozza, Mancini & Sozzetti (2016)), that is five times lower visibility probability (although the transit depth produced by such planet is significantly easier to detect as explained before). In any case, the intrinsic low probability of having a planet transit visible in the sky shows that building a statistically significant sample of exoplanets requires observing several thousands of stars over at least a few years.

The first known exoplanet to be discovered with transit photometry dates from the early 2000's. The authors of this discovery reported a transiting object causing a 1.2% transit depth in its host star's light curve, giving an estimated planetary radius of about 1.3 Jupiter radii. Later radial velocity measurements from that object suggested it to be a planet with around 90% of the Jupiter mass in an orbit at only 0.023 AU distant from the host star (Konacki *et al.* (2003)). Hence, given its proximity to the parent star and physical similarity to Jupiter, this planet enters in the category of planets called "hot Jupiters". Hot Jupiters are virtually the easiest planets to detect since they are very

large and close to the host star, thereby producing large transit depths and having large fractions of the celestial sphere through which observers can see it. Not surprisingly, there is a relatively high occurrence of hot Jupiters among the confirmed exoplanets shown in Figure 3. Terrestrial planets around Sun-like stars, in contrast, are very tough to be detected, in particular because of the strong requirements in terms of noise performance. Indeed, detecting such category of planets requires photometric precisions of the order of a few dozens parts-per-million (ppm), which requires therefore observing sufficiently bright targets.

1.4.2 The radial velocity spectroscopy method

A planet orbiting its host star causes the latter to wobble around the centre of mass of the system formed by both celestial bodies. From an observer's point of view, the light it receives from the wobbling star periodically shifts in wavelength over the time, going redder when it moves away and going bluer when it moves towards the observer (see representation in Figure 9), analogously to the change in wavelength that occurs in sound waves owing to the Doppler effect.

The radial velocity (or Doppler spectroscopy) method consists therefore in measuring the tiny ($\sim 10^{-4}$Å) wavelength shift in the light from a wobbling star, which is then translated into a corresponding stellar radial velocity projected in the direction of the radius connecting the star and the observer. That velocity depends on both star and planet masses. Since the former can be obtained from the stellar oscillation analysis through asteroseismology, the mass of the planet can finally be determined.

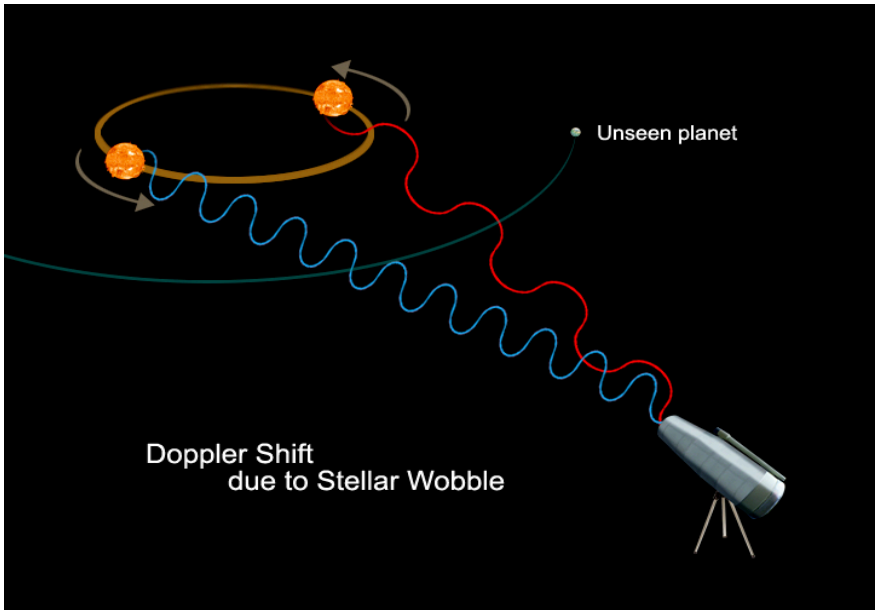
Radial velocity has shown to be a powerful method for detecting and characterizing masses of exoplanets. However, the main drawback of this method is the fact that it can only provide unambiguous mass estimations if the inclination angle between the observer's line of sight and the orbital plane of the planet is known. Otherwise, the method is limited to estimate only the planet's "projected mass" in the direction of the observer's line of sight, that is the minimum planet mass (Campante, Santos & Monteiro (2018)).

1.4.3 False transit signatures

Planets orbiting stars are not the only astrophysical sources capable of producing transit-like signals in light curves. Binary stars, which are known to exist since William Herschel back in the 1700's, can naturally produce transit dips as well. Therefore, if one looks at finding new exoplanets, then one should be capable of distinguishing legitimate planetary transits (i.e true positives) from those which are not (i.e. false positives).

In the context of the *Kepler* mission, a concept was created to designate statistically significant transit-like signatures marked for further data validation: the threshold crossing events (TCE) (see e.g Twicken *et al.* (2018)). Each of these events become a *Kepler* Object

Figure 9 – Representation of radial velocity (Doppler spectroscopy) method.



Source: NASA.

of Interest (KOI) (i.e. a planet candidate) which shall be subjected to proper vetting process in order to check whether they are true planets or not.

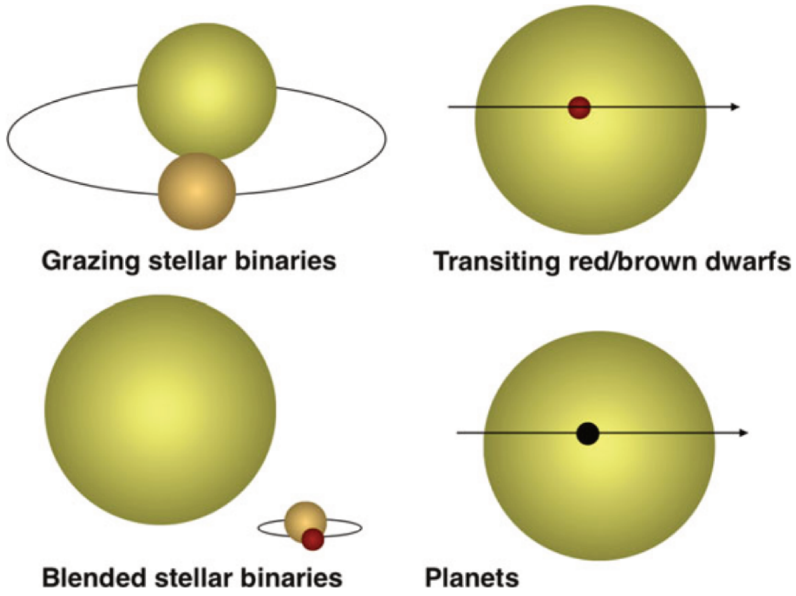
There are principally four astrophysical phenomena (Figure 10) capable of producing statistically significant transit-like signatures (Bozza, Mancini & Sozzetti (2016)):

1. Planets orbiting stars;
2. Grazing eclipsing binaries with similar mass and size;
3. Medium- or high-mass stars with low-mass (e.g. red dwarf) stellar companions;
4. Blended eclipsing binaries.

The latest three are examples of false positives. During vetting process, some of these false planet transits can be quickly identified by looking at their shape. For example, grazing stellar binaries produce transit signals that are V-shaped, whereas legitimate planet transits are U-shaped. Besides, a transit produced by eclipsing binaries with different effective temperatures is inevitably colour-dependant, which is naturally not the case of a true planet transit. Furthermore, eclipsing binaries periodically present two distinct transit depths, since both stars of the system are eclipsed alternately.

A blended eclipsing binary system consists of a isolate foreground (target) star, i.e. a star from which we are interested to extract photometry from, with a background

Figure 10 – Astrophysical phenomena producing transit-like signatures.



Source: Bozza, Mancini & Sozzetti (2016).

eclipsing binary (BEB) companion. Their separation in the sky is small enough so that the BEB can pollute the photometry of the target star on the detector of a photometer.

For the *Kepler* mission, Bryson *et al.* (2013) developed methods for detecting background false positives based on changes in the star image that occur during a transit. The authors found that, at low Galactic latitudes, background false positives account for near 40% of all *Kepler* transit-like signals.

1.5 Extracting photometry from stars

Analysing stellar oscillations through asteroseismology and detecting exoplanets with transit photometry requires extracting light curves from targets stars. One of the key point in this task consists in properly quantifying the amount of signal and noise that is embedded in the light curve. Since stars emits photons following a Poisson distribution, the photometric flux signal measured from a star has variance σ^2 and mean f such that $f = \sigma^2$. Accordingly, the SNR of the *raw* stellar flux is $\text{SNR} = f/\sigma = f/\sqrt{f} = \sqrt{f}$, i.e. equal to the square root of the mean flux. Higher SNR requires therefore a proportionally higher average flux, which can be obtained by augmenting the interval during which photons from the target star are collected. In contrast, measuring the total photon count from a star necessitate a detector – typically a charge couple devices (CCD) in spatial applications – which counts photons in discrete packages (pixels) that contribute, each, with additional noise in the photometry. That amount of extra noise depends on the quality of the detector and on the Point Spread Function (PSF) of the optics that drives the (point source) flux

from a star towards a given position of the detector. The more distorted the PSF, the larger the number of pixels required to register the star signal on the detector, thereby the greater is its noise contribution in the photometry. Furthermore, sources of photon flux from the sky such as contaminant stars and scattered zodiacal and Galactic lights also contribute to the total noise embedded in the photometry extracted from a star.

In all such context, extracting high precision light curves from stars is far from being a straightforward task. While several photometry extraction methods exist in the literature, they are all essentially derived from two major techniques: PSF fitting and aperture photometry. We provide a general description of both methods in the following sections.

1.5.1 PSF fitting photometry

The PSF fitting method consists in fitting a PSF-based image model of a star to its measured one, such as to minimize the following functional:

$$\chi^2 = \sum_{i,j} \frac{\left(I_{i,j} - \hat{I}_{i,j}\right)^2}{\sigma^2_{i,j}}, \tag{1.1}$$

where $I_{i,j}$ corresponds to the measured star image, $\hat{I}_{i,j}$ the modelled star image and $\sigma^2_{i,j}$ the variance of the measured flux at pixel coordinate (i, j). The modelled star image $\hat{I}_{i,j}$ can be defined as (Deheuvels & Ballot (2019)):

$$\hat{I}_{i,j} = a \times P_{i,j}(x_c, y_c) + b, \tag{1.2}$$

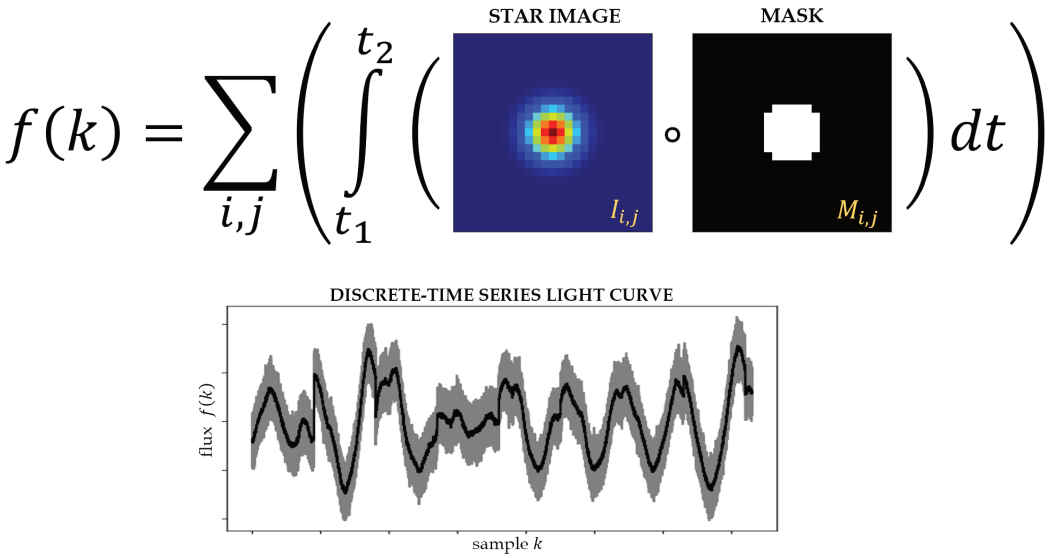
where $a_{i,j}$ is the stellar flux at pixel coordinate (i, j), $b_{i,j}$ is the background flux at pixel coordinate (i, j), and $P_{i,j}(x_c, y_c)$ is a PSF description at pixel coordinate (i, j) with centroid coordinate (x_c, y_c). Hence, the free parameters of the fit are x_c, y_c, a, and b.

The essential objective of this method is to determine, at pixel level, what are the contributions of the target star flux and the background flux to each pixel of the measured star image $I_{i,j}$. Going further, this method can be refined to include contributions from contaminant stars present in the star image scene. By doing so, one is then capable to optimally detach the average flux of the target star from that of contaminant sources and the diffuse background light. Moreover, the PSF fitting technique has also the advantage of providing relative accurate ($\sim 10^{-2}$ pixel) estimates on the centroid position of stars and is less sensitive to satellite jitter and long-term stellar position drift. In contrast, since – in real world scenario – uncertainties on the knowledge of the instrument PSF always exist, this method is mostly limited by the quality of the PSF description $P_{i,j}(x_c, y_c)$.

1.5.2 Aperture photometry

In aperture photometry, each light curve sample is generated by integrating the target flux over a limited number of pixels which shall be appropriately selected to maximize the scientific exploitability of the resulting time-series light curve.

Figure 11 – Schematic of aperture photometry method to produce a time-series light curve. The symbol "∘" stands for the Hadamard (element-wise) product applied between the (generic) star image $I_{i,j}$ and the (generic) aperture $M_{i,j}$, both of identical shape.



Source: author, with plot from Samadi *et al.* (2019).

Mathematically, the aperture photometry method can be described as

$$f(k) = \sum_{i,j} \left(\int_{t_1}^{t_2} \left(I_{i,j} \circ M_{i,j} \right) \, dt \right). \tag{1.3}$$

In the above expression, $I_{i,j}$ is a square matrix of indexes i and j representing the image of a star; $M_{i,j}$ is a square matrix of identical shape of $I_{i,j}$ representing an aperture (mask); t is the time in continuous domain; $(t_2 - t_1) > 0$ is the detector integration time; k represents the index of a single light curve sample; and $f(k)$ if the photometric flux relative to the light curve sample k. The symbol "∘" stands for the Hadamard (element-wise) product applied between the star image $I_{i,j}$ and the aperture $M_{i,j}$. A discrete-time series light curve results from a subsequent sequence of $f(k)$ as represented in Figure 11.

1.6 This thesis

The PLATO mission is expected to observe up to one million stars, depending on the final observation strategy. In contrast, transmitting to the ground individual images from each of these stars at sufficiently short cadence for further processing requires prohibitive telemetry resources. Hence, for a substantial fraction of the targets, an appropriate in-flight data reduction strategy (prior to data compression) needs to be executed. For that, the most suitable encountered solution consists in producing their light curves on board.

In view of its acknowledged high performance and straightforward implementation, mask-based (aperture) was adopted as photometry extraction method to produce light curves in flight. In such context, the present work unfolds the development carried out for defining the optimal collection of pixels (i.e. the aperture $M_{i,j}$, Equation 1.3) for extracting photometry on board from a significant fraction of the PLATO targets. Compared to the common approaches found in the literature to determine aperture shapes, this work brings a novel perspective through which greater importance is given to the problematic of background false positives (subsubsection 1.4.3). The major motivation for that is to provide ways of eliminating astrophysical false positives as early as possible in the planet discovery process. This is justifiable since although effective techniques exist for detecting false positives (e.g. Bryson *et al.* (2013)) from the observations, in most cases the vetting process of planet detections usually requires ground-based confirmations via radial velocity measurements that consume (costly) telescope time. Accordingly, using such infrastructure to identify false positives – that could be earlier rejected by the time of the detection – represent significant waste of money and time. Furthermore, the vast majority of light curves produced in flight will not have pixel data available on the ground for the identification of false positives. The main challenge involved in this work relies on the fact that it needs to propose an aperture photometry solution delivering sufficiently high photometric precision to be in agreement with the science requirements defined for the PLATO mission, and sufficiently low sensitivity to false planet detections; all that for a huge number of targets with the limited CPU and memory resources available in the spacecraft payload.

The following of this document[2] is organized as follows: Chapter 2 provides an overview of the PLATO mission including its science objectives and requirements, envisaged observation strategies and data products. This chapter also describes the main payload characteristics, including instrument point spread function (PSF), spectral response, and noise. Chapter 3 gives details on the extracted data from the adopted input catalogue (*Gaia* DR2). That information is used to build synthetic input images, called imagettes, to characterize the performance of aperture photometry. A synthetic PLATO P photometric

[2] This document uses the LaTeX typesetting package **abnTeX2** for technical and scientific documents. This package is based on ABNT rules that are required by the University of São Paulo in Brazil.

passband, calibrated on the VEGAMAG system, is derived to avoid the inconvenience of having colour dependency when estimating stellar fluxes – at detector level – from visual magnitudes. Colour relationships with Johnson's V and Gaia G magnitudes are thus provided. Moreover, an expression is derived to provide an estimation on the intensities of zodiacal light entering each PLATO telescope. Chapter 4 describes the methodology applied to find the optimal aperture model to extract photometry from stars in P5 sample. Three models are tested, including a novel direct method for computing a weighted aperture providing global lowest NSR. The chapter ends by showing comparative results between all aperture models with respect to their sensitivity in detecting true and false planet transits. Finally, Chapter 5 concludes with discussions on the presented results and perspectives/open issues for future work.

2 The PLATO space mission

PLAnetary Transits and Oscillations of stars (PLATO)[1] (Rauer *et al.* (2014)) is a space mission from ESA whose science objective is to discover and characterize new extrasolar planets and their host stars. Expected to be launched by the end 2026, this mission will focus on finding photometric transit signatures of Earth-like planets orbiting the habitable zone of main-sequence Sun-like stars. Thanks to its instrumental concept comprising multiple telescopes covering a very large field of view, PLATO will be able to extract long duration photometry from a significantly large sample of bright stars at very high photometric precision, allowing it to accurately characterize planetary and stellar parameters such as mass, radii, density and age.

Figure 12 – Artist's impression of PLATO spacecraft.

Source: OHB System AG.

[1] See <https://www.cosmos.esa.int/web/plato> and <https://platomission.com/>.

In this chapter, we provide an overview of the very foundation on which the PLATO mission and the present work are built. We start by giving a brief history of the project until its current development stage, then we move along the top science goals of the mission and derived requirements, covering the aspects of observational constraints and strategies. Next, a description of the instrument concept is given, including a summary of the main payload characteristics containing all the instrument parameters used in this work. We also provide a overview of the science data processing pipeline, in which context the present work takes place.

The work presented in this chapter is partially based on Marchiori *et al.* (2019). In-flight photometry extraction of PLATO targets: Optimal apertures for detecting extra-solar planets, A&A, 627, A71.

2.1 A brief history of the project

The PLATO mission was first proposed to ESA by a broad group of European scientists headed by Dr. Claude Catala (Paris Observatory), in response to the Call for (M-class) mission proposals of Cosmic Vision 2015-2025 released on March of 2007. From a total of 52 proposals, PLATO was pre-selected along with five other projects.

In June 2011, PLATO successfully completed phase A (feasibility evaluation), after having gone through assessment and definition studies which involved two independent industries to investigate the mission concept, and a Consortium of research Institutes and Universities to study the payload. The proposal however was not selected for the M1 or M2 launch opportunities as originally planned. Shortly after its non-selection, PLATO candidature was re-submitted for the M3 launch opportunity with renewed science case and mission design. One of the major changes in the new proposal was the transfer of the leading role from France to Germany, with Prof. Heike Rauer (DLR) taking place as PLATO Principal Investigator. On February 19th 2014, PLATO was selected by ESA for the M3 launch opportunity in 2022–2024.

After the selection, the mission entered in phase B (preliminary definition), this time involving three concurrent industrial contractors (Airbus Defence and Space, OHB System AG, and Thales Alenia Space) for designing the spacecraft. After being subjected to a thorough independent review at ESA, the new PLATO Science Management Plan was approved by the ESA Science Programme Committee in June of 2016. The approved plan included a (revised) baseline payload configuration comprising 26 telescopes with nominal science operations of fours years (plus a verified in-orbit lifetime of 6.5 years and eight years of consumables). The ESA Science Programme formally adopted the PLATO mission in June of 2017. In May of 2018, OHB System AG – a subsidiary of Bremen-based space and technology group OHB SE – was selected as the prime contractor of PLATO. The prime is responsible for constructing the spacecraft platform (which includes the main

satellite structure, the propulsion system, the solar panels, the attitude and orbit control system, among others features) and integrating it with the payload.

The mission is currently in end of phase B and shall be ready to start phase C (detailed definition and implementation) by end 2019. A recent important milestone for the mission was the delivery to ESA of a first batch of 20 charge-coupled devices (CCDs) in mid March of 2019.

2.2 Mission Consortium

The PLATO Mission Consortium (PMC) regroups several hundreds of scientists from almost all ESA Member States, including a few scientists from the United States and Brazil. The PMC is lead by the project Principal Investigator and is responsible for building, integrating, verifying, calibrating and delivering the PLATO payload to ESA. The payload subsystems include camera optical elements and detectors, flight hardware and software, electronics, as well as the pipeline algorithms and modules to generate high level scientific data products.

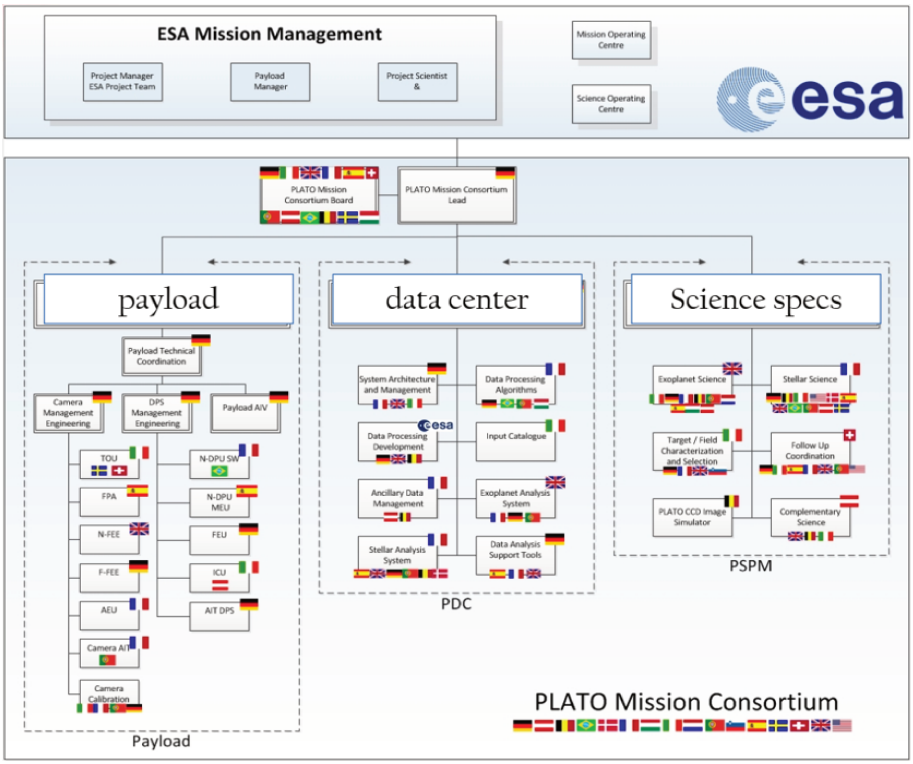
The PMC structure (Figure 13) includes the PLATO Data Centre (PDC) and the PLATO Science Management (PSM) teams. The PDC team is responsible for developing methods and tools for analysing, validating and calibrating the data collected by the spacecraft. The PSM provides scientific specifications for the PDC, specifies the PLATO input catalogue and organises preparatory and follow-up observations.

2.3 Science goals

The PLATO mission is in synergy with the fundamental questions that drives the ESA Cosmic Vision Program. Accordingly, the core design of the mission was established with top level science goals that can be broken down into the following specific objectives (PLATO Study Team (2017)):

1. Determine the bulk properties (mass, radii, and mean density) of planets in a wide range of systems, including terrestrial planets in the habitable zone of solar-like stars;
2. Study how planets and planet systems evolve with age;
3. Study the typical architectures of planetary systems;
4. Analyse the correlation of planet properties and their frequencies with stellar parameters;
5. Analyse the dependence of the frequency of terrestrial planets on the environment in which they formed;

Figure 13 – Overview of PLATO PMC structure.



Source: Dr. Heike Rauer, principal investigator of the PLATO Mission Consortium.

6. Study the internal structure of stars and how it evolves with age[2];

7. Identify suitable targets for spectroscopic follow-up measurements to investigate planet atmospheres.

The above science goals should allow PLATO – among other aptitudes – to better constrain planet formation and evolution models, identify potential candidates for habitable planets, improve stellar models and ages in general, correlate planet parameters with stellar properties and correlate planet occurrence frequency with environment.

To achieve its science goals, the PLATO mission aims at building and characterizing a statistically significant sample of planets down to Earth-size orbiting within the habitable zone of bright main sequence F, G, K Solar-type stars and M-stars. To accomplish it, the mission builds upon the transit method for detecting exoplanets, spectroscopic radial-velocity follow-up and TTVs measurements for determining exoplanet mass, and asteroseismology to determine masses, radii and age of exoplanet host stars.

Following that strategy, PLATO should be capable of delivering seismic characterization of a large sample of bright stars across the HR diagram (PLATO Study Team

[2] The age of the planet is assumed to be the age of the host star.

(2017)). Moreover, PLATO will be the first mission of its category being capable of providing a significant sample of (super) Earth-like planets orbiting within the habitable zone of Sun-like stars (Goupil (2017)). Furthermore, coupling accurate parameters of planets with those from their (bright) host stars should provide a consistent census of exoplanetary systems neighbour of our own Solar System.

2.4 Science requirements

We present here below a non-exhaustive list containing the major scientific requirements established for both asteroseismology and exoplanet search sciences of the PLATO mission, derived from the science goals described in the previous section. PLATO shall be capable of (PLATO Study Team (2017), ESA (2017), Goupil (2017)):

- detect a planet orbiting a G0V star with an orbital period of one year (this is roughly equivalent to being capable of detecting an Earth-like planet orbiting a Sun-like star at a distance of 1 AU);
- determine – with accuracy better than 2% – radius of G0V stars as bright as $V = 10$;
- determine – with accuracy better than 15% – mass of G0V stars as bright as $V = 10$;
- determine – with accuracy better than 10% – ages of G0V stars as bright as $V = 10$;
- detect and characterize terrestrial planets orbiting dwarf and sub-giants stars brighter then $V \sim 8$ and of spectral type F5 to late-K at distances including the habitable zone of such stars;
- provide dual photometric band information for stars brighter than $V \sim 8$;
- detect planets orbiting dwarf and sub-giants stars of spectral type F5 to late-K at distances including the habitable zone of such stars;
- detect terrestrial[3] planets orbiting M-dwarf stars at distances including the habitable zone of such stars;
- determine – with accuracy better than 3% – radius of detected planets down to Earth-size orbiting G0V stars as bright as $V = 10$;
- determine – with accuracy better than 2% – planetary-to-stellar radius of detected planets down to Earth-size orbiting G0V stars as bright a $V = 10$;

[3] A terrestrial planet is understood herein as being a planet whose radius $\mathcal{R}_p$ and mass $\mathcal{M}_p$ satisfy $\mathcal{R}_p < 2\mathcal{R}_\oplus$ and $\mathcal{M}_p < 10\mathcal{M}_\oplus$.

- deliver photometric data allowing to determine – through radial velocity measurements and with accuracy better than 10% – mass of terrestrial planets orbiting G0V stars;

- deliver photometric data allowing to determine – with precision of the order of $0.1\,\mu$Hz for main sequence stars – frequencies of normal oscillation modes above and below the mode with the maximum amplitude;

- observe between 10% and 50% of the sky with observation durations of at least two months;

- sustain in-orbit nominal science operations during at least four years.

The above science requirements translate – at instrument level – into noise requirements (see Figure 14) which can be summarized as follows

(A) the residual errors from systematic effects in the light curves of stars brighter than $V \sim 10$ must be limited to two thirds of their photon noise;

(B) The item (A) above must hold within the frequency range comprised between $20\,\mu$Hz and 40 mHz;

(C) The total random noise of stars brighter than $V \sim 10$ must be limited to $34\,\text{ppm}\,\text{hr}^{1/2}$.

(D) Noise requirements must be assured in the wavelength range between 500nm and 1000nm (visible and near infrared domains).

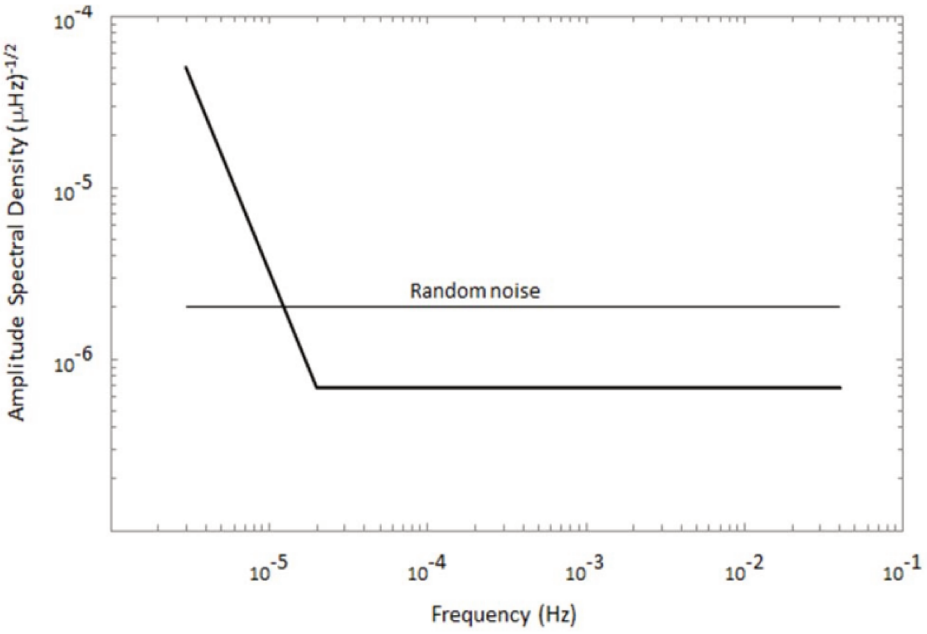
Satisfying the requirement (A) ensures that the total random noise in the light curve of a star brighter than $V \sim 10$ is dominated by its photon noise.

The frequency range of requirement (B) includes the time-domain interval comprised between a few minutes for the detection stellar oscillations and a few hours for the detection of planetary transits.

Satisfying the requirement (C) ensures that the oscillation modes of Solar-type stars can be identified, and that Earth-like planets orbiting in the habitable zone of Sun-like stars can be detected and their planetary-to-stellar radius characterized with accuracy better than 2% (ESA (2017)).

The requirements (A), (B), (C), and (D) altogether also work as drivers for linking science and engineering requirements (e.g. spacecraft pointing error, thermal control, electronics, spectral range etc.).

Figure 14 – Random noise requirements at instrument level.



Source: PLATO Study Team (2017).

2.5 Stellar samples

To achieve the science requirements, the target stars of the PLATO mission are categorized into four distinct samples (see Table 1), namely P1, P2, P4, and P5, following a criterion of scientific priority.

The Sample P1 represents the core science of the mission and holds therefore the highest priority. It is composed of dwarf and sub-giants stars of spectral type between F5 and K7, and visual magnitude $V \lesssim 11$. Since these stars are relatively bright and have images acquired at 25 seconds cadence, they will have very high photometric precision ($\lesssim 50\,\mathrm{ppm\,hr}^{1/2}$). As a consequence, ground-based radial velocity follow-up is expected to be more effective for this stellar sample.

The Sample P2, second in the order of priority, contains the brightest stars ($V \lesssim 8.2$) to be observed by the mission. Since their photometric fluxes exceed the saturation limit of the detectors of the normal cameras, their photometry will be extracted with the fast cameras, which have shorter image acquisition cadence (2.5 seconds). The spectral type of the stars in this sample is the same as that of P1 targets.

The Sample P4, third in the order of priority, regroups M-type dwarf stars with $V \lesssim 16$. Because these stars are cooler than those of the other samples, their habitable zones are closer. Therefore, the planets of interest orbiting these stars have shorter orbital periods, typically on the order of a few weeks.

Lastly, the Sample P5 includes a massive number ($> 245,000$) of F5 to late-K dwarf and sub-giants stars with visual magnitude in the range $8 \lesssim V \lesssim 13$. This sample aims to generate large statistical information on planet occurrence rate and systems evolution. For comparison, the *Kepler* and TESS missions – of the same category of PLATO – were designed to survey, in nominal terms, about 150,000 (Borucki *et al.* (2010)) and 200,000 (Ricker *et al.* (2014)) targets, respectively.

Table 1 – Summary of PLATO mission stellar samples.

Description	Sample P1	Sample P2	Sample P4	Sample P5
Number of stars	$\geqslant 15,000$	$\geqslant 1,000$	$\geqslant 5,000$	$\geqslant 245,000$
Spectral type	F5 to K7	F5 to K7	M	F5 to late-K
Magnitude in V band	$\lesssim 11$	$\lesssim 8.2$	$\lesssim 16$	$\lesssim 13$
Cadence of image acquisition	25 seconds	2.5 seconds	25 seconds	25 seconds
Data type	images	images	images	light curves (centroids @50 sec for 5% of the targets) (images @25 sec for at least 9,000 targets)
Data sampling	25 seconds	2.5 seconds	25 seconds	600 seconds (50 seconds for 10% of the targets)

Source: ESA (2017).

Within PLATO's mission design, light curves will be produced on board exclusively for the targets of the Sample P5. For all other stellar samples, which are primarily composed of the brightest targets, the photometry will be extracted on-ground from individual images downlinked from the spacecraft, thereby following the same principle as that of *Kepler* and TESS targets.

2.6 Observation strategies and envisaged stellar fields

To achieve the science requirements, the PLATO spacecraft shall be capable of carrying out uninterrupted long duration (few months to several years) photometric stellar observations of a very large sample of (bright) targets at very high photometric precision.

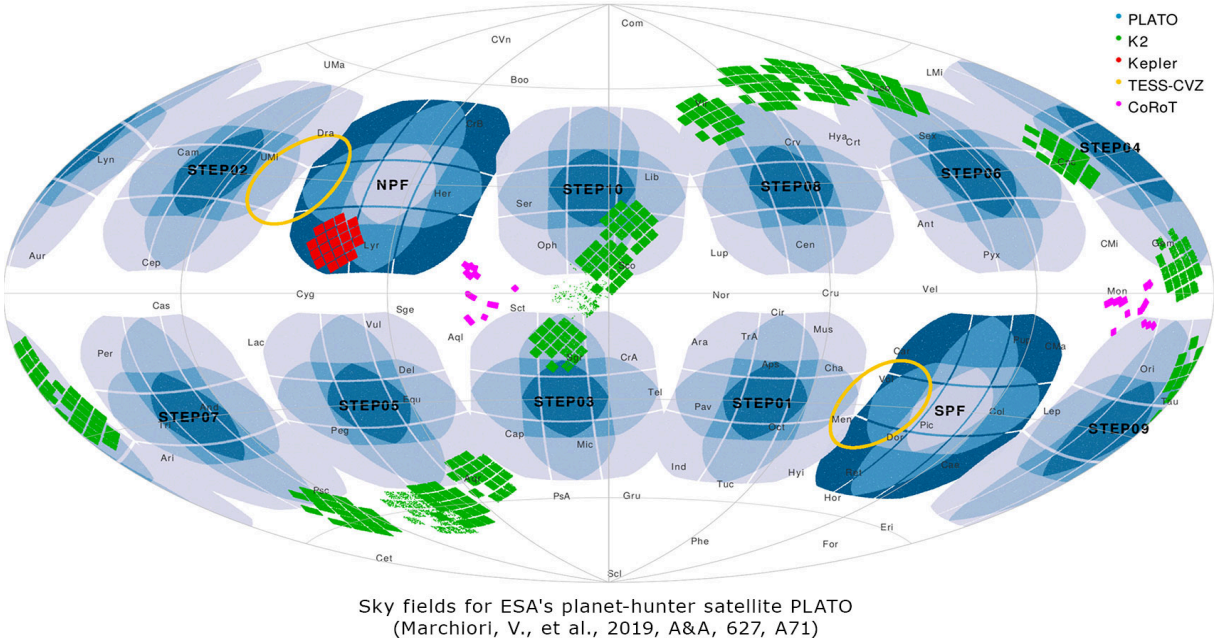
Considering a nominal mission duration of four years, two observation scenarios are considered for PLATO. The first consists of two long-duration (2+2 years) observation phases (LOP) with distinct sky fields. The second consists of a single LOP of three years plus one step-and-stare operation phase (SOP) of one year (i.e. a (3+1 years) observation configuration), covering multiple fields lasting a few months each.

Mission design constraints require the LOP fields to have absolute ecliptic latitude and declination above 63° and 40°, respectively. Under such conditions, two LOP fields are actually envisaged: a southern PLATO field (SPF) centred at Galactic coordinates $l = 253°$ and $b = -30°$ (towards the Pictor constellation) and a northern PLATO field

(NPF) centred at $l = 65°$ and $b = 30°$ (towards the Lyra and Hercules constellations and also including the *Kepler* target field).

An illustration containing the locations of both SPF and NPF is shown in Figure 15, as well as the possible locations of the SOP fields. The definite pointing coordinates of the PLATO target fields will be defined two years before launch.

Figure 15 – Sky coverage in Galactic coordinates of PLATO's provisional SPF and NPF long-duration LOP fields, including the possible locations of the short-duration SOP fields (STEP 01-10). The illustration also shows some sky areas covered by the surveys: *Kepler* (red), *Kepler*-K2 (green), TESS (Continuous Viewing Zones-CVZ; yellow) and CoRoT (magenta).



Source: courtesy of Valerio Nascimbeni (INAF-OAPD, Italy), on behalf of the PLATO Mission Consortium. This image is front page of A&A's Volume 627 (July 2019).

To give an idea of the potential of PLATO observations in the matter of exoplanet detections, with an observation baseline of 2+2 years, the estimated planet yield from stars brighter than $V \sim 13$ is of the order of 4,600 planets (all sizes and orbital periods comprised). This represents more than the total number of discovered exoplanets to date. Yet, if a mission extension of two more years is granted, the total planet yield – with an observation baseline of 3+1+2 years – might achieve expressive $\sim$ 13,000 planets from stars brighter than $V \sim 13$ (ESA (2017)). A mission extension of at least two years is possible considering that the spacecraft is designed for a (nominal) in-orbit operation lifetime of 6.5 years and will carry consumables for 8 years.

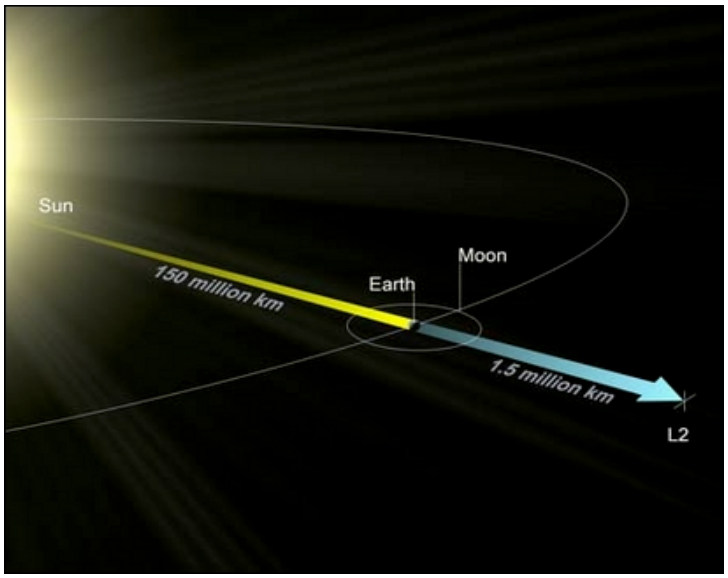
2.7 Launch, orbit and science operations

Uninterrupted long duration and high photometric precision observations as those specified for PLATO requires a sufficiently stable spatial environment. That is, the satellite must be placed into an orbit where it can keep its sight permanently turned to the target field without any viewing obstruction, and preferably with (relative) high thermal stability.

Low-Earth orbits suffer from important flux gradients of energetic particles (e.g. due to the South Atlantic Anomaly (Baglin, Chaintreuil & Vandermarcq (2016), Nasuddin, Abdullah & Hamid (2019)), high levels of scattered Sun-light reflected by the Earth, frequent observation interruptions, and strong thermal variations at timescales of minutes to hours. Hence, the PLATO spacecraft will be placed into a (sufficiently away from the Earth) Lissajous orbit[4] around the L2 Lagrangian point (Figure 16). The L2 point, which is about 1.5 million kilometres (0.01 au) beyond the Earth orbit distance with respect to the Sun, fulfils all the observational needs established for the mission. In L2 orbit, PLATO will be permanently on the line that passes through the Sun and the Earth; as a consequence, the spacecraft will have the exact same orbital period as that of the Earth.

The launch of PLATO satellite is expected to take place by end 2026 from the Kourou base at French Guiana, possibly on a Soyuz-Fregat2-1b launcher. Once in space, the PLATO satellite will be subjected to regular (every 30 days) station-keep manoeuvres to compensate for dynamic orbital instabilities.

Figure 16 – Orbit location (Lagrange L2 point) of PLATO spacecraft.



Source: ESA/NASA.

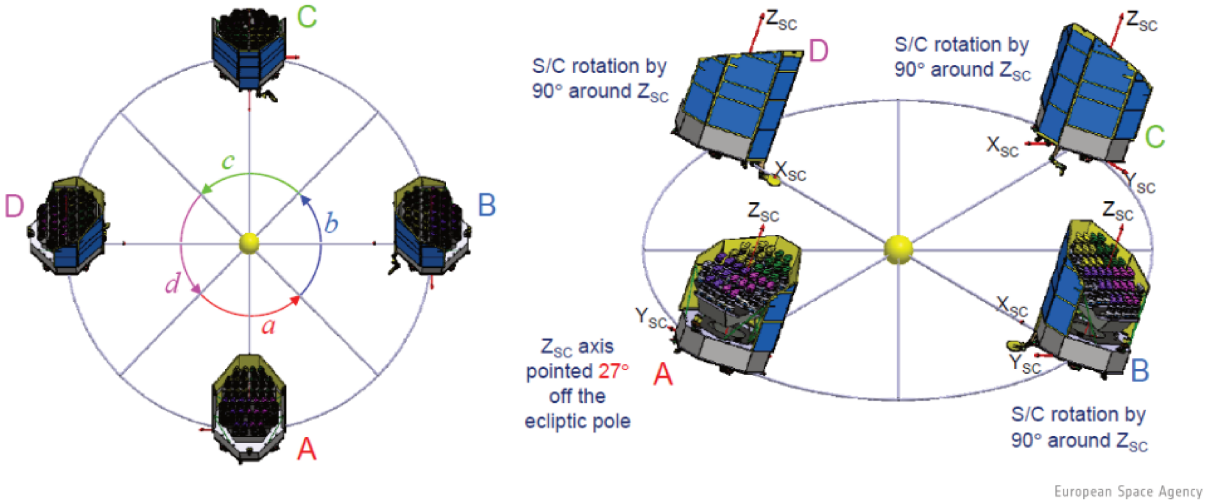
The PLATO satellite will operate according to two distinct operation modes: the

[4] <https://en.wikipedia.org/wiki/Lissajous_orbit>

observation mode and the calibration mode. The observation mode corresponds to the nominal science operation phase during which the instrument is pointing continuously towards a given direction to collect all useful data that are necessary for the mission to cover its core science objectives. The calibration mode, to be triggered every three months, corresponds to the period during which several correction procedures will be carried out to keep both spacecraft and payload in conditions to follow the mission requirements.

For example, during instrument calibration phases the satellite will be rotated by 90° around the payload line of sight to keep the solar panels and shields facing the Sun (Figure 17). Also, images from the whole instrument field of view will be downlinked to the ground and used by the PMC to account for variations across the field of view such as the zodiacal light[5]. Furthermore, changes in the pointing direction of the instrument for switching between different LOP and SOP fields will also be performed during the calibration phases.

Figure 17 – Schematic of spacecraft rotation around payload line of sight to keep solar panels and shields facing the Sun.



Source: ESA.

2.8 Data products

All science data produced by the PLATO mission are categorized into three product levels that can be summarized[6] as follows:

[5] A detailed description of how to quantify the intensities of zodiacal light on PLATO cameras is provided in subsection 3.1 of this document.

[6] A complete description of the science data products is given in PLATO Study Team (2017).

- Level-0: comprises light curves, images and centroid curves of target stars from all individual telescopes, including eventual instrumental corrections applied on board only;

- Level-1: basically consists of Level-0 data that was calibrated and corrected for instrumental errors. These data will be mostly public available after calibration;

- Level-2: comprises planetary candidates along with their respective transit depth, transit duration, and estimated radius; list of planetary systems confirmed with TTVS, to be characterized by combining planetary transits and asteroseismology; results of asteroseismology analysis including stellar rotation periods, masses, radii and ages; and associated uncertainties;

- Level-3: list of confirmed planetary systems to be characterized by combining planetary transits, asteroseismology, and ground-based radial velocity follow-up.

2.9 Instrument description

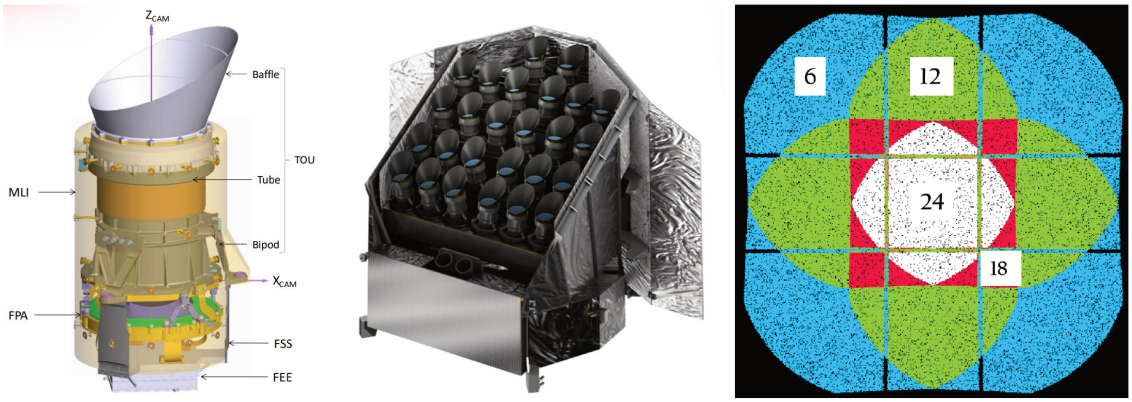
2.9.1 Overall characteristics

The PLATO payload relies on an innovative multi-telescope concept consisting of 26 small aperture (12 cm pupil diameter) and wide circular field of view (∼1,037 deg^2) telescopes mounted in a single optical bench. Each telescope is composed of an optical unit (TOU), a focal plane assembly holding the detectors, and a front-end electronics (FEE) unit. The whole set is divided into 4 groups of 6 telescopes (herein called normal telescopes or N-CAM) dedicated to the core science and 1 group of 2 telescopes (herein called fast telescopes or F-CAM) used as fine guidance sensors by the attitude and orbit control system. The normal telescope assembly results in a overlapped field of view arrangement (see Figure 18), allowing them to cover a total sky extent of about 2,132 deg^2, which represents almost 20 times the active field of the *Kepler* instrument. The N-CAM and F-CAM designs are essentially the same, except for their distinct readout cadence (25 and 2.5 seconds, respectively) and operating mode (full-frame and frame-transfer, respectively). In addition, each of the two F-CAM includes a bandpass filter (one bluish and the other reddish) for measuring stellar flux in two distinct wavelength bands. Table 2 gives an overview of the main payload characteristics based on (ESA, 2017).

2.9.2 Point Spread Function (PSF)

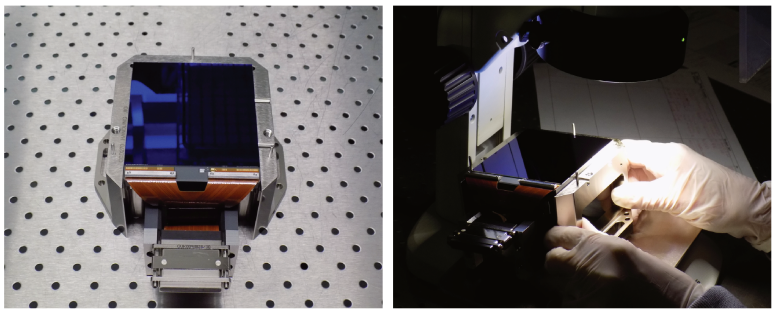
Starlight reaching the focal plane of PLATO cameras will inevitably suffer from distortions caused by both optics and detectors, causing this signal to be non-homogeneously spread out over several pixels. The physical model describing such effects is the PSF, from which one can determine – at subpixel level – how stellar signals are distributed over the

Figure 18 – **Left**: Schematic of one PLATO telescope. **Centre**: Representation of the PLATO spacecraft with 24+2 telescopes. **Right**: Layout of the resulting field of view obtained by grouping the normal telescopes into a 4×6 overlapping configuration. The colour code indicates the number of telescopes covering the corresponding fractional areas (Table 2): 24 (white), 18 (red), 12 (green), and 6 (cyan).



Source: OHB-System AG (**centre**). The PLATO Mission Consortium (**left** and **right**).

Figure 19 – PLATO's Teledyne-e2v 270 CCD.



Source: ESA.

pixels of the detector. This work uses synthetic optical PSF models obtained from the baseline telescope optical layout (Figure 20) simulated on ZEMAX® software. Estimated assembly errors such as lens misalignment and focal plane defocus are included.

Beyond optics, the detectors also degrade the spatial resolution of stellar images through charge disturbances processes such as the charge transfer inefficiency (CTI) (Short *et al.* (2013), Massey *et al.* (2014)), "brighter-fatter" (Guyonnet *et al.* (2015)), and diffusion (Widenhorn (2010)). Several tests are being carried out by ESA to characterize such effects for the charge coupled devices (CCD) of PLATO cameras, so at the present date no formal specifications for the corresponding parameters are available. However, the optical PSFs alone are known to be a non-realistic final representation of the star signals. Therefore, to obtain a first order approximation of the real physics behind the PSF enlargement taking

Table 2 – Summary of main payload characteristics.

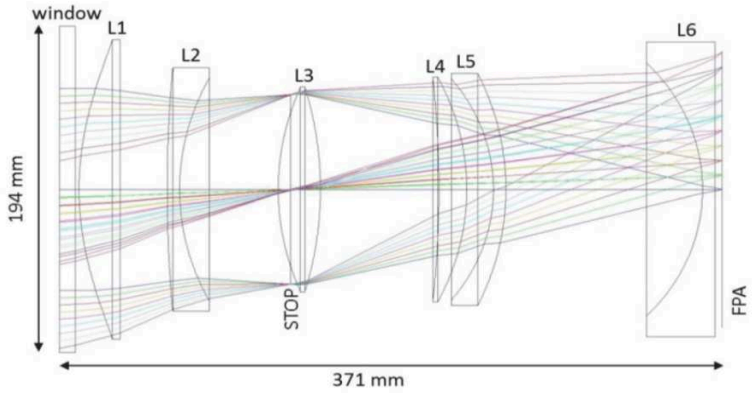
Description	Value
Optics	(24+2) telescopes axisymmetric dioptric design
TOU Spectral range	$500 - 1000\,\mathrm{nm}$
Pupil diameter (per telescope)	$12\,\mathrm{cm}$
Detector	back-illuminated Teledyne-e2v CCD 270 (Figure 19)
N-CAM Focal Plane	4 full-frame CCDs (4510 × 4510 pixels each)
F-CAM Focal Plane	4 frame-transfer CCDs (4510 × 2255 pixels each)
Pixel size	$18\,\mu\mathrm{m}$ square
On-axis plate scale (pixel field of view)	15 arcsec
Quantization noise	$\sim 7.2\,\mathrm{e^-rms\,px^{-1}}$
Readout noise (CCD+FEE)	$\sim 50.2\,\mathrm{e^-rms\,px^{-1}}$
Focal length	24.5 cm
Detector smearing noise	$\sim 45\,\mathrm{e^-\,px^{-1}\,s^{-1}}$
Detector dark current noise	$\sim 4.5\,\mathrm{e^-\,px^{-1}\,s^{-1}}$
N-CAM cadence	25 s
N-CAM exposure time	21 s
N-CAM readout time	4 s
F-CAM cadence	2.5 s
N-CAM field of view	$\sim 1037\,\mathrm{deg^2}$ (circular)
F-CAM field of view	$\sim 619\,\mathrm{deg^2}$
Full field of view	$\sim 2132\,\mathrm{deg^2}$
Fractional field of view	$294\,\mathrm{deg^2}$ (24 telescopes) $171\,\mathrm{deg^2}$ (18 telescopes) $796\,\mathrm{deg^2}$ (12 telescopes) $871\,\mathrm{deg^2}$ (6 telescopes)

Source: ESA (2017).

place at PLATO detectors with respect to the diffusion, the optical PSFs are convolved to a Gaussian kernel with a standard deviation of 0.2 pixel. The resulting simulated PSF models are shown in Figure 21 for 15 angular positions, ξ, within the field of view of one camera. In this work, PSF shape variations due to target colour are assumed to be of second order and are thus ignored.

To reduce the overlap of multiple stellar signals and increase photometric precision, PLATO cameras are primarily designed to ensure that about 77% of the PSF flux is enclosed, on average, within $\sim 2.5 \times 2.5$ pixels across the field of view, or 99% within

Figure 20 – Baseline optical layout of each PLATO telescope.



Source: the PLATO Mission Consortium.

$\sim 5 \times 5$ pixels. As a consequence, the size of the pixels are relatively large compared to that of the PSF, making the distribution of energy from stars very sensitive to their barycentre location within a pixel (see Figure 22).

2.9.3 Spectral response

The spectral response of a photometer represents its efficiency in converting incident photons into effective counts of electrons at detector level. It is derived from the combined effect of optical transmission and CCD quantum efficiency.

The spectral response curve $\mathcal{S}_P(\lambda)$ of PLATO cameras (see Figure 23) is defined as

$$\mathcal{S}_P(\lambda) = \mathcal{O}_P(\lambda) \times \mathcal{Q}_P(\lambda), \tag{2.1}$$

where $\mathcal{O}_P(\lambda)$ is the optical transmission and $\mathcal{Q}_P(\lambda)$ is the quantum efficiency of PLATO cameras. For comparison, Figure 23 also includes the *Gaia* G passband[7], Johnson's V filter from Bessell (1990), Kurucz template (`alpha_lyr_stis_008`) of Vega A0V star from the CALSPEC[8] database, E-490[9] reference solar spectrum from ASTM and a M2V-type star synthetic spectrum from the Pickles atlas[10] (Pickles (1998)).

Numerical values of $\mathcal{O}_P(\lambda)$, $\mathcal{Q}_P(\lambda)$, and $\mathcal{S}_P(\lambda)$ – as they are known to date – are provided in Table 3. In Chapter 3, $\mathcal{S}_P(\lambda)$ is used to derive a PLATO P photometric passband calibrated in the VEGAMAG system (i.e. normalized by the flux of Vega star).

[7] <https://www.cosmos.esa.int/web/gaia/auxiliary-data>
[8] <http://www.stsci.edu/hst/observatory/crds/calspec.html>
[9] <https://www.nrel.gov/grid/data-tools.html>
[10] <http://www.stsci.edu/hst/observatory/crds/pickles_atlas.html>

Table 3 – Optical transmission $\mathcal{O}_P(\lambda)$, detector quantum efficiency $\mathcal{Q}_P(\lambda)$ and resulting spectral response $\mathcal{S}_P(\lambda)$ of PLATO normal cameras at beginning of life, as a function of the wavelength λ.

λ [nm]	500	550	600	650	700	750	800	850	900	950	1000
$\mathcal{O}_P(\lambda)$	0.658	0.711	0.747	0.756	0.771	0.779	0.785	0.788	0.791	0.792	0.794
$\mathcal{Q}_P(\lambda)$	0.7370	0.8070	0.8770	0.8675	0.8580	0.7565	0.6550	0.4845	0.3140	0.1650	0.0620
$\mathcal{S}_P(\lambda)$	0.4849	0.5738	0.6551	0.6558	0.6615	0.5893	0.5142	0.3818	0.2484	0.1307	0.0492

Source: Pertenais, Wohlfeil & Peter (2018).

2.9.4 Vignetting

Another parameter impacting instrument efficiency is vignetting, an inherent optical feature that causes attenuation of image brightness. Such an effect increases non-linearly as the angular position, ξ, of the source augments with respect to the optical axis ($\xi = 0$) of the instrument. The total vignetting, $\mathrm{f_{vig}}$, is the combined result of both natural ($\mathrm{f_{vig}^N}$) and mechanical ($\mathrm{f_{vig}^M}$) vignetting intensities, such that

$$\mathrm{f_{vig}} = 1 - (1 - \mathrm{f_{vig}^N}) \times (1 - \mathrm{f_{vig}^M}). \tag{2.2}$$

For PLATO cameras, the natural vignetting $\mathrm{f_{vig}^N}$ follows a cosine square law resulting in

$$\mathrm{f_{vig}^N} = 1 - \cos^2(\xi), \tag{2.3}$$

whereas the mechanical vignetting $\mathrm{f_{vig}^M}$ is computed numerically from the baseline optical layout. Figure 24 shows the vignetting intensities of PLATO normal cameras as a function of the off-axis angle ξ of the target. The corresponding numerical values are provided in Table 4.

Table 4 – Combined natural and mechanical obscuration vignetting, $\mathrm{f_{vig}}$, as a function of the off-axis angle, ξ, of the target.

ξ [deg]	$\mathrm{f_{vig}}$ [%]	ξ [deg]	$\mathrm{f_{vig}}$ [%]	ξ [deg]	$\mathrm{f_{vig}}$ [%]
0.000	0.00	7.053	1.51	14.001	5.85
1.414	0.06	8.454	2.16	15.370	7.03
2.827	0.24	9.850	2.93	16.730	8.53
4.238	0.55	11.241	3.80	18.081	11.58
5.647	0.97	12.625	4.78	18.887	13.69

Source: author.

2.10 Data processing algorithms

The PLATO data processing pipeline (Figure 25) is a critical component of the payload. This is composed of multiple ground- and flight-based algorithms that are necessary to convert the raw data collected by the instrument, which inevitably carries unwanted systematic disturbances, into scientifically exploitable (Level-1) light curves as outlined in subsection 2.8.

The major sources of systematic errors expected for PLATO are

- Uncertainties on PSF model;
- Long-term star position drift;
- Sky background light;
- Charge disturbance processes occurring in the detectors (CTI, brighter-fatter and diffusion);
- Pixel saturation;
- Light curve outliers;
- Satellite jitter.

Extensive studies have been carried out focussed on the definition of the on-board and on-ground data processing algorithms (Samadi *et al.* (2018), Grolleau *et al.* (2018)) that shall be applied to PLATO images and light curves to correct these for the above sources of systematic errors. In the next few sections, we provide important considerations with regard to some of the effects listed above, in particular to those having a stronger connection to the present work.

2.10.1 PSF modelling

During and after launch, the space environment unavoidably causes overall changes in the instrument response that cannot always be accurately predicted, including variations in the PSF model. Nevertheless, accurate knowledge of the PSFs is imperative for proper correction of systematic errors in the light curves and computing the photometric apertures, so a strategy for reconstructing the PSFs is needed. As the individual raw images downlinked from the spacecraft cannot describe the distribution of stellar flux on the detectors with sufficient resolution, high resolution PSFs such as those in Figure 21 will be reconstructed on the ground from micro-scanning sessions (Samadi *et al.* (2019), Mathé & Samadi (2015)). This process, which will occur every three months during instrument calibration phases, basically consists of collecting (in flight) a series of raw images from subpixel

displacements following an Archimedean spiral (Figure 26). These images are then used on ground by inverse methods to reconstruct the high resolution PSFs.

The efficiency of the inversions is mostly affected by the magnitude of the target star and the off-axis angular position, ξ, of the camera group observing it. The amount of flux from nearby contaminant stars also impacts on the inversion efficacy. However, recent simulations results show that having a sufficiently good knowledge on the positions and magnitude of the stars within the micro-scanning images makes the inversions viable even under the presence of several contaminant stars (Reese & Marchiori (2018)). In that sense, the high precision data from the *Gaia* catalogue will be of major importance.

2.10.2 Long-term star position drift

As explained in subsubsection 2.9.2, the dimensions of PLATO CCD pixels are relatively large compared to the PSF size, making the distribution of energy from stars very sensitive to their barycentre location within a pixel. In parallel, two physical effects are expected to cause apparent star motion of stars on the focal plane: thermo-elastic distortions and aberration of stars.

Thermo-elastic distortions stand for small expansion and contraction displacements of the spacecraft platform that are caused by variations on the temperature gradient across its structure. These mechanical disturbances propagates then to the optical bench where the cameras are installed, thereby causing small but not negligible changes in their positions with respect to the stars in the sky. Overall, thermo-elastic distortions are expected to contribute with apparent stellar position drifts of about 0.5 pixel in three months.

The aberration of stars, also known as differential velocity (kinematic) aberration (DVA), is an astronomical relativistic phenomenon that causes apparent motion of celestial objects about their true positions (see Figure 27). The amplitude of this effect depends on the component of the spacecraft's velocity vector along the line of sight. The greater this velocity component, the smallest the DVA amplitude and vice-versa. For PLATO stars, the DVA amplitude is expected to be up to ~ 8 arcsec (~ 0.5 pixel) in three months.

Considering both thermo-elastic distortions and DVA effects combined, stars in PLATO field are therefore expected to manifest apparent position motions of the order of ~ 1.3 pixel in three months in worst case scenarios.

2.10.3 Sky background light

In stellar photometry, fluxes originated from sources other than the stars of interest (target stars) are considered backgrounds. Background fluxes can be distinguished between those arising from celestial sources (e.g. zodiacal light and contaminant stars) from those

originated at the instrument level (e.g. electronic gain and internal optical light reflections called “ghosts”).

High-precision-stellar-photometry-based missions, like PLATO, are significantly affected by backgrounds in basically two ways: on one hand, background random noise increases the noise-to-signal ratio of the light curves; and on the other hand, background average flux dilutes astrophysical transit depths, making these appear in the light curves smaller than they naturally are.

With regards to the PLATO light curves, since the satellite will be positioned in L2 orbit, the major celestial contributors to background flux will be the zodiacal light and the flux from contaminant stars.

2.10.4 Pixel saturation

The total number of electrons that a CCD pixel can accommodate is finite. Beyond that limit, known as the pixel full well capacity (FWC), any additional charge falling into that pixel will inevitably be spilled out into adjacent pixels along the same column.

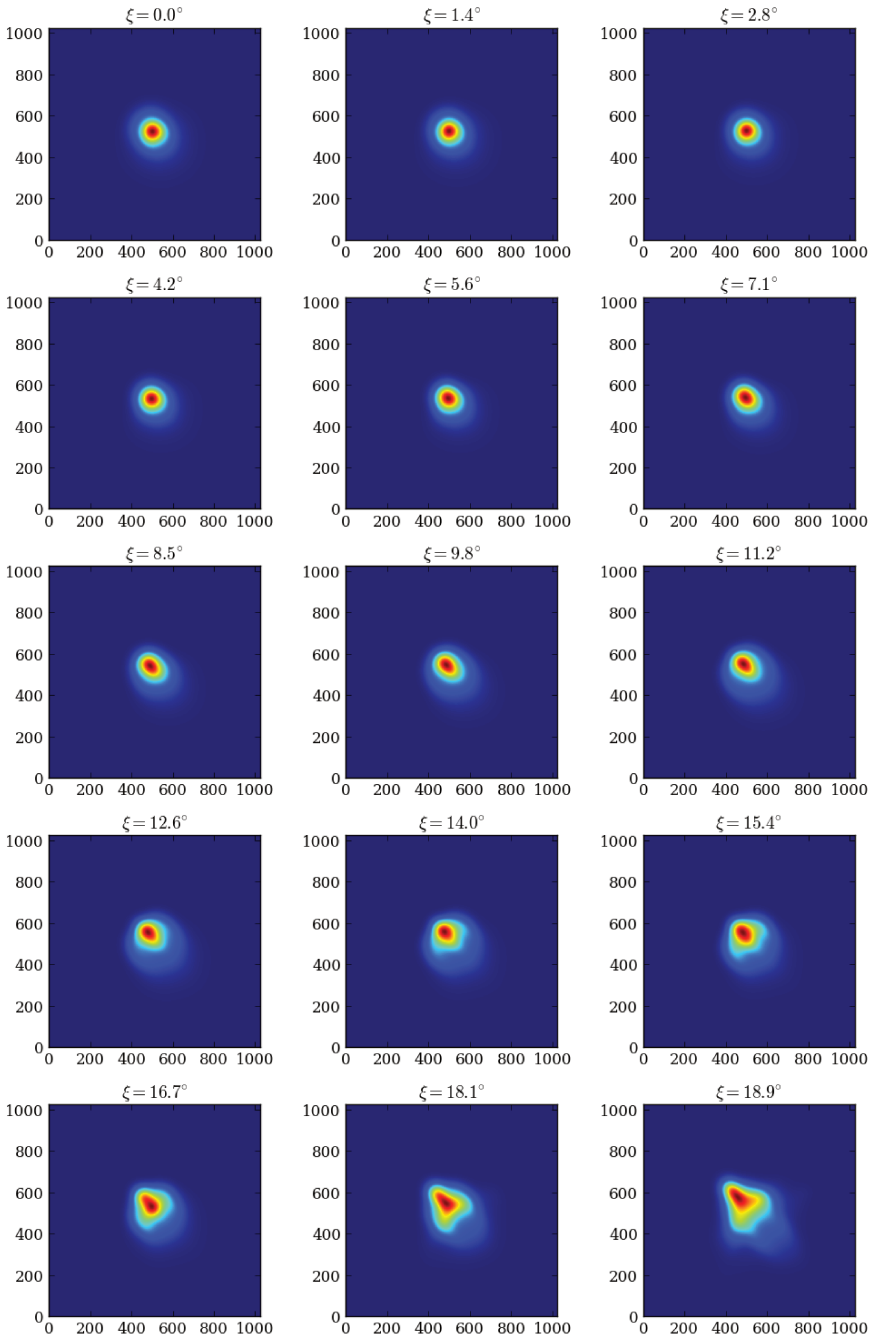
Stars that are bright enough to saturate one or more pixels are called “saturated” stars. Extracting photometry from such stars requires a different pixel selection strategy with respect to that applied to non-saturated stars. In that case, the typical solution consists of collecting additional flux from an extended number of pixels along the columns, depending on the level of saturation.

In the current instrument design, PLATO detectors are expected to exhibit saturation at pixels observing stars brighter than $V \sim 8.5$ beyond 25s exposure on normal cameras. The exact saturation limit however depends on the location of the star in the CCD and where its barycentre falls within a pixel.

2.10.5 Photometry extraction

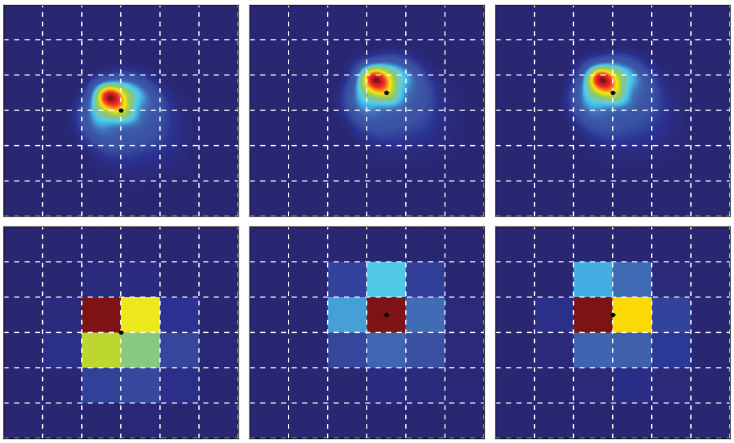
As pointed out in subsection 2.5, light curves will be produced on board exclusively for the targets of the Sample P5. In such context, this thesis presents the research work carried out – within the data processing algorithm working group – to define optimal apertures for extracting photometry of P5 targets. Within the WBS structure of the data processing algorithms working group (Figure 25), this thesis represents the work package with code WP 323 100 (Mask-based photometry).

Figure 21 – Simulated PSF shapes of PLATO telescopes (1/128 pixel resolution) as a function of the angular position ξ in the sky of a source at $-45°$ azimuth. Angular positions range from $\xi = 0°$ (centre) to $\xi = 18.9°$ (edge) of camera field-of view. Each optical PSF is convolved to a Gaussian diffusion kernel with standard deviation of 0.2 pixel to simulate the CCD behaviour. Each image above correspond to a CCD surface of 8×8 pixels, which is enough to encompass $\sim 99.99\%$ of the total PSF energy.



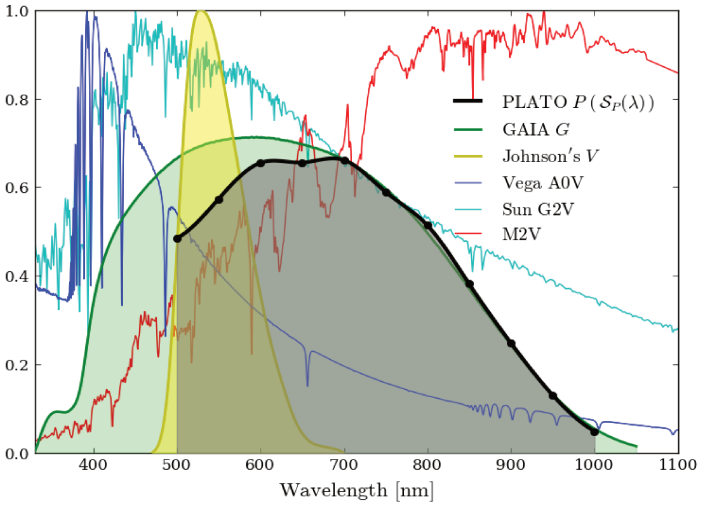
Source: Martin Pertenais from the DLR, on behalf of the PLATO Mission Consortium.

Figure 22 – Energy distribution of PSF across the pixels for three distinct intra-pixel target barycentre locations (black dots): at pixel corner (left), at pixel centre (middle), and at the border of two adjacent pixels (right). Dashed white lines represent pixel borders. **Top**: High resolution PSF at $\xi = 14°$. **Bottom**: Corresponding low resolution PSF.



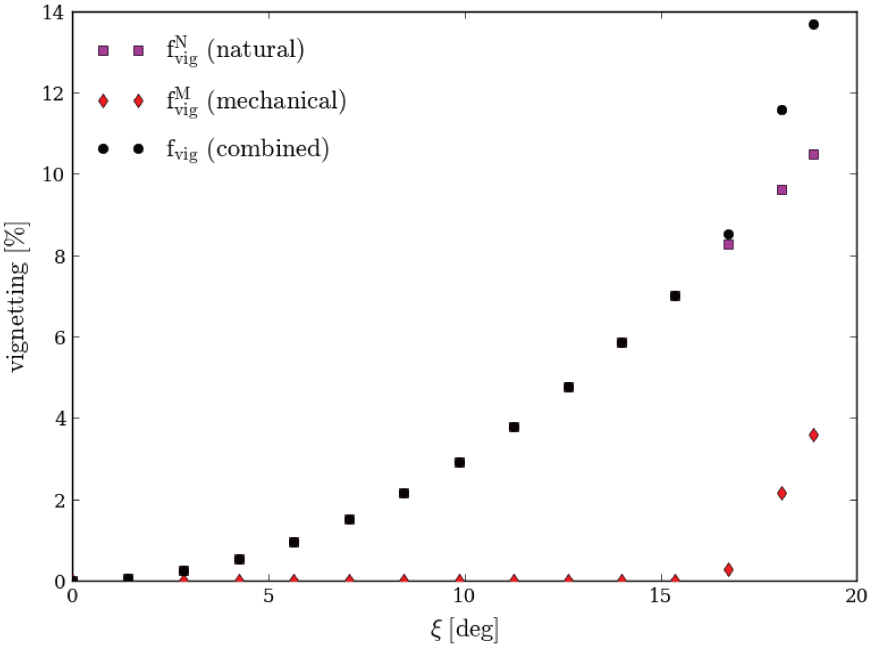
Source: author.

Figure 23 – **Black**: Preliminary spectral response $\mathcal{S}_P(\lambda)$ of PLATO N-CAM at beginning of life. Values are currently known at the black dots c; these are crossed by a cubic spline interpolation curve. **Green**: *Gaia G* band. **Yellow**: Johnson's *V* filter. **Blue**: Vega (A0V) normalized spectrum. **Cyan**: Sun (G2V) normalized spectrum. **Red**: Normalized spectrum of a M2V-type star.



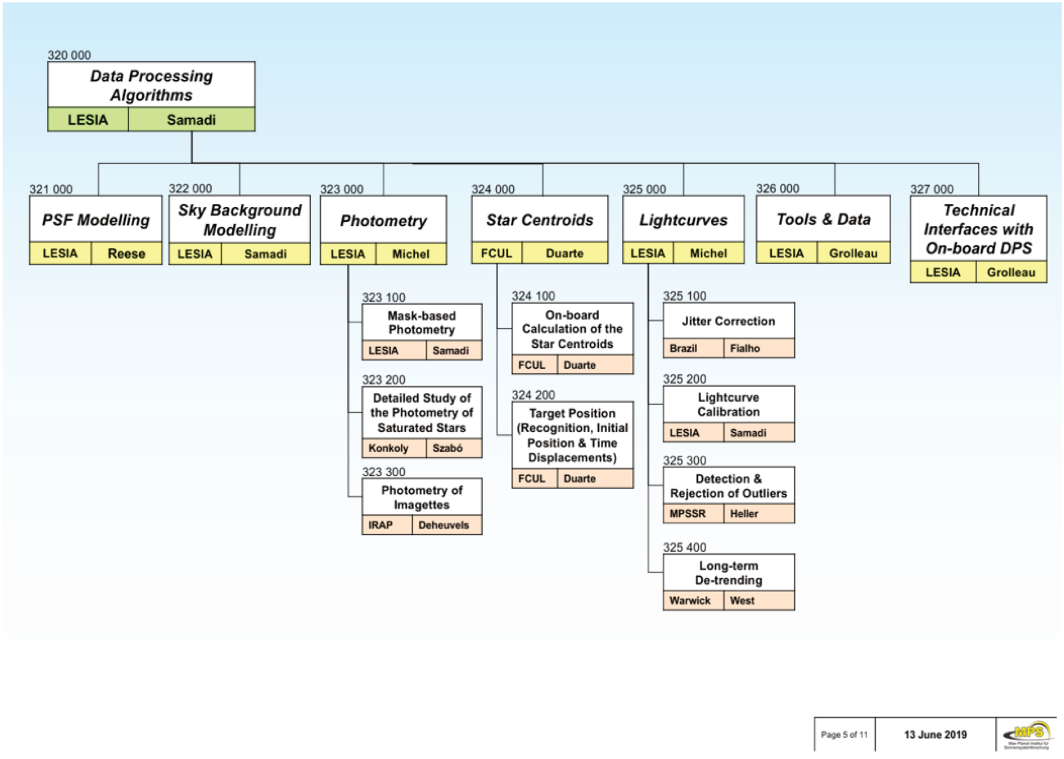
Source: author.

Figure 24 – Combined natural and mechanical obscuration vignetting, f_{vig}, as a function of the off-axis angle, ξ, of the target.



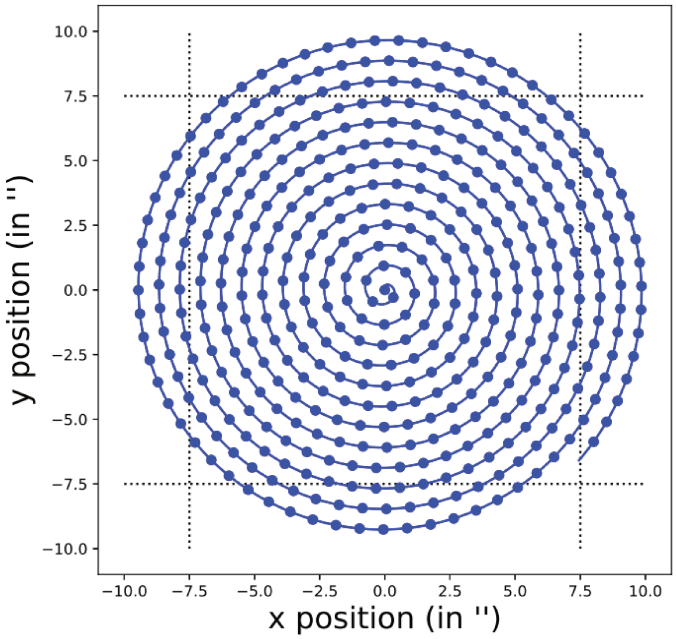
Source: author.

Figure 25 – Work breakdown structure of Data Processing Algorithms.



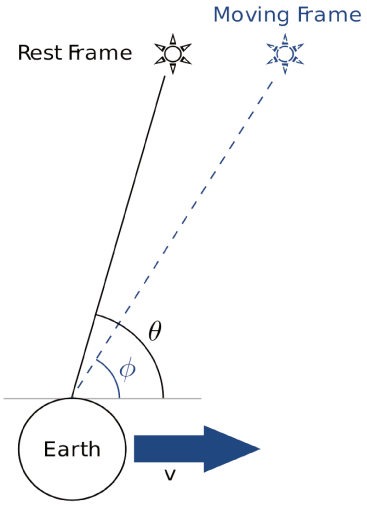
Source: the Max Planck institute for Solar system research (MPS).

Figure 26 – Archimedean spiral sub-pixel depointing trajectory used for micro-scanning strategy. The dotted lines correspond to the CCD pixel borders.



Source: Samadi *et al.* (2019).

Figure 27 – .



Source: <https://en.wikipedia.org/wiki/Aberration_of_light>.

3 Image formation at PLATO detectors

During science observations, in-flight photometry extraction will be performed independently for each target by integrating its flux over a set of selected pixels (aperture or mask). Such pixel collection is to be chosen from a 6×6 pixels square window called imagette, assigned uniquely to each target. Characterizing the expected performance of mask-based photometry requires building up such imagettes (i.e. realistic models of $I_{i,j}$; Equation 1.3), which shall be composed of realistic stellar content (targets and respective contaminants). This chapter presents in detail all the elements considered for constructing the imagettes that are used as input for the simulations of aperture photometry presented in Chapter 4.

The work presented in this chapter is strongly based on Marchiori *et al.* (2019). In-flight photometry extraction of PLATO targets: Optimal apertures for detecting extra-solar planets, A&A, 627, A71.

3.1 Zodiacal light

As the spacecraft will be positioned in L2 orbit (located at approximately 1.01 au from the Sun), sky background flux entering its cameras will be dominated by the zodiacal light, i.e. sunlight scattered by interplanetary dust particles agglomerated across the ecliptic plane. Zodiacal light brightness is conventionally expressed in counts of 10th visual magnitude solar-type stars per square degree, also known as S_{10} unit.

By denoting $\mathcal{I}_{\odot}(\lambda)$ as the solar spectral irradiance at 1AU and adopting a corresponding apparent visual magnitude $V_{\mathrm{Sun}} = -26.74$ mag, the S_{10} unit is formally defined[1] as

$$S_{10} = 10^{-0.4(10-V_{\mathrm{Sun}})}\,\mathcal{I}_{\odot}(\lambda)\ \mathrm{deg}^{-2} = 6.61 \times 10^{-12}\,\mathcal{I}_{\odot}(\lambda)\ \mathrm{sr}^{-1}. \tag{3.1}$$

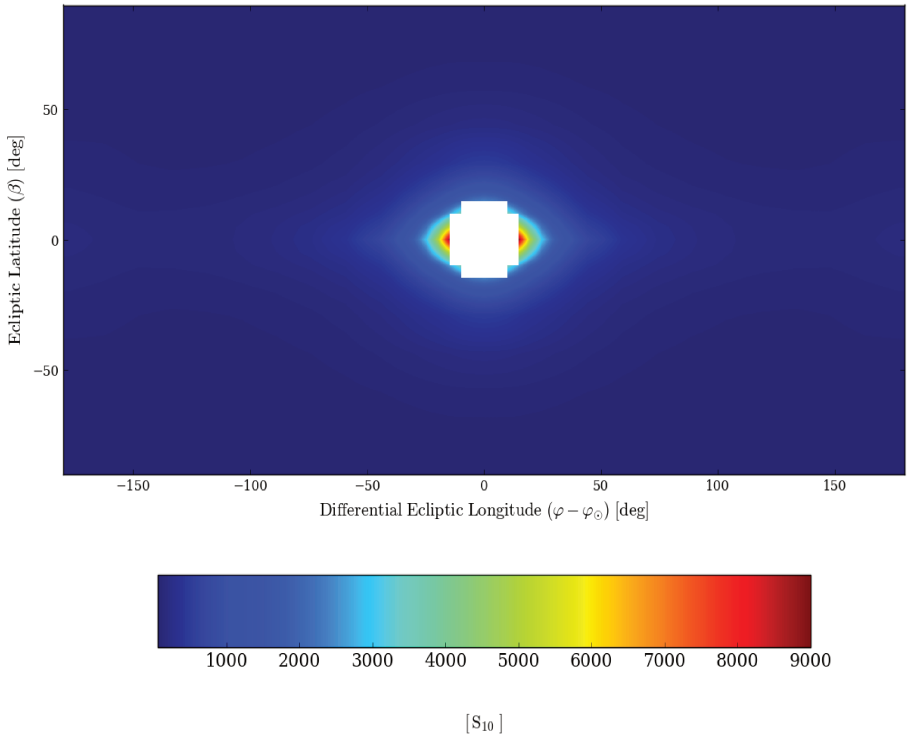
Tabular data containing zodiacal light measurements in S_{10} units are available in Leinert *et al.* (1998). The published data (see visual representation in Figure 28) are given as a function of viewing directions specified in ecliptic coordinates with zero point in the Sun (Figure 29). The corresponding values are valid for an observer located in the vicinity of Earth and at monochromatic wavelength (500 nm). Outside these conditions, a semi-analytical model containing a few correction factors shall be applied. Based on that model,

[1] The steradian (sr), also called square radian, is a unit of solid angle in the International System of Units (SI). Accordingly, 1 sr is equal to $(180/\pi)^2\,\mathrm{deg}^2$.

we have built an expression (see Table 5 for parameters description) for estimating the amount of zodiacal light flux f_P^{ZL} on one N-CAM

$$f_P^{\mathrm{ZL}} = f_{\mathrm{ZL}} \left(6.61 \times 10^{-12}\ \mathrm{sr}^{-1}\right) (h\,c)^{-1}\, \Omega\, \Theta_P\, \mathrm{f_{L2}}\, (1 - \mathrm{f_{vig}}) \int_{\lambda_1}^{\lambda_2} \mathcal{I}_{\odot}(\lambda)\, \mathrm{f_{red}}(\lambda)\, \mathcal{S}_P(\lambda)\, \lambda\, d\lambda. \tag{3.2}$$

Figure 28 – Zodiacal light model on the celestial sphere.



Source: author, with data extracted from Leinert *et al.* (1998).

In the above expression, $\mathcal{I}_{\odot}(\lambda)$ is modelled with ASTM's E-490[2] reference solar spectrum (Figure 23) within the spectral range $[\lambda_1, \lambda_2] = [500, 1000]$nm of PLATO cameras, resulting in (expressed in units of $\mathrm{e^-\,px^{-1}\,s^{-1}}$)

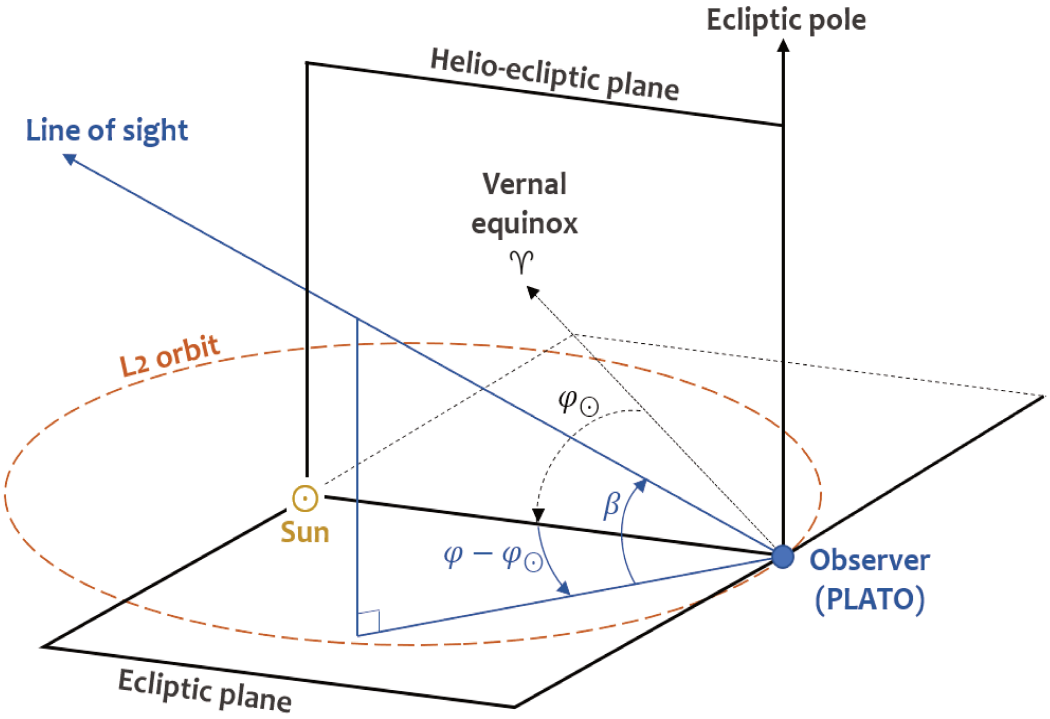
$$f_P^{\mathrm{ZL}} \sim 0.39\, f_{\mathrm{ZL}}\, (1 - \mathrm{f_{vig}})\,. \tag{3.3}$$

Excluding vignetting ($f_{\mathrm{vig}} = 0$), $1\,\mathrm{S_{10}}$ of zodiacal light ($f_{\mathrm{ZL}} = 1$) corresponds thus to about $0.39\ \mathrm{e^-\,px^{-1}\,s^{-1}}$ being generated at the detectors of one PLATO camera.

For illustration, Figure 30 shows a 360-degree zodiacal light panorama[3] taken from Mauna Kea in 2011. In this photograph, the *gegenschein*[4] (radiation backscattering)

[2] <https://www.nrel.gov/grid/data-tools.html>
[3] <http://www.zam.fme.vutbr.cz/~druck/Astro/Mauna_kea_11/0-info.htm>
[4] <https://www.swisseduc.ch/stromboli/volcano/photoastro/gegenschein-en.html>

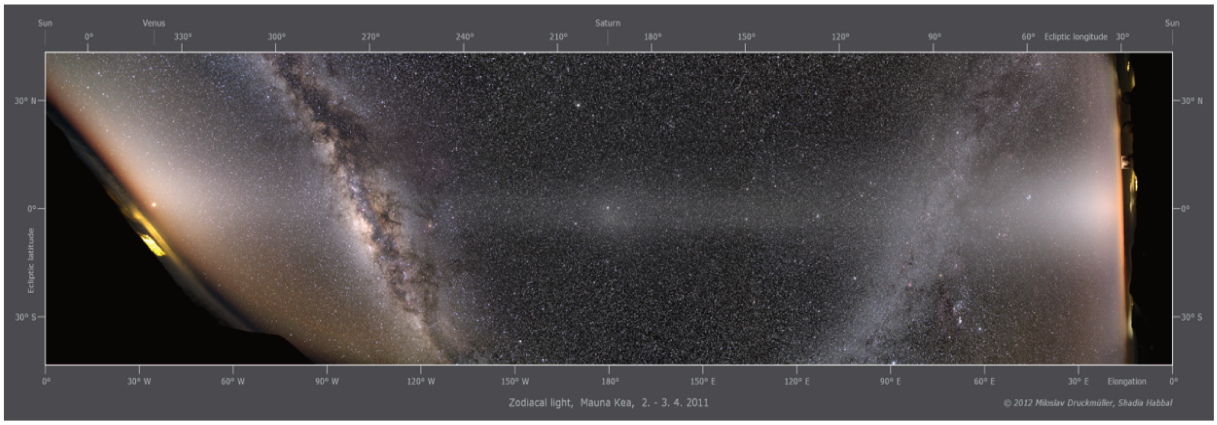
Figure 29 – Ecliptic coordinates with zero point in the Sun.



Source: author, adapted from Grün *et al.* (2001) and Molina & dallAmico (2016).

effect in zodiacal light has elongation of 180°, which coincides with $\varphi_{\text{LoS}} - \varphi_\odot = \pm 180°$ in Figure 28.

Figure 30 – 360-degree zodiacal light panorama taken from Mauna Kea, Hawaii.



Source: Miloslav Druckmüller, Shadia Habbal.

Table 5 – Description of the parameters of Equation 3.2.

Description	Symbol	Value	Unit
Zodiacal light tabulated data	f_{ZL}	see Leinert *et al.* (1998)	S_{10}
Planck's constant	h	6.63×10^{-34}	J s
Speed of light in vacuum	c	2.99×10^{8}	$\mathrm{m\,s^{-1}}$
Field-of-view solid angle (per pixel)	Ω	4.2×10^{-9}	sr
Entrance pupil surface (per camera)	Θ_P	113.1	$\mathrm{cm^2}$
Spectral range	$[\lambda_1, \lambda_2]$	$[500, 1000]$	nm
Sun's spectral irradiance	$\mathcal{I}_{\odot}(\lambda)$	ASTM's E-490 spectrum	$\mathrm{W/cm^2/nm}$
Correction factor for L2 point	$\mathrm{f_{L2}}$	0.975	adim
Instrument vignetting	$\mathrm{f_{vig}}$	see Table 4	adim
Redenning correction factor	$\mathrm{f_{red}}(\lambda)$	see Leinert *et al.* (1998)	adim
Spectral response	$\mathcal{S}_P(\lambda)$	see Figure 23	adim

Source: author.

3.2 The input stellar catalogue

Input stellar catalogues are essential tools for space missions dedicated to asteroseismology and exoplanet searches. Besides its crucial role in field and target selection, it is also noticeably useful for estimating and characterizing the performance of photometry extraction methods prior to mission launch. Indeed, an input catalogue allows one to produce synthetic sky images containing realistic stellar distributions, including their relative positions, apparent magnitudes, effective temperatures, gravities, metallicities and more. At the present date, a PLATO Input Catalogue (PIC) is being developed based on the ultra-high precision astrometric data from the *Gaia* mission (Gaia Collaboration *et al.*, 2016). In the future, the PIC might also include information available from other sky surveys such as the Large Synoptic Survey Telescope (LSST) (IVEZIĆ *et al.*, 2019), Panoramic Survey Telescope and Rapid Response System (PanSTARSS) (Chambers *et al.*, 2016), and SkyMapper Wolf *et al.* (2018). The PIC will provide abundant and detailed stellar information for optimized target selection vis-à-vis mission science goals. As the PIC was not yet available by the time that the present work was started, we have adopted the *Gaia* DR2 (Gaia Collaboration *et al.*, 2018b) as input catalogue, which provides all the information needed for the present work.

The development of this thesis is fully based on the *Gaia* stars that are located in the fraction of SPF (see Figure 15) that is equivalent to the area covered by a single PLATO camera ($\sim 1037\,\mathrm{deg^2}$) centred at SPF centre. This stellar field is denoted hereafter as the input field (IF). That represents roughly half of the SPF area in the sky and encompasses about 12.8 million stars listed in the *Gaia* DR2 catalogue with G magnitude comprised between 2.45 and 21. Table 6 presents, in different reference systems, the coordinates of

IF centre, hereafter referred to as IF line of sight (IF_{LoS}). Figure 31 illustrates the sky area covered by IF in ecliptic coordinates with zero point $\varphi_\odot$ in the Sun. Three cases are represented: φ_{LoS} shifted by $-110°$ with respect to $\varphi_\odot$; φ_{LoS} aligned with $\varphi_\odot$; and φ_{LoS} shifted by $+110°$ with respect to $\varphi_\odot$. It is also indicated in the same figure, i the zodiacal light intensities – as described in subsection 3.1 – perceived by an observer located in L2 orbit. Figure 32 shows the expected temporal variations of zodiacal light during a one-year period, at three different latitudes within SPF: (1) at the latitude closest to the ecliptic plane ($\beta \sim -46°$), i.e. where the zodiacal intensities are expected to reach their maximum in the field; (2) at IF_{LoS} latitude ($\beta \sim -70°$), i.e. around the centre of the field; and (3) at the south ecliptic pole ($\beta = -90°$), i.e. where the zodiacal intensities are expected to be invariant and near minimum in the field. The plots show that, for stars located on the lowest absolute latitudes within SPF, the temporal variation of zodiacal light might be up to $\sim 160\%$ peak to peak over three months. Around IF_{LoS}, this variation might reach up to $\sim 40\%$ peak to peak over the same time scale.

Table 6 – Coordinates of the input field line of sight (IF_{LoS}) in different reference systems.

Reference System	Longitude	Latitude
Galactic	$l_{LoS} = 253°$	$b_{LoS} = -30°$
Equatorial	$\alpha_{LoS} = 86.80°$	$\delta_{LoS} = -46.40°$
Ecliptic	$\varphi_{LoS} = 83.62°$	$\beta_{LoS} = -69.77°$

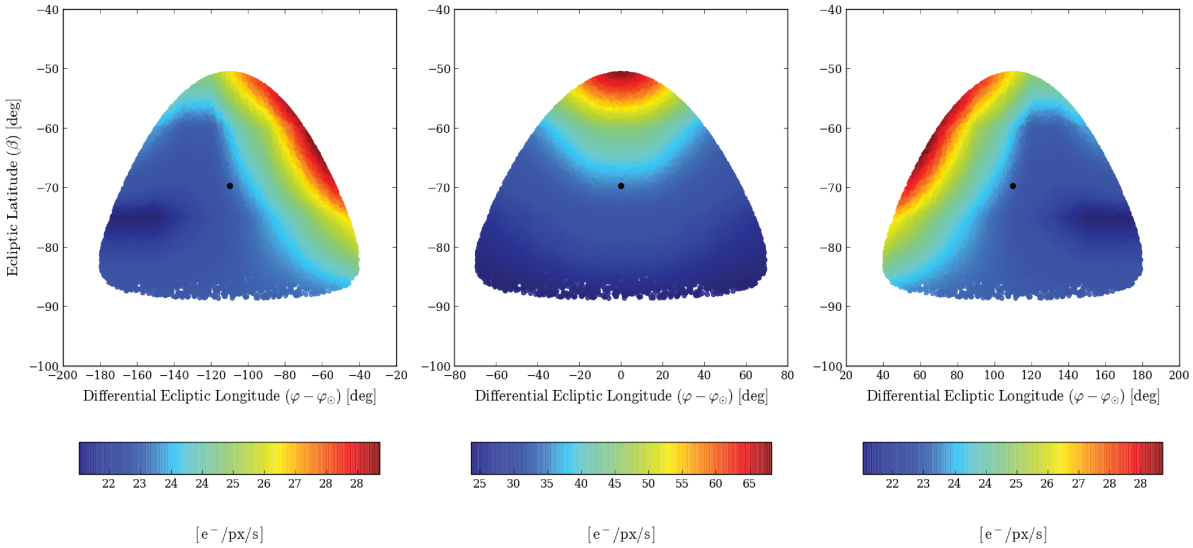
Source: author.

3.3 Determining stellar positions on the focal plane

Accurately evaluating photometric performance requires properly determining stellar positions at detector (pixel) level. In this subsection, mathematical expressions are provided for transforming star positions from celestial (equatorial) coordinates to Cartesian coordinates at the focal plane of one PLATO camera.

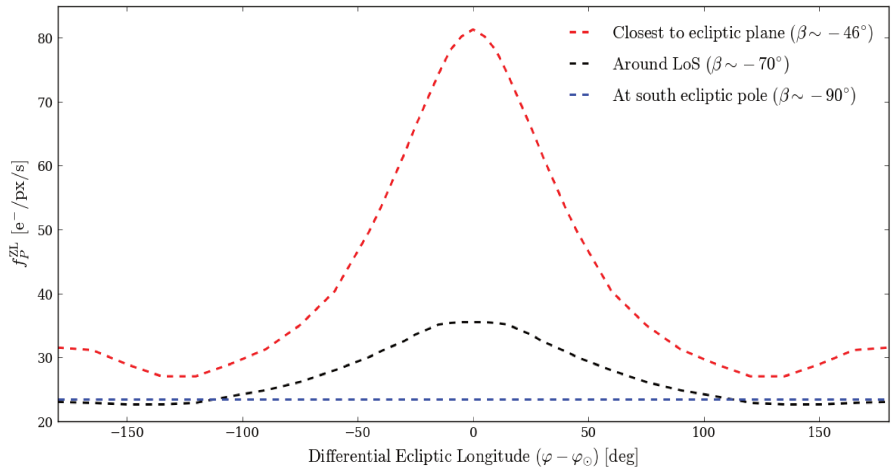
We define herein an arbitrary camera reference frame (see Figure 33) with angular coordinates (ξ, ζ), where ξ is the off-axis angle, i.e. the angular position of the star in the sky (see Figure 21) with respect to the camera optical axis, and ζ is the azimuthal angle. In this reference frame, the corresponding Cartesian coordinates with axes (x, y, z) are such that the z-axis is superimposed to the camera optical axis, and both x and y axes have origin in the centre of the focal plane, i.e. in the centre of the total surface covered by the (2×2) CCD array. The azimuthal angle ζ is measured in the counter-clockwise sense from the x-axis. A star with nominal position (ξ_0, ζ_0) in the camera's reference frame

Figure 31 – Scatter plot of zodiacal light across IF, with $\mathrm{IF}_{\mathrm{LoS}}$ indicated by black dots. Instrument vignetting is not included ($f_{vig} = 0$). **Left**: $\varphi_{\mathrm{LoS}} - \varphi_{\odot} = -110°$. **Centre**: $\varphi_{\odot} = \varphi_{LoS}$. **Right**: $\varphi_{\mathrm{LoS}} - \varphi_{\odot} = +110°$.



Source: author.

Figure 32 – Expected temporal variation of zodiacal light at three different latitudes within SPF over a 1-year period.

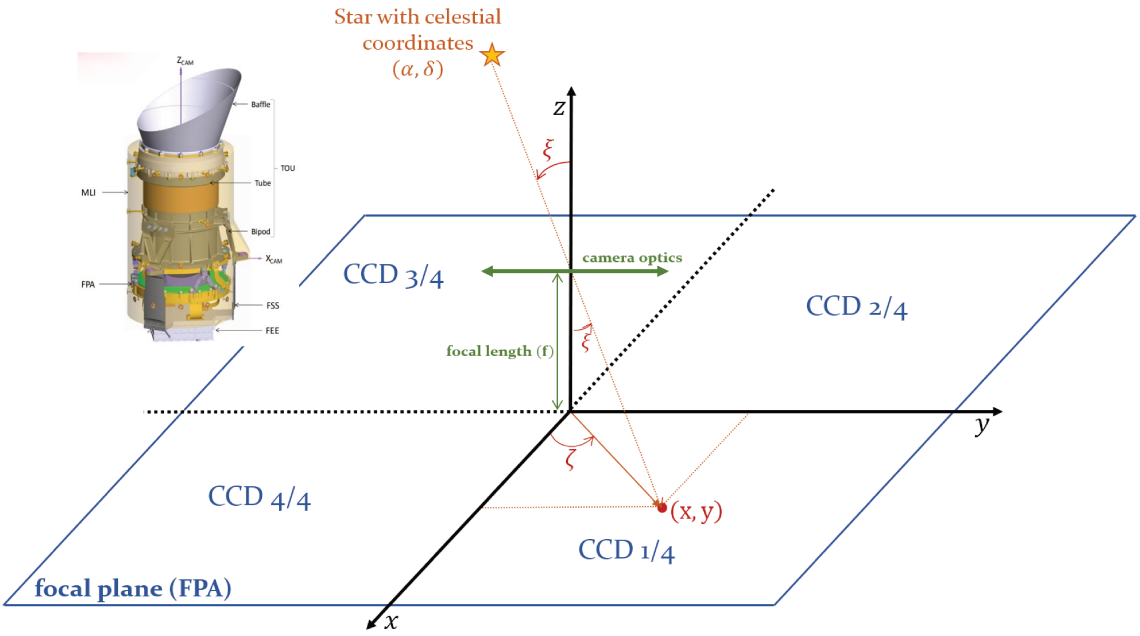


Source: author.

has a corresponding nominal position $(\mathbf{x}_0, \mathbf{y}_0)$ in the focal plane that can be determined through a gnomonic projection of the form

$$\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix} = \left(\frac{10^4\,\mu\mathrm{m/cm}}{18\,\mu\mathrm{m/px}} \right) \times \begin{bmatrix} \mathsf{f}\,\tan(\xi_0)\cos\zeta_0 \\ \mathsf{f}\,\tan(\xi_0)\sin\zeta_0 \end{bmatrix}, \tag{3.4}$$

Figure 33 – Arbitrary camera reference frame (z-axis ≡ camera optical axis).



Source: author.

where $\mathsf{f} = 24.5$ cm is the focal length (Table 2) of the PLATO cameras and $\mathbf{x}$ and $\mathbf{y}$ are given in units of pixel. A star's nominal position (ξ_0, ζ_0) in the camera reference frame can be determined from its position in celestial (equatorial) coordinates (α_0, δ_0) using

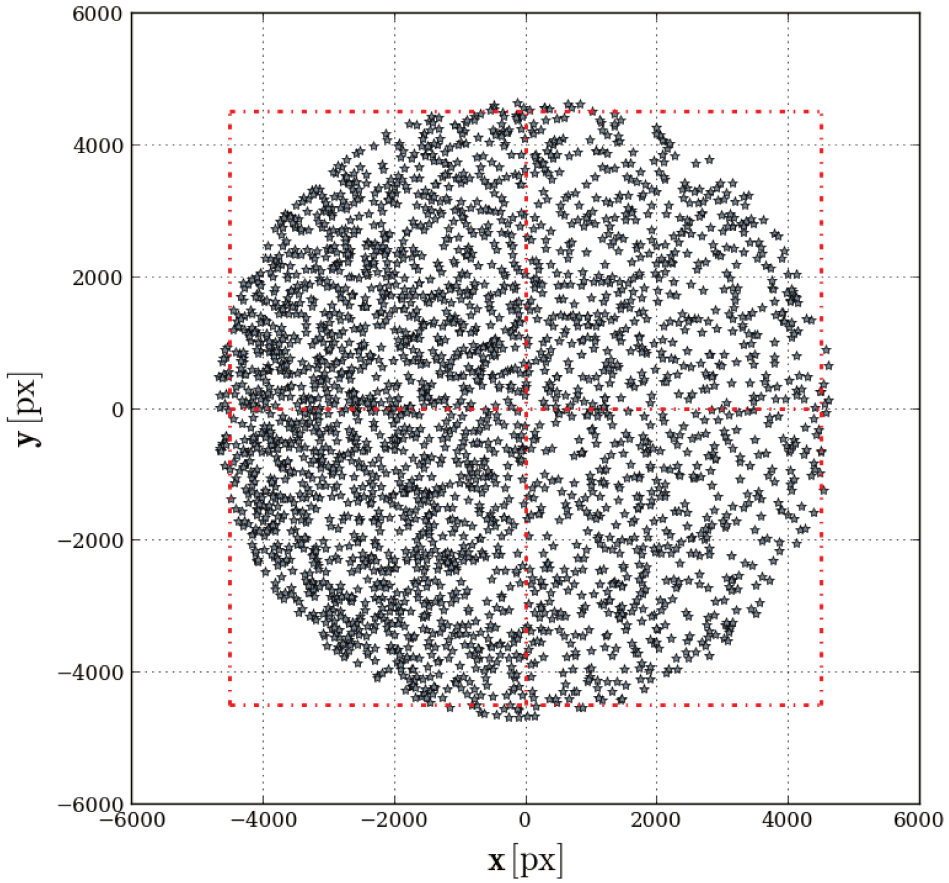
$$\begin{bmatrix} \xi_0 \\ \zeta_0 \end{bmatrix} = \begin{bmatrix} \pi/2 - \operatorname{asin}\left(\sin(\delta_0)\cos(\delta_{\mathrm{LoS}} - \pi/2) - \cos(\delta_0)\sin(\delta_{\mathrm{LoS}} - \pi/2)\sin(\alpha_0 - \alpha_{\mathrm{LoS}} + \pi/2)\right) \\ \operatorname{atan2}\left(\dfrac{\cos(\delta_0)\cos(\delta_{\mathrm{LoS}} - \pi/2)\sin(\alpha_0 - \alpha_{\mathrm{LoS}} + \pi/2) + \sin(\delta_0)\sin(\delta_{\mathrm{LoS}} - \pi/2)}{\cos(\delta_0)\cos(\alpha_0 - \alpha_{\mathrm{LoS}} + \pi/2)}\right) \end{bmatrix}. \tag{3.5}$$

We note that above transformation assumes that the z-axis is aligned with the direction of $\mathrm{IF}_{\mathrm{LoS}}$. This is consistent with the fact that the adopted input field in this work is equivalent to the sky area covered by a single camera centred at SPF centre, as explained in the previous section. Hence, switching stellar positions between (α, δ) to $(\mathbf{x}, \mathbf{y})$ coordinates on the focal plane can be broken into two steps: one for applying the transformation $(\alpha, \delta) \rightarrow (\xi, \zeta)$ (Equation 3.5), and another one for applying the transformation $(\xi, \zeta) \rightarrow (\mathbf{x}, \mathbf{y})$ (Equation 3.4). Accordingly, Figure 34 is obtained by applying this procedure to the stars located within the input field IF. In this figure, readers might notice that the camera field of view slightly overrun the CCD array area at its edges, while the areas towards its corners are unexposed to light.

3.4 Satellite jitter

High spacecraft pointing stability is key for achieving low photometric noise. In contrast, although the attitude and orbit control systems of spacecrafts typically deliver

Figure 34 – A few stars randomly chosen from the sources located within IF, with respective positions projected onto the focal plane of a single PLATO camera. Red dashed lines represent the borders of the (2×2) CCD array. The increased stellar density leftwards is natural and due to higher proximity to the galactic plane.

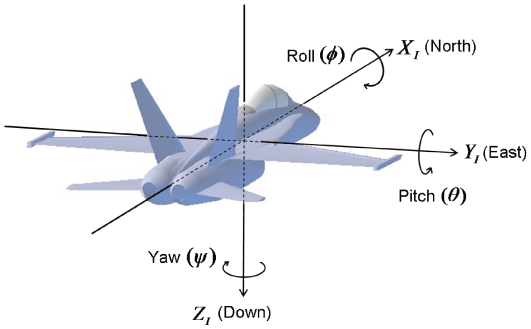


Source: auhtor.

average null pointing errors, some non-null residual errors always exist. Such an unwanted satellite motion is also known as jitter, which because of fixed aperture photometry method (subsubsection 1.5.2) inevitably propagates to the photometry, thereby degrading its noise performance. Hence, to estimate the amount of photometric noise induced by satellite jitter in the light curves, we describe in this section the calculation steps for quantifying the pointing error amplitudes at the focal plane (i.e. CCD) coordinates $\mathbf{x}$ and $\mathbf{y}$, from the pointing error signal given in the reference frame of the Fine Guidance System (FGS).

Pointing errors are typically specified in terms of the three Euler angles (ϕ, θ, ψ) (see Figure 35), which represent rotations about spacecraft's Cartesian axes x (roll), y (pitch) and z (yaw), respectively. Following that representation of orientation in space,

Figure 35 – Representation of the Euler angles (ϕ, θ, ψ) in the body reference frame.



Source: CH Robotics (<http://www.chrobotics.com/library>).

Figure 36 shows a jitter time-series realization which is representative of the current status of PLATO spacecraft's pointing requirements. In the reference frame of one PLATO camera (Figure 33), the nominal angular position (ξ_0, ζ_0) of a star can be translated into a corresponding nominal position (x_0, y_0, z_0) in Cartesian coordinates with

$$\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = \begin{bmatrix} \sin(\xi_0)\cos\zeta_0 \\ \sin(\xi_0)\sin\zeta_0 \\ \cos(\xi_0) \end{bmatrix}. \tag{3.6}$$

In the above expression, we adopt for simplicity a canonic three dimensional description of the Cartesian coordinates (x, y, z), that is (x, y, z) are points located within a unit radius sphere.

Because of satellite jitter, the nominal position (x_0, y_0, z_0) of the star deviates, at each instant k, to a new position (x_k, y_k, z_k) which depends on the corresponding amplitudes of the Euler angles $(\phi_k, \theta_k, \psi_k)$ (see schematic of Figure 37). Each new star position (x_k, y_k, z_k) can be determined through its respective rotation matrix[5] $R(\phi_k, \theta_k, \psi_k)$ doing

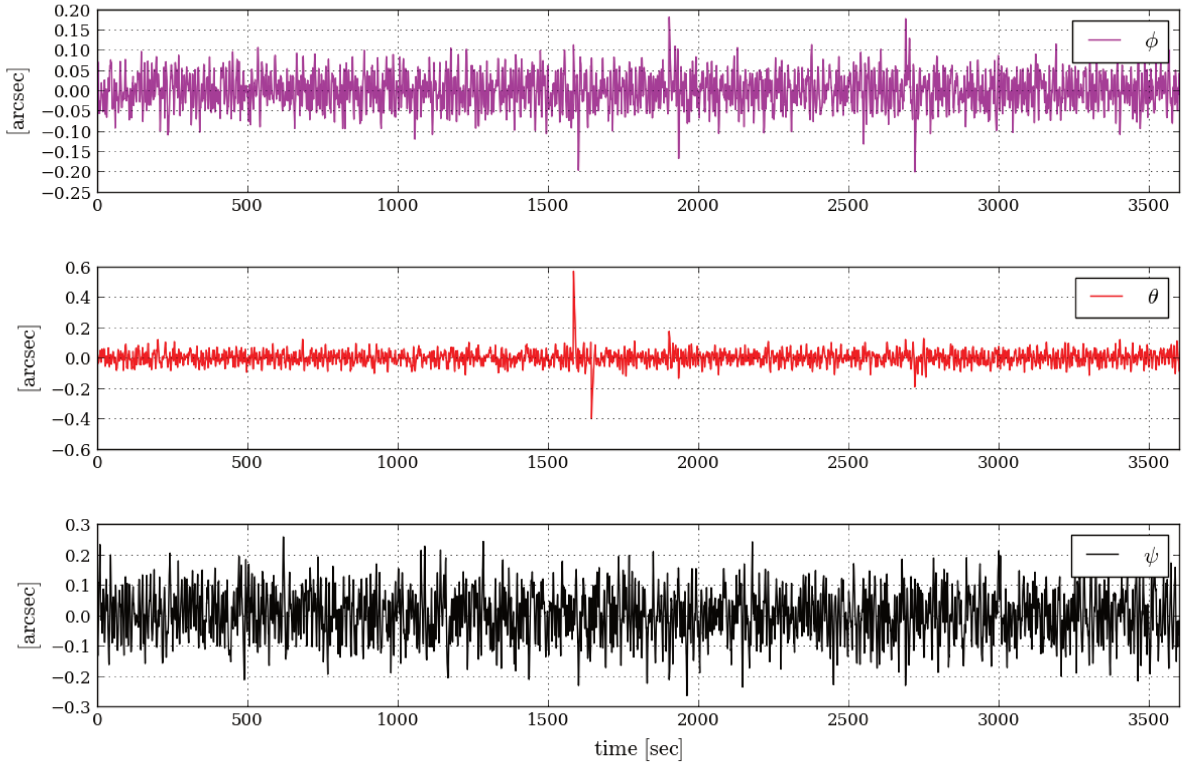
$$\begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix} = R(\phi_k, \theta_k, \psi_k) \begin{bmatrix} x_{\text{ref}} \\ y_{\text{ref}} \\ z_{\text{ref}} \end{bmatrix}. \tag{3.7}$$

The rotation matrix $R(\phi_k, \theta_k, \psi_k)$ is the scalar product of the rotation matrices $R_x(\phi_k)$, $R_y(\theta_k)$ and $R_z(\psi_k)$, respectively relative to the Euler angles ϕ_k, θ_k and ψ_k at instant k, i.e.

$$R(\phi_k, \theta_k, \psi_k) = R_z(\psi_k) \cdot R_y(\theta_k) \cdot R_x(\phi_k). \tag{3.8}$$

[5] <http://www.gregslabaugh.net/publications/euler.pdf>

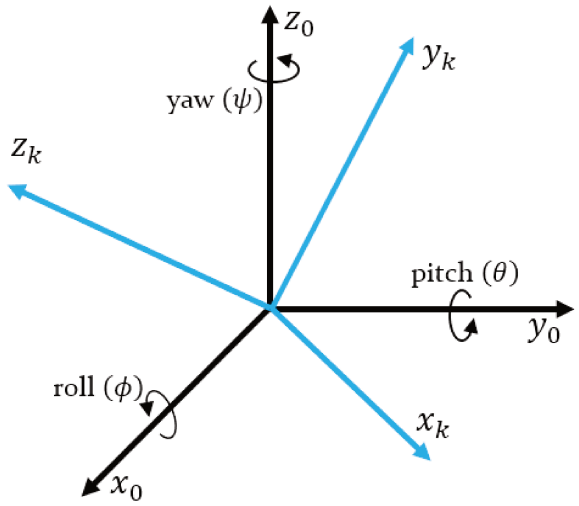
Figure 36 – Illustration of a jitter time-series realization for the PLATO satellite. This time-series has one hour duration, sample frequency of 8 Hz, and is representative of the current status of instrument's pointing requirements.



Source: PLATO Industrial Core Team (OHB-System AG, TAS, RUAG Space).

The vector $(x_{\text{ref}}, y_{\text{ref}}, z_{\text{ref}})$ depends on how the Euler angles $(\phi_k, \theta_k, \psi_k)$ are specified. If rotations are given with respect to the respective previous states, then $(x_{\text{ref}}, y_{\text{ref}}, z_{\text{ref}}) = (x_{k-1}, y_{k-1}, z_{k-1})$. If rotations are given with respect to the respective nominal positions, then $(x_{\text{ref}}, y_{\text{ref}}, z_{\text{ref}}) = (x_0, y_0, z_0)$.; this is the case of the jitter time-series shown in Figure 36. $R_x(\phi_k)$, $R_y(\theta_k)$ and $R_z(\psi_k)$ are defined as

Figure 37 – Representation of the Euler angles (ϕ, θ, ψ) in the body reference frame.



Source: author.

$$
\begin{aligned}
R_x(\phi_k) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_k & -\sin\phi_k \\ 0 & \sin\phi_k & \cos\phi_k \end{bmatrix}, \\
R_y(\theta_k) &= \begin{bmatrix} \cos\theta_k & 0 & \sin\theta_k \\ 0 & 1 & 0 \\ -\sin\theta_k & 0 & \cos\theta_k \end{bmatrix}, \\
R_z(\psi_k) &= \begin{bmatrix} \cos\phi_k & -\sin\phi_k & 0 \\ \sin\phi_k & \cos\phi_k & 0 \\ 0 & 0 & 1 \end{bmatrix}.
\end{aligned} \tag{3.9}
$$

Joining Equations 3.6, 3.7, 3.8 and 3.9 yields

$$
\begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix} = \begin{bmatrix} \cos\theta_k \cos\psi_k & \sin\phi_k \sin\theta_k \cos\psi_k - \cos\phi_k \sin\psi_k & \cos\phi_k \sin\theta_k \cos\psi_k + \sin\phi_k \sin\psi_k \\ \cos\theta_k \sin\psi_k & \sin\phi_k \sin\theta_k \sin\psi_k + \cos\phi_k \cos\psi_k & \cos\phi_k \sin\theta_k \sin\psi_k - \sin\phi_k \cos\psi_k \\ -\sin\theta_k & \sin\phi_k \cos\psi_k & \cos\phi_k \cos\theta_k \end{bmatrix} \begin{bmatrix} \sin(\xi_0)\cos\zeta_0 \\ \sin(\xi_0)\sin\zeta_0 \\ \cos(\xi_0) \end{bmatrix}. \tag{3.10}
$$

Each new star position (ξ_k, ζ_k) in the camera's reference frame can be determined from (x_k, y_k, z_k) by applying

$$
\begin{bmatrix} \xi_k \\ \zeta_k \end{bmatrix} = \begin{bmatrix} \pi/2 - \operatorname{atan2}\left(z_k \sin(\zeta_k)/y_k\right) \\ \operatorname{atan2}\left(y_k/x_k\right) \end{bmatrix}, \tag{3.11}
$$

with corresponding position deviations $(\Delta\xi_k, \Delta\zeta_k)$ given by

$$\begin{bmatrix} \Delta\xi_k \\ \Delta\zeta_k \end{bmatrix} = \begin{bmatrix} \xi_k - \xi_0 \\ \zeta_k - \zeta_0 \end{bmatrix}. \tag{3.12}$$

Finally, Equation 3.4 is employed for determining the new star position $(\mathbf{x}_k, \mathbf{y}_k)$ – in units of pixel – projected on the CCD

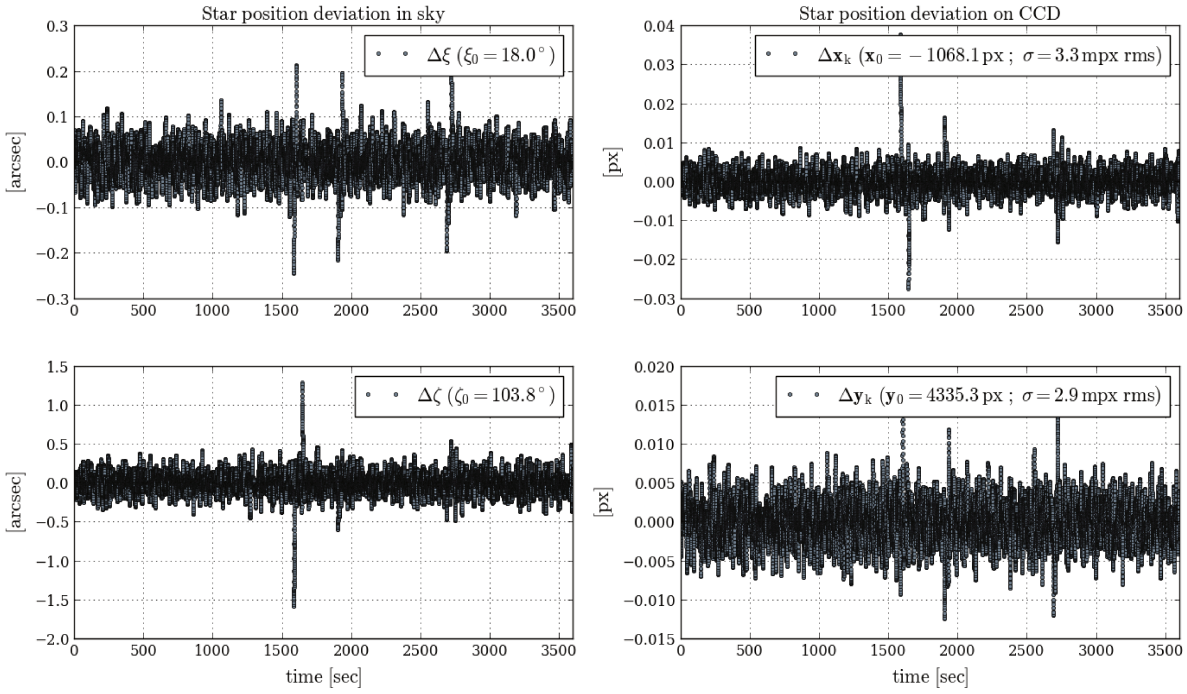
$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathsf{f}\ \tan(\xi_k 0)\cos\zeta_k \\ \mathsf{f}\ \tan(\xi_k)\sin\zeta_k \end{bmatrix}, \tag{3.13}$$

with corresponding position deviations $(\Delta\mathbf{x}_k, \Delta\mathbf{y}_k)$ given by

$$\begin{bmatrix} \Delta\mathbf{x}_k \\ \Delta\mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{x}_k - \mathbf{x}_0 \\ \mathbf{y}_k - \mathbf{y}_0 \end{bmatrix}. \tag{3.14}$$

For illustration, Figure 38 shows the jitter time-series of Figure 36 in the camera reference frame using Equation 3.12 and Equation 3.14. The calculation was performed considering a star near the edge of the camera field of view ($\xi_0 = 18°$). We note that, at CCD level, the nominal satellite jitter has a scatter of only a few milipixel.

Figure 38 – Representation of the jitter time-series realization of PLATO satellite (Figure 36) in the camera reference frame using Equation 3.12 and Equation 3.14.



Source: author.

3.5 A synthetic PLATO P photometric passband

3.5.1 Definition

The PLATO mission was designed based on stellar magnitudes specified in the visible band. Nevertheless, to avoid the inconvenience of having colour dependency when estimating stellar fluxes at camera detectors, from the visual magnitudes, it is more appropriate to work in a proper instrument photometric band. Therefore, a synthetic P magnitude calibrated in the VEGAMAG system is established herein

$$P = -2.5 \log_{10} \left(\frac{\int_{\lambda_1}^{\lambda_2} \mathcal{I}_\star(\lambda)\, \mathcal{S}_P(\lambda)\, \lambda\, d\lambda}{\int_{\lambda_1}^{\lambda_2} \mathcal{I}_{\text{Vega}}(\lambda)\, \mathcal{S}_P(\lambda)\, \lambda\, d\lambda} \right) + P_{\text{Vega}}, \tag{3.15}$$

where $\mathcal{I}_\star(\lambda)$ is the spectral irradiance of a given star, $\mathcal{I}_{\text{Vega}}(\lambda)$ is the spectral irradiance of the Vega star and P_{Vega} is its magnitude in the P band, assumed to be equal to $V_{\text{Vega}} = 0.023$ mag (Bohlin (2007)). P band zero point is given by

$$zp = 2.5 \log_{10} \left((h\, c)^{-1}\, \Theta_P \int_{\lambda_1}^{\lambda_2} \mathcal{I}_{\text{Vega}}(\lambda)\, \mathcal{S}_P(\lambda)\, \lambda\, d\lambda \right) + P_{\text{Vega}}. \tag{3.16}$$

This constant (see Table 7) provides a straightforward way for switching between stellar flux and magnitudes using

$$P = -2.5 \log_{10} \left((h\, c)^{-1}\, \Theta_P \int_{\lambda_1}^{\lambda_2} \mathcal{I}_\star(\lambda)\, \mathcal{S}_P(\lambda)\, \lambda\, d\lambda \right) + zp. \tag{3.17}$$

Thus, having the zero point constant zp and the magnitude P of a given star, its corresponding photometric flux f_P (per camera and expressed in units of $\text{e}^-\ \text{s}^{-1}$) can be estimated with

$$f_P = 10^{-0.4(P - zp)}. \tag{3.18}$$

3.5.2 Relationship with Johnson's V band

For switching between P and V magnitudes, a $V - P$ relationship is determined using the Johnson-Cousins V filter (Figure 23) and modelling $\mathcal{I}_\star(\lambda)$ with (A- to M-type) synthetic stellar spectra extracted from the POLLUX database[6] (Palacios *et al.* (2010)). The template `alpha_lyr_stis_008` (Figure 23) from CALSPEC is adopted for modelling $\mathcal{I}_{\text{Vega}}(\lambda)$, i.e. the calibration star Vega. The resulting $V - P$ samples are shown in Figure 39 as a function of the effective temperature $\mathcal{T}_\star$ of the synthetic spectra $\mathcal{I}_\star(\lambda)$, the former

[6] Extensive description of its content is available at <http://npollux.lupm.univ-montp2.fr/user-s-guide>.

Table 7 – Zero points zp of the synthetic P, G, $G_{\rm BP}$ and $G_{\rm RP}$ photometric passbands calibrated with Vega `alpha_lyr_stis_008` model.

Photometric Passband	Vega [mag]	zp [mag]	zp dev. (A) [mag]	zp dev. (B) [mag]
P	0.023	20.62		
G	0.029	25.6879	4.6×10^{-4}	4.70×10^{-2}
$G_{\rm BP}$	0.039	25.3510	4.3×10^{-4}	1.10×10^{-2}
$G_{\rm RP}$	0.023	24.7450	1.69×10^{-2}	1.50×10^{-2}

Source: author.

Note: Vega magnitudes for *Gaia* passbands are extracted from Casagrande & VandenBerg (2018). Deviations (zp dev.) of G, $G_{\rm BP}$ and $G_{\rm RP}$ zero points are computed with respect to the reference DR2 magnitudes presented in Evans *et al.* (2018) (A) and the revised versions in Weiler (2018) (B).

ranging from 4,000K to 15,000K by steps of 500K. The corresponding third order fitted polynomial for $V - P$ is

$$V - P = -1.184 \times 10^{-12}(\mathcal{T}_\star)^3 + 4.526 \times 10^{-8}(\mathcal{T}_\star)^2 - 5.805 \times 10^{-4}\mathcal{T}_\star + 2.449. \quad (3.19)$$

Therefore, for a star with known visual magnitude V and effective temperature $\mathcal{T}_\star$, one can determine its P magnitude with Equation 3.19 and then apply Equation 3.18 for estimating the corresponding photometric flux at PLATO detectors. Table 8 shows the predicted flux $f_P^{\rm ref}$ for a reference PLATO target – i.e. a 6,000K G0V spectral type star – as a function of its V and P magnitudes. The values include brightness attenuation due to vignetting for a source at $\xi = 14°$ (see Figure 21). In this scenario, a reference PLATO star with $V = 11$ has $P = 10.66$ and $f_P^{\rm ref} = 9.074\,{\rm ke^-\,s^{-1}}$ per camera, or $\sim 218\,{\rm ke^-\,s^{-1}}$ when cumulating over 24 cameras.

3.5.3 Obtaining P and V from *Gaia* magnitudes

Since the present work is based on stars observed with *Gaia*, expressions need to be derived for converting from the magnitude scales available in this catalogue to our synthetic V and P magnitudes. *Gaia* collects data in three photometric systems: G, $G_{\rm BP}$, and $G_{\rm RP}$. As defined in Jordi *et al.* (2010), all of them are calibrated in the VEGAMAG system, following therefore the same philosophy as Equation 3.15, Equation 3.16, and Equation 3.17, i.e. with fluxes being integrated over the corresponding bandwidth limited spectral responses $\mathcal{S}_G(\lambda)$, $\mathcal{S}_{G_{\rm BP}}(\lambda)$ and $\mathcal{S}_{G_{\rm RP}}(\lambda)$. To keep consistency with the definitions of our V and P bands, we apply the same Vega model for producing synthetic *Gaia*

Table 8 – Normal camera (N-CAM) predicted flux f_P^{ref} for a reference 6,000K G0V star as a function of its V and P magnitudes. Respective photon noise is computed by integrating the flux over one exposure and sampling it at each cadence, resulting in 144 samples over one hour.

V [mag]	P [mag]	f_P^{ref} (per camera) [$10^3\,\mathrm{e^-/s}$]	f_P^{ref} (24 cameras) [$10^3\,\mathrm{e^-/s}$]	Photon noise (24 cameras) [ppm $\mathrm{hr}^{1/2}$]
8.0	**7.66**	143.820	3,451.7	9.8
8.5	**8.16**	90.745	2,177.9	12.3
9.0	**8.66**	57.256	1,374.1	15.5
9.5	**9.16**	36.126	867.0	19.5
10.0	**9.66**	22.794	547.1	24.6
10.5	**10.16**	14.382	345.2	31.0
11.0	**10.66**	9.074	217.8	39.0
11.5	**11.16**	5.726	137.4	49.1
12.0	**11.66**	3.613	86.7	61.8
12.5	**12.16**	2.279	54.7	77.7
13.0	**12.66**	1.438	34.5	97.9
13.5	**13.16**	0.907	21.8	123.2
14.0	**13.66**	0.573	13.7	155.1
14.5	**14.16**	0.361	8.7	195.3
15.0	**14.66**	0.228	5.5	245.9
15.5	**15.16**	0.144	3.5	309.5
16.0	**15.66**	0.091	2.2	389.7

Source: author.

Note: Values include vignetting (see Table 4) for a source at $\xi = 14°$ (see Figure 21) and are consistent with the current status of PLATO's instrument parameters described in Börner *et al.* (2018).

bands. Consequently, we impose to this model the corresponding Vega magnitudes listed in Casagrande & VandenBerg (2018). Table 7 summarizes the obtained zero points for the synthetic G, G_{BP}, and G_{RP} bands from this approach. We note that these zero points present satisfactorily low deviations with respect to the reference DR2 magnitudes published in Evans *et al.* (2018), and the later improved versions in Weiler (2018).

For obtaining both P and V magnitudes from the *Gaia* G magnitude, we determine the $G-P$ and $V-P$ relationships by means of the $G_{\mathrm{BP}}-G_{\mathrm{RP}}$ colour index, resulting in the plots shown in Figure 39. The corresponding fitted polynomials are

$$G-P = 0.00652\,(G_{\mathrm{BP}}-G_{\mathrm{RP}})^3 - 0.08863\,(G_{\mathrm{BP}}-G_{\mathrm{RP}})^2 + 0.37112\,(G_{\mathrm{BP}}-G_{\mathrm{RP}}) + 0.00895. \quad (3.20)$$

$$V-P = -0.00292\,(G_{\rm BP}-G_{\rm RP})^3+0.10027\,(G_{\rm BP}-G_{\rm RP})^2+0.37919\,(G_{\rm BP}-G_{\rm RP})+0.00267. \quad (3.21)$$

Unlike Equation 3.19, the expressions in Equation 3.20 and Equation 3.21 are described as a function of the $G_{\rm BP}-G_{\rm RP}$ color index, rather than the effective temperature $\mathcal{T}_\star$. The reason for that is the low availability of effective temperatures in DR2 (less than 10% of the sources). In contrast, $G_{\rm BP}$ and $G_{\rm RP}$ magnitudes are simultaneously available for more than 80% of the sources. To verify the consistency of the synthetic calibrations derived from synthetic stellar spectra, we compare our $V-G=(V-P)-(G-P)$ relationship with the $V-G$ polynomial fit (Busso *et al.* (2018)) derived from Landolt[7] standard stars (398 sources) observed with *Gaia*. As shown in Figure 39, the synthetic $V-G$ curve from this work exhibits satisfactory agreement with the $V-G$ polynomial fit obtained from the true *Gaia* observations. The maximum absolute error between both curves is 9.8×10^{-2} mag at $G_{\rm BP}-G_{\rm RP}=2.75$ mag. Hence, for the purposes of this work, we assume that the polynomials of Equation 3.20 and Equation 3.21 provide sufficiently accurate estimates of P and V magnitudes from the G magnitude of the DR2 catalogue.

3.6 Identifying target and contaminant stars

This subsection delineates the ensemble of target and contaminant stars, from the input catalogue, that shall be used to build input images for simulating aperture photometry. First, the position of each star within IF is determined at the focal plane array of one PLATO camera (as explained in subsection 3.2, IF covers exactly the field of a single camera). Next, following the definition of the P5 sample, we assign as targets all stars located within IF that have $0.57\leqslant G_{\rm BP}-G_{\rm RP}\leqslant 1.84$ (F5 to late-K spectral types) and P magnitude in the range $7.66\leqslant P\leqslant 12.66$, the latter corresponding to $8.0\leqslant V\leqslant 13.0$ for a reference PLATO target, i.e. a 6,000K G0V star. This accounts for about 127,000 sources. Target selection based on the P band is more convenient than the V band, as it allows us to overcome the colour dependency of the latter. In other words, this approach ensures that all targets assume flux values within a fixed range (that of Table 8), regardless of their effective temperature. This is thus consistent with a target selection strategy driven by noise performance, magnitude, and, spectral type. As for the contaminant stars, they correspond to all existing sources in the input catalogue located within 10 pixel radius around all targets. This accounts for about 8.3 million stars with P magnitude comprised in the range $2.1\lesssim P\lesssim 21.1$. We note that only sources satisfying $-0.227\leqslant G_{\rm BP}-G_{\rm RP}\leqslant 4.524$ are included in the ensemble of contaminant stars to conform with the range of applicability of the polynomials described in Equation 3.20 and Equation 3.21. According to the above description, Figure 40 presents some statistics

[7] <https://www.eso.org/sci/observing/tools/standards/Landolt.html>.

(distances and differential magnitudes) on the distribution of contaminant stars relative to their corresponding targets.

3.7 Setting up the imagettes

Now that both target and contaminant stars from IF are properly identified and their respective fluxes can be accurately predicted using Equation 3.20 and Equation 3.18, the input imagettes can finally be computed. For that, the following procedure is applied:

a) Use the input catalogue and derived properties to obtain magnitudes, fluxes and the locations of target and contaminant stars at intrapixel level. From that, we consider a stellar subset composed of 50,000 targets (from the total of $\sim$ 127k targets within IF). These are neighboured by $\sim$3.25 million contaminant stars;

b) Employ the PSFs presented in Section 2.9.2 as models of both instrument optical and detector responses to stellar flux;

c) By convention, the pixels of an imagette are selected such that the centre of the resulting imagette is located at no more than an absolute euclidean distance of 0.5 pixel from any of the x and y coordinates of the target barycentre (see examples in Figure 22). This is done to maximize the amount of target energy falling within its imagette;

d) Translate each star position deviation $(\Delta\mathbf{x}_k, \Delta\mathbf{y}_k)$ of the satellite pointing time-series from Figure 38 into a corresponding shifted imagette with respect to the nominal position (zero). The resulting set of shifted imagettes shall be used as input to compute the amount of jitter noise in the photometry.

Following this process, an input image (reference frame) like that illustrated in Figure 41 was generated for each target (including respective contaminants). Shifted images to account for satellite motion are therefore computed target by target with respect to their respective reference frames.

3.8 Discussions

In this chapter, we described in detail the primary inputs needed for constructing images (imagettes) to be used for simulating the performance of aperture photometry, including

a) an expression for estimating the per camera intensities of zodiacal light, which is the dominant source of diffuse sky light at Lagrange-L2 orbit where PLATO spacecraft will be launched in;

b) the input stellar catalogue (*Gaia* DR2) and the input field, the latter corresponding to the sky area covered by a single PLATO camera centred at the centre of the Southern PLATO Field (SPF);

c) a PLATO P photometric passband calibrated in the VEGAMAG system for consistently estimating stellar flux on PLATO detectors, including color relationships (polynomials) with *Gaia* and Johnson's V magnitude scales;

d) identification of potential PLATO target stars based on their P magnitude and spectral type (F5 to late-K type stars), according to the scientific requirements defined for the P5 sample (see subsection 2.5);

e) identification of all contaminant stars located at up to 10-pixel radius around all considered targets stars;

f) mathematical expressions for projecting stellar positions from the sky onto the focal plane of one PLATO camera centred at SPF centre;

g) a procedure for generating imagettes considering stellar positions at both pixel and intrapixel level, and including satellite jitter.

A few important considerations shall be taken into account about some of the items from the above list. In relation to item a), the expression for estimating the zodiacal light at PLATO detectors considers the current status of the instrument design. Hence, such an expression is subjected to change in the future, depending on the evolution of the PLATO instrument parameters, notably its spectral response $\mathcal{S}_P(\lambda)$.

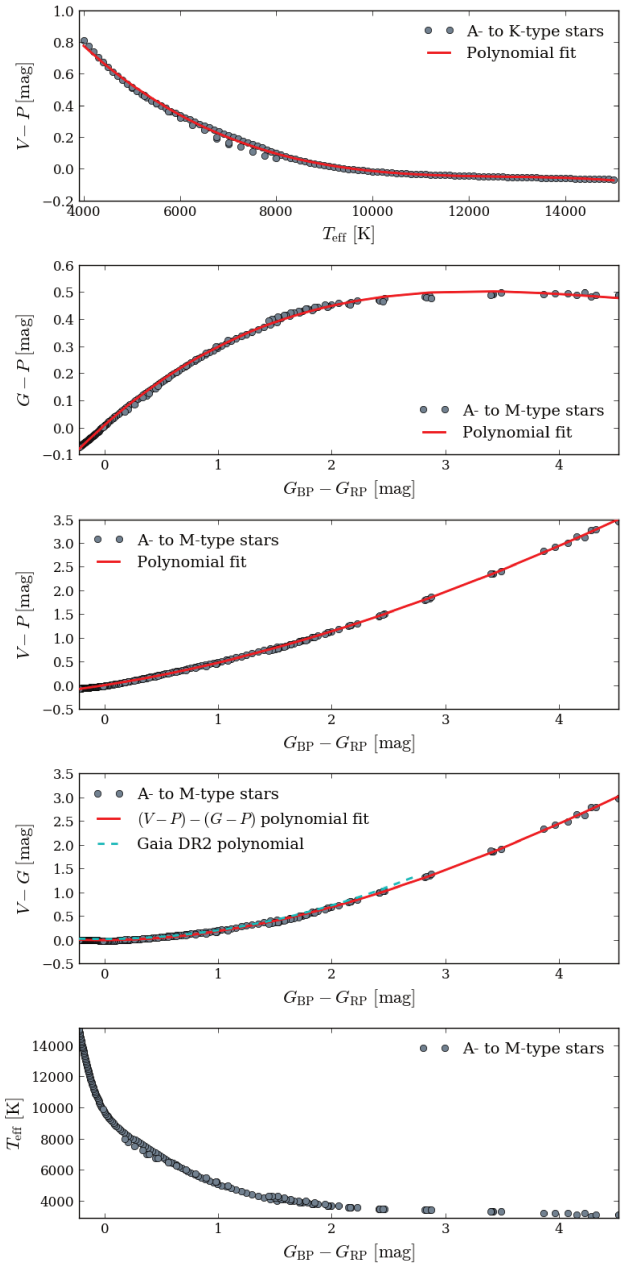
In regard to item c), the synthetic PLATO P photometric passband shown in this chapter assumes that an absolute empty space exists between the instrument and the stars it observes. That is, the calibration of the P band derived in this work does not account for spectral changes on stellar fluxes caused by the Interstellar Medium (ISM), such as the total stellar abortion A (extinction). However, as long as the flux of a given star is sufficiently well characterized in other magnitude systems (such as *Gaia* G and Johnson's V), converting it then to the corresponding P magnitude – using the calibration adopted in this work – should logically give sufficiently reliable results in the PLATO magnitude. Besides, PLATO targets are mostly composed of bright (close) stars, so the effects of stellar extinction are not expected to represent a significant impact on the statistical noise performance characterization of aperture photometry presented in this thesis. Moreover, the current P magnitude will inevitably require further recalibration once the final spectral response of the instrument is characterized. At a certain extent hence, the actual uncertainties on the instrument design are still likely to represent greater impact on stellar flux estimations than those related to stellar abortion. Concerning item d), selecting targets based on the P band, compared to visual V, avoids the inconvenience of obtaining colour-dependant noise performance. We can infer from Figure 39 that different stars with

the same V magnitude may effectively have identical fluxes on PLATO detectors (and vice-versa), which is ambiguous.

In regard to item e), the particular choice of considering contaminant stars located at up to 10-pixel radius around the targets was made to make sure that all contaminant stars with sufficient flux to generate false background transits, should they be eclipsed, are taken into account in the present study. In the next chapter, some results are presented and discussed evidencing that such a threshold is more than enough for the purposes of this work.

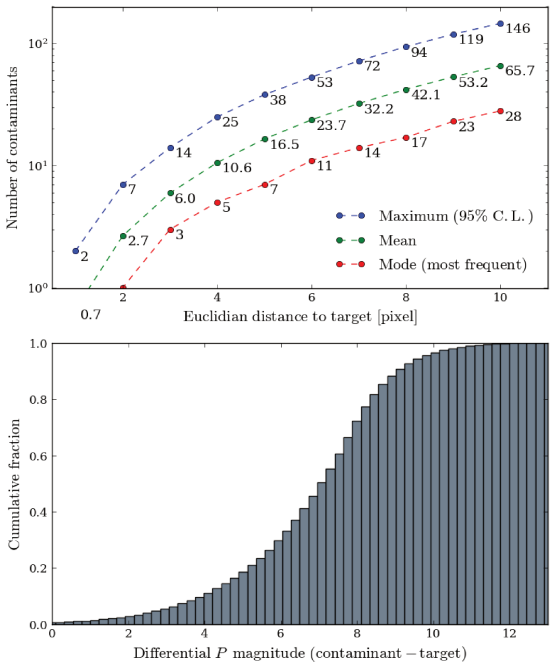
Finally, a few considerations are necessary concerning the sources in *Gaia* DR2. Evans *et al.* (2018) reported some very likely saturation and imperfect background subtraction issues affecting sources with $G \lesssim 3.5$ and $G \gtrsim 17$, respectively. Since the central point in this work is to establish a relative performance comparison between different photometric aperture models – particularly in scenarios of high stellar crowding – those sources were not removed from the working subset of stars. The inaccuracies resulting from the mentioned issues will ultimately be evenly propagated to all tested mask models, having therefore no potential to significantly impact the comparative basis analysis.

Figure 39 – Relationships between the photometric passbands V, P, and G, obtained by modelling $f(\lambda)$ with A- to M-type synthetic stellar spectra extracted from the POLLUX database. Red polynomials are derived from Eqs. 3.19, 3.20, and 3.21, and are applicable within the range $4,000 \leqslant T_{\rm eff} \leqslant 15,000$ (top frame) and $-0.227 \leqslant G_{\rm BP} - G_{\rm RP} \leqslant 4.524$. *Gaia* DR2 polynomial is based on Landolt stars observed with *Gaia* and is applicable within the range $-0.5 \leqslant G_{\rm BP} - G_{\rm RP} \leqslant 2.75$. It has a scatter of 4.6×10^{-2} mag (Busso *et al.*, 2018). The relationship between $T_{\rm eff}$ and the synthetic $G_{\rm BP} - G_{\rm RP}$ color index (bottom frame) is consistent with the color–temperature relations published in Andrae *et al.* (2018).



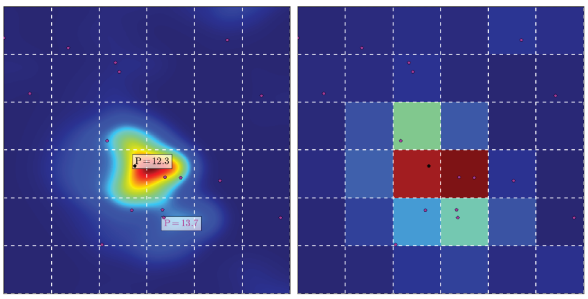
Source: author.

Figure 40 – **Top**: Number of contaminants brighter than $P \sim 21.1$ as a function of the Euclidean distance from the target stars ($7.66 \leqslant P \leqslant 12.66$). Maximum values have 95% confidence level. **Bottom**: Cumulative distribution function of the differential P magnitude between contaminant and target stars, the former located at up to 10 pixels distant from the latter.



Source: author.

Figure 41 – Example of input image. **Left**: high resolution PSF ($\xi = 18°$) for a target with $P = 12.3$ (barycentre designated by the black dot) surrounded by several contaminants within a 3-pixel radius (respective barycentres designated by the magenta dots). The brightest contaminant in the frame (below the target) has $P = 13.7$. All other contaminants are at least 2 mag fainter than the target. **Right**: corresponding low resolution PSF (imagette) at pixel level. Dashed white lines represent pixel borders.



Source: author.

4 Aperture photometry

As explained in subsection 1.6, in-flight (aperture) photometry extraction will be applied to a significantly large number of targets from PLATO's P5 stellar sample (Table 1). In that context, this chapter describes in detail the approach adopted for determining the optimal collection of pixels (aperture) to perform such task, with special attention given to the problematic of false positives caused by background eclipsing objects.

The work presented in this chapter is strongly based on Marchiori *et al.* (2019). In-flight photometry extraction of PLATO targets: Optimal apertures for detecting extra-solar planets, A&A, 627, A71.

4.1 State of the art

There is a noteworthy number of publications on the theme of photometric masks. Among the oldest, some emphasis should be given to the work of Howell (1989), in which the idea of a growth curve (signal-to-noise ratio as a function of aperture radius) for point-source observations is introduced, showing how much the photometric performance of an aperture varies as a function of its size in pixels. This simple and useful feature also helped, at the time, better evidencing that the collection of pixels providing maximum signal-to-noise ratio is greatly dependant on the brightness of the source. Also dating from the late 1980s, the stellar photometry package DAOPHOT[1] (Stetson (1987)) is still used today to perform point-source photometry using circular apertures approximated as irregular polygons (Bajaj & Khandrika (2017)). Similarly, the Python package PhotUtils[2] allows us to extract photometry from astronomical sources based on circular, elliptical or rectangular apertures. Coming next, Naylor (1998) proposes employing weighted masks for imaging photometry, providing improved ($\sim 10\%$) noise-to-signal ratio (NSR) performance compared to non-weighted (binary) masks. Later on, and oriented to planet transit finding and asteroseismology, Llebaria & Guterman (2006) and Bryson *et al.* (2010) developed strategies to compute optimized binary masks for extracting light curves from CoRoT and *Kepler* targets, respectively. In the particular context of *Kepler's* data processing pipeline, binary masks are referred to as simple aperture photometry. They were primarily designed to minimize noise for maximum transit detection sensitivity and as input for determining a halo of pixels to be downlinked along with the aperture pixels. More recently, Smith *et al.* (2016) propose a new method to assign apertures for *Kepler* targets, focused on planet detection and mitigation of systematic errors, through an optimization scheme based on

[1] <http://www.star.bris.ac.uk/~mbt/daophot/>
[2] <https://photutils.readthedocs.io/en/stable/index.html>

NSR and Combined Differential Photometric Precision (CDPP)[3] (Jenkins *et al.* (2010b)). As described in *Kepler's* Data Processing Handbook[4], this method is implemented within the Photometry Analysis (PA) component of *Kepler's* science pipeline. Alongside, Aigrain *et al.* (2015) and Lund *et al.* (2015) provide techniques for mask pixel selection for *Kepler* K2 targets. The former proposes soft-edging top-hat-like apertures, which has satisfactory performance for sufficiently bright targets and is relatively robust to uncertainties on the knowledge of the PSF. The latter uses clustering of pixels, which best fits the flux distribution of several nearby targets, being therefore more suitable for dense fields. A modified version of this method is employed in Handberg & Lund (2016) for reducing the data of *Kepler* K2 targets from campaigns 0 to 4. Besides, it is also considered as one of the possible solutions for extracting light curves from TESS targets (Lund *et al.* (2017)).

4.2 Proposed approach for PLATO P5 targets

As pointed out in subsection 2.5, the primary focus of the P5 stellar sample is to generate large statistics on planet occurrence rate. Hence, the aperture photometry methods developed for the space missions CoRoT, *Kepler* and TESS are of greater interest in the context of the present work, as these are the greatest references in the domain of exoplanet search. Considering these three examples, it is noticeable that the term optimal is recurrently employed to distinguish apertures that minimize NSR or some noise-related metric such as CDPP. That is, of course, a reasonable way to proceed because the sensitivity at which a planet transit can be found in a light curve, for instance, is strongly correlated to its noise level. On the other hand, the higher the ease in identifying a transit-like signal, either because of sufficiently low NSR or CDPP, the higher the probability that a background object in the scene generates a TCE. This background object could be, for example, an EB mimicking a true planet transit. Although background false positives may be efficiently identified in certain cases when, besides the light curves, the corresponding pixel data is also available – as demonstrated by Bryson *et al.* (2013) – most of the stars in P5 unfortunately lack that extra information (see Table 1) because of telemetry constraints related to the large number of tagetes in this stellar sample. In this particular unfavoured scenario, conceiving photometric masks based uniquely on how well a transit-like signal can be detected, ignoring potential false positives may not be the best strategy. To verify the consistence of this hypothesis, two science metrics are introduced for directly quantifying the sensitivity of an aperture in detecting true and false[5] planet transits. The idea is then to verify whether or not the best compromise between these two parameters is obtained from apertures having overall lower NSR.

[3] Roughly speaking, CDPP is an estimate of how well a transit-like signal can be detected (Smith *et al.* (2016)).

[4] <https://archive.stsci.edu/kepler/manuals/>

[5] This work addresses the occurrence of false planet transits caused by background eclipsing objects, in particular EBs.

4.3 Photometric precision

4.3.1 Nominal NSR

The NSR is the principal performance indicator for evaluating the exploitability of photometry signals. For PLATO stellar light curves derived from aperture photometry applied to imagettes, we use the following metric to compute the per cadence NSR (NSR_*; see parameters description in Table 9):

$$\mathrm{NSR}_* = \frac{\sqrt{\sum_{n=1}^{36}\left(\sigma^2_{F_{T_n}} + \sum_{k=1}^{N_C}\sigma^2_{F_{C_{n,k}}} + \sigma^2_{B_n} + \sigma^2_{D_n} + \sigma^2_{Q_n}\right) w_n^2}}{\sum_{n=1}^{36} F_{T_n}\, w_n}. \tag{4.1}$$

A per cadence light curve sample corresponds to the integrated mask flux over one exposure interval of the detectors, which corresponds to 21 seconds (Table 2) for PLATO N-CAM. In the context of PLATO, NSR scales over multiple independent samples and measurements,

$$\mathrm{NSR} = \frac{10^6}{12\sqrt{t_d\, N_T}}\,\mathrm{NSR}_*, \tag{4.2}$$

where t_d is the observation duration in hours and N_T is the number of telescopes observing the star. The constant in the denominator of the above expression stands for the square root of the number of samples in one hour, i.e. $\sqrt{3600\mathrm{s}/25\mathrm{s}} = 12$, based on the 25 seconds cadence (Table 2) of the PLATO N-CAMs. The NSR of a signal with duration of one hour ($\mathrm{NSR}_{1\mathrm{h}}$) is therefore (expressed in units of $\mathrm{ppm\,hr}^{1/2}$)

$$\mathrm{NSR}_{1\mathrm{h}} = \frac{10^6}{12\sqrt{N_T}}\,\mathrm{NSR}_*. \tag{4.3}$$

We note that the flux noise induced by satellite jitter is not included in Eq. 4.1. In fact, attempting to do so would be a fairly complicated task because jitter contribution depends on the final shape of the aperture (Fialho *et al.* (2007)). Considering that the nominal jitter amplitudes of PLATO spacecraft have a scatter of the order of a few milipixel at detector level (see Figure 38), jitter noise can be included a posteriori to the determination of the apertures. Later in subsection 4.9, we discuss in what circumstances this might be an acceptable approach.

4.3.2 NSR including satellite jitter

Once the apertures are determined, we include jitter noise in the photometry using the shifted imagettes described in Section 3.7. Let NSR_*^{jitter} be the per cadence NSR, which includes star motion due to satellite jitter. NSR_*^{jitter} is given by

$$\mathrm{NSR}_*^{jitter} = \mathrm{NSR}_*\sqrt{1 + \left(\frac{\sigma_J^2}{\sigma_*^2}\right)}, \tag{4.4}$$

Table 9 – Description of the parameters of Equation 4.1.

Description	Symbol
Photon noise from the target star	$\sigma^2_{F_T}$
Photon noise from a contaminant star	$\sigma^2_{F_C}$
Background noise from the zodiacal light	σ^2_B
Overall detector noise (including readout, smearing and dark current)	σ^2_D
Quantization noise	σ^2_Q
Average flux from the target star	F_T
Average flux from a contaminant star	F_C
Mask weight in the interval $[0, 1]$	w
imagette pixel index $= \{1, 2, 3, \ldots, 36\}$	n
Contaminant star index $= \{1, 2, 3, \ldots, N_C\}$	k
Number of contaminant stars within 10-pixel radius around the target	N_C

Source: author.

where σ_J is the photometric jitter noise obtained from the shifted imagettes and σ_* corresponds to the numerator of the expression in Equation 4.1. The above expression assumes stationary random noise for both photometric flux and satellite jitter.

4.4 Confusion

We present herein the SPR. This factor allows us to quantify the average fractional contaminant flux from background stars captured by an aperture. Letting $F_{C,k}$ be the photometric flux contribution from a single contaminant star k and F_{tot} the total flux, we have

$$F_{C,k} = \sum_{n=1}^{36} F_{C_{n,k}}\, w_n, \tag{4.5}$$

$$F_{tot} = \sum_{n=1}^{36} \left(F_{T_n} + B_n + \sum_{k=1}^{N_C} F_{C_{n,k}} \right) w_n, \tag{4.6}$$

where B_n is the average background flux at pixel n from the zodiacal light. SPR_k is denoted as the fractional flux from the contaminant star k with respect to the total photometric flux (target plus contaminants and zodiacal light), i.e.

$$\mathrm{SPR}_k = \frac{F_{C,k}}{F_{tot}}. \tag{4.7}$$

Accordingly, the fractional flux from all contaminant stars is

$$\mathrm{SPR}_{tot} = \sum_{k=1}^{N_C} \mathrm{SPR}_k. \tag{4.8}$$

It is worthy noticing that SPR_{tot} is complementary to the crowding metric r defined in Batalha *et al.* (2010), i.e. $\mathrm{SPR}_{tot} = 1 - r$.

4.5 Detectability of planet transits

When a planet eclipses its host star, it produces a maximum transit depth δ_p which is, at first order approximation[6], equal to the square of the ratio between the planet radius and the star radius

$$\delta_p = \left(R_p / R_\star\right)^2. \tag{4.9}$$

In practice, δ_p is always diluted by the contaminant flux from surrounding stars and background light, such that the observed transit depth δ_{obs} is a fraction of the original transit depth δ_p

$$\delta_{obs} = \left(1 - \mathrm{SPR}_{tot}\right) \delta_p. \tag{4.10}$$

Traditionally, a planet detection is not considered scientifically exploitable unless it has been observed at least three times. Furthermore, observed transits must reach a certain level of statistical significance, η, of the total noise, σ. In this work, we adopt the threshold[7] of 7.1σ ($\eta_{min} = 7.1$) as the minimum condition for characterizing a TCE with three transits. This yields

$$\delta_{obs} \geqslant \eta_{min}\, \sigma = 7.1\sigma. \tag{4.11}$$

The total noise σ scales with the signal (transit) duration t_d and with the number of transit events n_{tr}, resulting

$$\sigma = \mathrm{NSR}_{1\mathrm{h}} / \sqrt{t_d\, n_{tr}}. \tag{4.12}$$

By combining the above expressions, we can determine the range of detectable planet radius (c.f. Batalha *et al.* (2010))

$$R_p \geqslant R_\star \sqrt{\frac{\eta}{\left(1 - \mathrm{SPR}_{tot}\right)} \frac{\mathrm{NSR}_{1\mathrm{h}}}{\sqrt{t_d\, n_{tr}}}}. \tag{4.13}$$

[6] Discussions on how to compute planetary transit depths more accurately are provided in Heller (2019).

[7] This criterion was established to ensure that no more than one false positive due to random statistical fluctuations occurs over the course of the *Kepler* mission (Jenkins *et al.* (2010)).

Earth-like planets located at about 1au from Sun-like stars have $\delta_p \sim 84$ ppm and $t_d \sim 13$ hours. Consequently, it is required that $\mathrm{NSR_{1h}} \lesssim 74\mathrm{ppm\,hr^{1/2}}$ for that type of planet to be detected at $\eta = \eta_{min} = 7.1$, $n_{tr} = 3$ and $\mathrm{SPR}_{tot} = 0$. From Equation 4.13, we can obtain the statistical significance η at which a planet can be detected

$$\eta = \delta_p \sqrt{t_d\, n_{tr}}\,(1 - \mathrm{SPR}_{tot})\,/\mathrm{NSR_{1h}}. \tag{4.14}$$

We denote hereafter N_{TCE}^{good} as the frequency of target stars satisfying $\eta \geqslant \eta_{min}$ for $n_{tr} \geqslant 3$. Accordingly, the aperture model providing the largest N_{TCE}^{good} is that being more likely in a statistical sense to detect true planet transits.

4.6 Sensitivity to background false transits

In this section, we derive a metric for evaluating the sensitivity of an aperture in detecting false planet transits originated from astrophysical eclipses of contaminant stars. Such events may occur, in particular, when the contaminant star in question is part of an EB system and is sufficiently bright and sufficiently close to a target star. False planet transits caused by grazing EBs are thus not addressed herein.

When a given contaminant star k is eventually eclipsed, the raw photometry reveals a within aperture fractional flux decrease $\Delta F_{C,k}^{raw}$ and a corresponding within aperture fractional magnitude increase $\Delta \mathrm{m}_{C,k}^{raw}$, such that

$$\Delta \mathrm{m}_{C,k}^{raw} = -2.5 \log_{10} \left(\frac{F_{C,k}^{raw} - \Delta F_{C,k}^{raw}}{F_{C,k}^{raw}} \right). \tag{4.15}$$

By denoting $\Delta \mathrm{m}_{C,k}^{raw}$ as the background transit depth $\delta_{back,k}$ in mag units and $\Delta F_{C,k}^{raw}/F_{tot}$ as the resulting observed transit depth $\delta_{obs,k}$ in the raw light curve, relative to the contaminant star k, we obtain

$$\delta_{obs,k} = \mathrm{SPR}_k^{raw} \left(1 - 10^{-0.4\delta_{back,k}}\right), \tag{4.16}$$

with

$$\mathrm{SPR}_k^{raw} = \frac{F_{C,k}^{raw}}{F_{tot}}. \tag{4.17}$$

This expression shows that the background transit depth $\delta_{back,k}$ affects the light curve as an observed transit depth $\delta_{obs,k}$, which is proportional to SPR_k^{raw}, i.e. the SPR of the contaminant star k in the raw photometry. Because $\delta_{obs,k}$ is the result of a false planet transit, we want this depth to be sufficiently small to prevent a TCE to be triggered, i.e.

$$\delta_{obs,k} < \eta_{min}\,\sigma. \tag{4.18}$$

Although the above statement holds if, and only if, the SPR_k is below a certain level for given $\delta_{back,k}$, η, t_d, and n_{tr}. We denote such a threshold as the critical SPR (SPR_k^{crit}) of

the contaminant star k. It can be determined with

$$\mathrm{SPR}_k^{crit} = \frac{\eta}{\left(1 - 10^{-0.4\delta_{back,k}}\right)} \frac{\mathrm{NSR}_{1\mathrm{h}}}{\sqrt{t_d\, n_{tr}}}. \tag{4.19}$$

We denote hereafter N_{TCE}^{bad} as the frequency of target stars satisfying $\mathrm{SPR}_k^{raw} \geqslant \mathrm{SPR}_k^{crit}$ for $\eta \geqslant \eta_{min} = 7.1$ and $n_{tr} \geqslant 3$. Accordingly, the aperture model providing the smallest N_{TCE}^{bad} is that more likely in a statistical sense to naturally reject false planet transits caused by background eclipsing objects.

4.7 Background flux correction

Background correction refers to subtracting, from the raw photometry, flux contributions from contaminant sources and scattered stray light (e.g. zodiacal and Galactic lights). The spatial distribution of background light is commonly described using polynomial models, whose coefficients are determined based on flux measurements taken at strategically selected pixels Drummond *et al.* (2008), Twicken *et al.* (2010). For PLATO, the strategy for background correction is not yet characterized at the present date, thus no accurate information on this subject is available for inclusion in this study. Notwithstanding, we investigate in this section the impact of an ideally perfect background correction on the science metrics N_{TCE}^{good} and N_{TCE}^{bad}, i.e. what happens when $B_n = F_{C,k} = \mathrm{SPR}_k = \mathrm{SPR}_{tot} = 0$.

In this case, the observed depth of a legitimate planet transit simply converges to its true depth, i.e. $\delta_{obs} = \delta_p$ (the transit dilution is completely cancelled). In parallel, the parameter η (Eq. 4.14) increases, meaning that the apertures become more sensitive to detect true planet transits, which ultimately implies an increase in N_{TCE}^{good} as well.

Analysing the impact on N_{TCE}^{bad} is not as straightforward as it is for N_{TCE}^{good}. First, we denote hereafter F_{tot}^{corr} as the total photometric flux resulted after the background correction, which only contains signal from the target

$$F_{tot}^{corr} = \sum_{n=1}^{36} F_{T_n}\, w_n. \tag{4.20}$$

Next, we define $\Delta F_{C,k}^{raw}/F_{tot}^{corr}$ as the resulting observed transit depth $\delta_{obs,k}^{corr}$, after background correction, caused by an eclipse of the contaminant star k. This leads us, using Equation 4.15, to an expression for $\delta_{obs,k}^{corr}$ which is similar to that of Equation 4.16, except that the term $(F_{C,k}^{raw}/F_{tot}^{corr})$ appears in place of SPR_k^{raw}, resulting in

$$\delta_{obs,k}^{corr} = \left(F_{C,k}^{raw}/F_{tot}^{corr}\right)\left(1 - 10^{-0.4\delta_{back,k}}\right). \tag{4.21}$$

The above identity shows that removing the background flux from the photometry does not suppress the false transit caused by a background EB. Indeed, although the average flux from the eclipsing contaminant star goes to zero ($F_{C,k} = 0$) in the corrected photometry,

the transit depth $\delta_{obs,k}^{corr}$ depends on the intrinsic (raw) contaminant flux $F_{C,k}^{raw}$ that is present in the scene, which is thus independent of any further processing applied in the photometry. Besides, this result is consistent with the fact that the background correction only removes the nominal (out-of-transit) average flux of the contaminant source from the photometry, therefore becoming no longer effective if such signal changes after the correction (e.g. owing to an eclipse, i.e. when $\delta_{back,k} \neq 0$).

For convenience, we define hereafter the apparent SPR (SPR_k^{app}), which is manifested during the eclipse of a contaminant star k in a light curve with flux fully corrected for the background

$$\mathrm{SPR}_k^{app} = \frac{F_{C,k}^{raw}}{F_{tot}^{corr}}. \tag{4.22}$$

This yields

$$\delta_{obs,k}^{corr} = \mathrm{SPR}_k^{app}\left(1 - 10^{-0.4\delta_{back,k}}\right). \tag{4.23}$$

Comparing Equation 4.16 and Equation 4.23, it is simple to verify that $\delta_{obs,k}^{corr} > \delta_{obs,k}$ because $\mathrm{SPR}_k^{app} > \mathrm{SPR}_k^{raw}$. This means that the apertures become more sensitive to detect false planet transits from background eclipsing objects when the corresponding photometry is corrected for the average background flux. This happens because the background correction reduces the dilution of such transits.

From all the above considerations, it is possible to state therefore that the background correction is expected to increase both N_{TCE}^{good} and N_{TCE}^{bad} metrics.

4.8 Aperture models

4.8.1 Gradient mask

As noise-to-signal ratio is the main performance parameter to be evaluated, a logical mask model to experiment with is the one having weights w_n providing, for each target, the global minimum NSR_*. Since the masks have by definition the same dimension of the imagettes, i.e. 6×6 pixels, those which are optimal could be determined without much effort by exhaustive search, i.e. by simple trials of several w_n combinations, keeping the one giving the smallest NSR_*. Naturally, that kind of approach is far from being efficient, especially considering that this procedure must be executed for tens of thousands of target stars. To avoid this inconvenience, we developed a direct method for calculating the optimal (minimum) NSR weights. To determine such a mask, we rely on the fact that NSR_*, at its minimum, shall have a gradient identically equal to zero ($\nabla\mathrm{NSR}_* = 0$) with respect to its ($6 \times 6 = 36$) weights, i.e.

$$\frac{\partial \mathrm{NSR}_*}{\partial w_1} = \frac{\partial \mathrm{NSR}_*}{\partial w_2} = \frac{\partial \mathrm{NSR}_*}{\partial w_3} = \cdots = \frac{\partial \mathrm{NSR}_*}{\partial w_{36}} = 0. \tag{4.24}$$

From the above expression we obtain 36 non-linear equations of the form

$$w_n \, \sigma_n^2 \sum_{i=1}^{36} w_i \, F_{T_i} = F_{T_n} \sum_{i=1}^{36} w_i^2 \, \sigma_i^2, \tag{4.25}$$

where n is the mask weight index $= \{1, 2, 3, \ldots, 36\}$, i.e. one for each partial derivative of Equation 4.24, and σ^2 corresponds to the total variance resulted from all noise contributors of Equation 4.1. One simple solution beyond the trivial one for w_n satisfying the above equality can be calculated directly with

$$w_n = \frac{F_{T_n}}{\sigma_n^2}. \tag{4.26}$$

Conventionally, all w_n are then normalized by $\max[w_n]$ to satisfy $0 \leqslant w_n \leqslant 1$, so that each weight w_n directly represents the fraction of the imagette flux being caught by the aperture at the corresponding pixel n. For illustration, Figure 43a shows the resulting gradient mask for the input image example of Figure 41.

In order to simplify terminology, the masks w_n obtained from Equation 4.26 are hereafter referred to as *gradient* masks, based on the fact that they are determined from the mathematical gradient of NSR_* expression. Each time they are mentioned however, we should keep in mind that they correspond to the masks providing the global minimum noise-to-signal ratio from all the possible combinations of mask weights w_n in Equation 4.1. In other words, gradient masks are the optimal NSR masks in the strict sense.

4.8.2 Gaussian mask

Having examined the shape of gradient masks applied to several stars, we noticed that they look quite similar to a bell shaped curve. Therefore, we decided to test Gaussian-like masks to verify whether they could provide near-best NSRs when compared to gradient masks. Depending on the performance difference, the advantage of having an analytical mask that requires fewer parameters to be computed could justify its choice over the gradient mask. On these terms, we calculate the weights w_n of a Gaussian mask using the conventional symmetric Gaussian function expression

$$w_{x,y} = \exp\left(-\frac{(x - x_\star)^2 + (y - y_\star)^2}{2\sigma_w^2} \right), \tag{4.27}$$

where

- (x, y) are Cartesian coordinates of the imagette pixels with shape 6×6;
- $(x_\star, y_\star)$ are the coordinates of the target barycentre within the imagette;
- σ_w is the mask width in pixels on both x and y dimensions;
- $w_{x,y}$ is the mask weight in the interval $[0, 1]$ at (x, y).

As the imagette dimension is fixed and the target position within it is well known thanks to the input catalogue, choosing a Gaussian mask for a given target reduces to finding a proper width. To accomplish it, we simply iterate over different values of σ_w and keep that giving the lowest NSR_*, as shown in Figure 42. For illustration, Figure 43b shows the resulting Gaussian mask for the input image example of Figure 41.

4.8.3 Binary mask

Binary masks are non-weighted apertures, meaning that the photometry is extracted by fully integrating pixel fluxes within the mask domain and discarding those which are outside it. This type of aperture was extensively employed to produce light curves of CoRoT and *Kepler* targets, so this solution is well known for delivering satisfactory performance. In the context of PLATO, we applied the following routine to compute a binary mask for each target imagette.

1. Arrange all pixels n from the target imagette in increasing order of NSR_n

$$\mathrm{NSR}_n = \frac{\sqrt{\sigma^2_{F_{T_n}} + \sum_{k=1}^{N_C} \sigma^2_{F_{C_{n,k}}} + \sigma^2_{B_n} + \sigma^2_{D_n} + \sigma^2_{Q_n}}}{F_{T_n}}. \tag{4.28}$$

2. Compute the aggregate noise-to-signal $\mathrm{NSR}_{agg}(m)$, as a function of the increasing number of pixels $m = \{1, 2, 3, \ldots, 36\}$, stacking them to conform to the arrangement in the previous step and starting with the pixel owning the smallest NSR_n

$$\mathrm{NSR}_{agg}(m) = \frac{\sqrt{\sum_{n=1}^{m} \left(\sigma^2_{F_{T_n}} + \sum_{k=1}^{N_C} \sigma^2_{F_{C_{n,k}}} + \sigma^2_{B_n} + \sigma^2_{D_n} + \sigma^2_{Q_n} \right)}}{\sum_{n=1}^{m} F_{T_n}}. \tag{4.29}$$

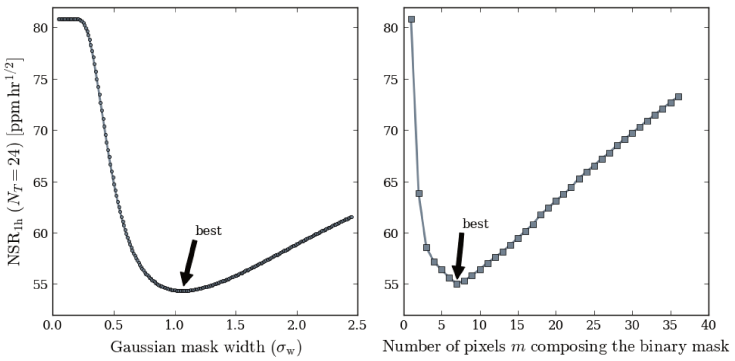
3. Define as the aperture the collection of pixels m providing minimum $\mathrm{NSR}_{agg}(m)$.

As the binary mask gets larger following the above routine, the NSR typically evolves as illustrated in Figure 42. Accordingly, the resulting best NSR binary mask for the input image example of Figure 41 is shown in Figure 43c.

4.9 Performance results

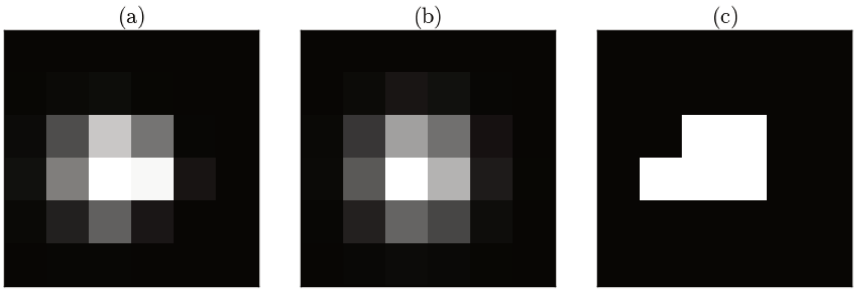
This section presents the photometric performance results of the three aperture models defined in subsection 4.8. These results are given in terms of NSR, SPR, N^{good}_{TCE}, and N^{bad}_{TCE}; and were obtained by applying each aperture model to all 50,000 input imagettes from subsection 3.7.

Figure 42 – Example of NSR evolution curve as a function of the increasing aperture size for a target star with $P = 11$. **Left**: Gaussian mask. **Right**: Binary mask.



Source: author.

Figure 43 – Aperture shapes computed as described in subsection 4.8, for the input image example of Figure 41. **Left (a)**: gradient mask. **Centre (b)**: Gaussian mask. **Right (c)**: binary mask.



Source: author.

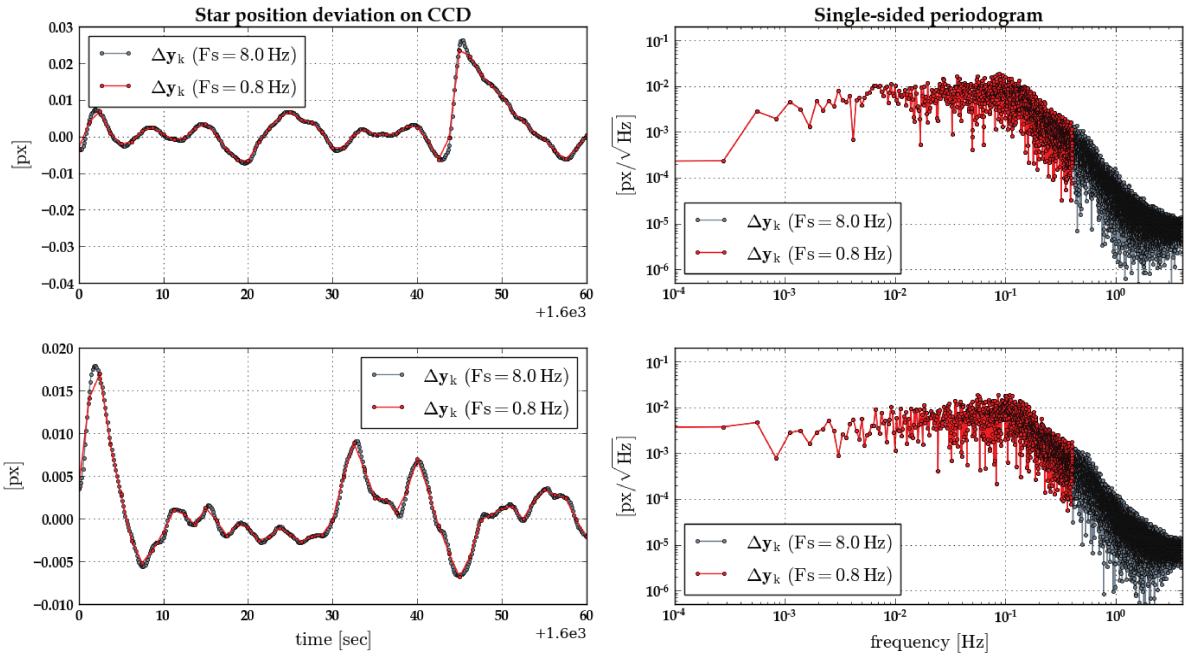
4.9.1 Photometric jitter noise

As early mentioned in subsection 3.7, in order to quantify the impact of spacecraft jitter in the noise performance of aperture photometry, this work adopts an approach that uses a set of shifted images to account for satellite motion. These shifted imagettes are computed target by target with respect to their respective reference frames (nominal positions). By applying this scheme, it was possible to quantify how much larger $\mathrm{NSR}_{1\mathrm{h}}^{jitter}$ (Equation 4.3) is in relation to $\mathrm{NSR}_{1\mathrm{h}}$ (Equation 4.4) under nominal and degraded scenarios of pointing performance. The results are presented in Table 10 and were obtained based on a set of 10,000 target stars. To increase simulation speed, the jitter time series of Figure 36 was down-sampled by a factor of 10, resulting in a 0.8 Hz signal keeping the same statistical properties (mean, variance, and spectral energy distribution) as the original signal (see comparison in Figure 44). Based on that, a total of 2,880 shifted images (1h duration signal) were produced per reference image, resulting in a total of $50,000 \times (2880 + 1) \sim 144 \times 10^6$

synthetic imagettes.

The results show that in nominal conditions the impact of jitter on the photometry is practically negligible ($\lesssim 0.5\%$), meaning that including jitter in the calculation scheme of the apertures would not only represent a complicated procedure, but also an useless effort in that particular case.

Figure 44 – Nominal jitter time-series at sampling frequencies of 8Hz and 0.8Hz.



Source: author.

Table 10 – Maximum noise-to-signal degradation at 95% confidence level as a function of satellite jitter amplitude, computed from a sample of 10,000 targets. Four scenarios are considered: nominal (Figure 36), three times (3×) nominal, five times (5×) nominal, and seven times (7×) nominal jitter.

Aperture model	Nominal	3× Nominal	5× Nominal	7× Nominal
Gradient	0.31%	2.9%	8.1%	16.2%
Gaussian	0.41%	3.6%	10.0%	19.43%
Binary	0.50%	4.7%	12.3%	23.2%

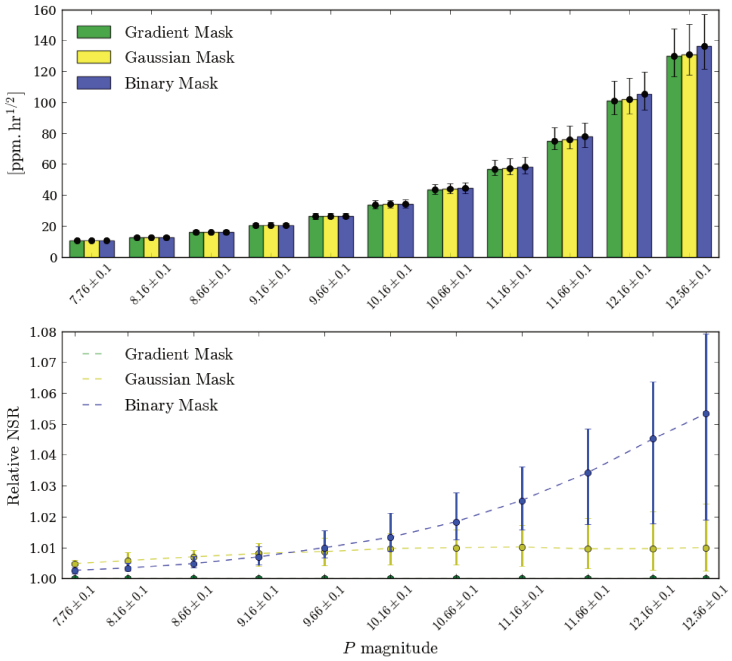
Source: author.

4.9.2 Noise-to-signal ratio

The performance parameter $\mathrm{NSR}_{1\mathrm{h}}^{jitter}$ was computed for our subset of input images assuming a nominal satellite jitter. The results are shown in Figure 45. Overall, the three

aperture models present comparable results for targets brighter than $P \sim 10.5$, with differences of less than 2% on average. The Gaussian mask has consistent suboptimal NSR performance over the entire P5 magnitude range, that is only $\sim 1\%$ higher on average than the gradient mask. The binary mask has better performance on average than the Gaussian mask for targets brighter than $P \sim 9$, but its performance degrades rapidly with increasing magnitude. For the faintest P5 targets, the binary mask presents NSR values about 6% higher on average and $\sim 8\%$ higher in the worst scenarios with respect to the gradient mask. Therefore, looking exclusively in terms of NSR, weighted masks are clearly the best choice.

Figure 45 – **Top**: Median values (black dots) of $\mathrm{NSR}_{1\mathrm{h}}^{jitter}$ ($N_T = 24$) as a function of target P magnitude and the applied mask model. **Bottom**: Relative $\mathrm{NSR}_{1\mathrm{h}}^{jitter}$, where the unit stands for the best NSR. In both plots, interval bars represent dispersions at 90% confidence level.



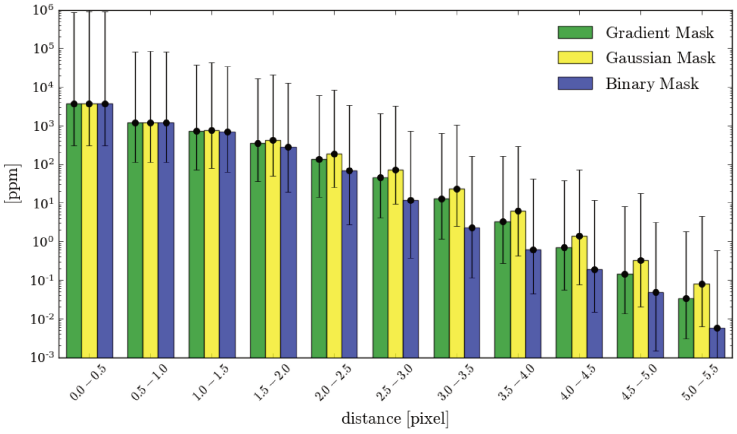
Source: author.

4.9.3 Stellar pollution ratio

It is presented in Figure 46 and Figure 47, respectively, the results of SPR_k and SPR_{tot}, respectively. The total SPR (SPR_{tot}) was computed for all 50,000 sources of our working subset of targets, while the per contaminant SPR (SPR_k) was computed for all $\sim$ 3.25 million stars located within a 10 pixel radius from those targets. Both plots show that the binary mask collects significantly less contaminant flux overall, and more particularly when the contaminant sources are located at more than 2 pixels distant from

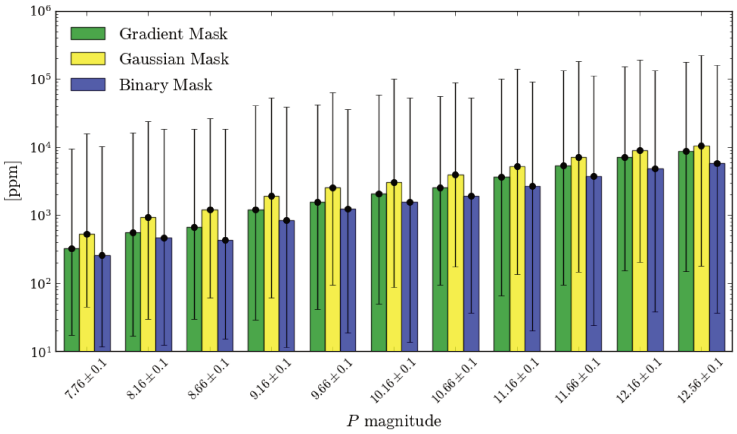
the targets. To give a rough idea, for about 80% of the contaminant sources SPR_k is at least three times greater for the weighted masks. This result is however not surprising because gradient and Gaussian masks are typically larger to best fit the shape of the PSF. This is the reason why they typically give lower NSR, as shown in the previous section.

Figure 46 – Median values (black dots) of SPR_k (Equation 4.7) as a function of the distance in pixels between the contaminants sources and their respective targets, and the applied mask model. Interval bars represent dispersions at 90% confidence level.



Source: author.

Figure 47 – Median values (black dots) of SPR_{tot} (Equation 4.8) as a function of target P magnitude and the applied mask model. Interval bars represent dispersions at 90% confidence level.



Source: author.

4.9.4 Detectability of planet transits

With both NSR and SPR determined, it is now possible to estimate the number N_{TCE}^{good} of target stars with sufficiently low NSR permitting the detection of eventual planets orbiting them. Table 11 and Table 12 show the values for N_{TCE}^{good} for the case of an Earth-like planet orbiting a Sun-like star, respectively, for the scenarios of SPR_{tot} as given by Figure 47 (no background correction) and $\mathrm{SPR}_{tot} = 0$ (perfect background correction).

The results show that the advantage of weighted masks regarding NSR performance, which is up to $\sim 7.5\%$ better with respect to the binary mask for the faintest and most numerous targets (see Figure 45), does not translate into a proportionally better sensitivity in detecting true planet transits. Indeed, the mask with lowest NSR, called the gradient mask, provides only $\sim$0.8% more chance of detecting Earth-like planets orbiting Sun-like stars at 1au. The difference between Gaussian and binary masks is even smaller, i.e. $\sim$0.4%. All three masks are equally capable of detecting Jupiter-like planets, no matter the number of telescopes observing the host star. To understand this, we need to compare Figure 45 and Figure 48. Taking the case of detecting Earth-like planets at about 1 au from Sun-like stars, the analyses show that the limiting magnitude[8] for aperture photometry is of the order of $P \sim 11.7$ ($V \sim 12$ @6,000K) at 7.1σ, $n_{tr} = 3$ and $N_T = 24$. Therefore, for most of the magnitude range ($11 \lesssim P \leqslant 12.66$) where the binary mask present the most degraded NSR performance with respect to the weighted masks, the latter do not provide any advantage in detecting such planets after all. Thus the small differences in N_{TCE}^{good} between binary and weighted masks, for the considered scenario, are consistent.

Correcting for the background results in an almost negligible impact ($\lesssim 0.6\%$ increase) in the overall sensitivity of the apertures in detecting true planet transits. Also, it has no significant impact in the comparative basis analysis between the different aperture models. It is important to stress however that inefficient background correction may significantly limit the accuracy with which planet transit depths can be determined.

Hence, from a planet transit finding perspective, designating an optimal solution for extracting photometry from the P5 stellar sample now becomes substantially less obvious. To this extent, looking at how each aperture performs in terms of false planet transit rejection may give us a hint about which is effectively the most appropriate choice.

4.9.5 Sensitivity to background false transits

The parameters SPR_k^{raw} (Equation 4.17) and SPR_k^{app} (Equation 4.22) are now compared with SPR_k^{crit} (Equation 4.19), to determine the number N_{TCE}^{bad} of contaminant stars with sufficiently high average flux to generate false positives. Two scenarios are considered herein: N_{TCE}^{bad} representing the number of contaminant sources for which $\mathrm{SPR}_k^{raw} \geqslant \mathrm{SPR}_k^{crit}$,

[8] This threshold is likely to be diminished by the presence of stellar activity in the noise (Gilliland *et al.* (2011)).

Table 11 – Number N_{TCE}^{good} of target stars for which $\eta \geqslant \eta_{min}$, as a function of the number N_T of telescopes observing them and the applied aperture model.

N_T	Gradient Mask	Gaussian Mask	Binary Mask
24	19,063 (38.1%)	18,674 (37.3%)	18,201 (36.4%)
18	15,105 (30.2%)	14,753 (29.5%)	14,469 (28.9%)
12	10,629 (21.3%)	10,368 (20.7%)	10,202 (20.4%)
6	5,528 (11.1%)	5,395 (10.8%)	5,357 (10.7%)
weighted	**10,067 (20.1%)**	**9,833 (19.7%)**	**9,667 (19.3%)**

Source: author.

Note: the values in this table corresponds to the case of an Earth-like planet with $\delta_p = 84$ ppm, $t_d = 13$h, $n_{tr} = 3$ and SPR_{tot} given by the simulated values presented in Figure 47 (i.e. assuming no background correction). They were determined from the dataset of 50,000 target stars. The weighted values correspond to the effective N_{TCE}^{good}, obtained by assuming uniform star distribution and a fractional field of view as given in Figure 47.

Table 12 – Same as Table 11, but for $\mathrm{SPR}_{tot} = 0$ (i.e. assuming a perfect background correction).

N_T	Gradient Mask	Gaussian Mask	Binary Mask
24	19,608 (39.2%)	19,319 (38.6%)	18,637 (37.3%)
18	15,510 (31.0%)	15,264 (30.5%)	14,806 (29.6%)
12	10,909 (21.8%)	10,701 (21.4%)	10,441 (20.9%)
6	5,625 (11.2%)	5,527 (11.1%)	5,456 (10.9%)
weighted	**10,318 (20.6%)**	**10,141 (20.3%)**	**9,884 (19.8%)**

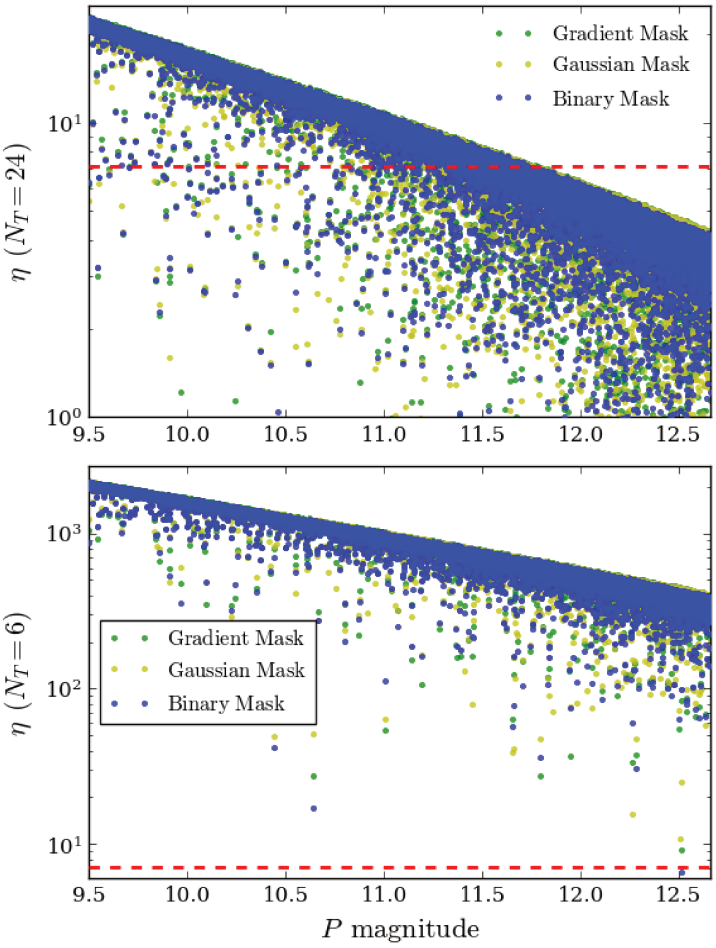
Source: author.

Note: a scatter plot of η, as a function of target P magnitude, is illustrated in Figure 48 for $N_T = 24$.

which supposes no background correction in the photometry; and N_{TCE}^{bad} representing the number of contaminant sources for which $\mathrm{SPR}_k^{app} \geqslant \mathrm{SPR}_k^{crit}$, which supposes a perfect background correction in the photometry. In both cases, we define SPR_k^{crit} with $\delta_{back,k} = 8.5\%$ (~ 0.1 mag), $t_d = 4$h, $\eta = 7.1$, and $n_{tr} = 3$. The chosen value for $\delta_{back,k}$ corresponds to the median depth (Figure 49) of the sources in the *Kepler* Eclipsing Binary Catalogue (Third Revision)[9], considering both primary (pdepth) and secondary (sdepth) depths together. The chosen value for t_d corresponds to the median transit duration (Figure 49) of the offset false positive sources listed in the Certified False Positive Table at NASA Exoplanet Archive. The transit duration values themselves were retrieved form the Threshold Crossing Events Table, by crossmatching the ID columns (KepID) from both tables. It is worthy

[9] <http://keplerebs.villanova.edu>

Figure 48 – Scatter plot of the statistical significance η (Equation 4.14) computed for 50,000 target stars, as a function of their respective P magnitude and the applied aperture model. The red dashed line represents the threshold $\eta_{min} = 7.1$. Values of N_{TCE}^{good} are provided in Table 11 Table 12. **Top**: Statistics for an Earth-like planet with $\delta_p = 84$ ppm, $t_d = 13$h, $n_{tr} = 3$, $\mathrm{SPR}_{tot} = 0$, and $N_T = 24$. **Bottom**: Statistics for an Jupiter-like planet with $\delta_p = 0.1$, $t_d = 29.6$h, $n_{tr} = 3$, $\mathrm{SPR}_{tot} = 0$, and $N_T = 6$.

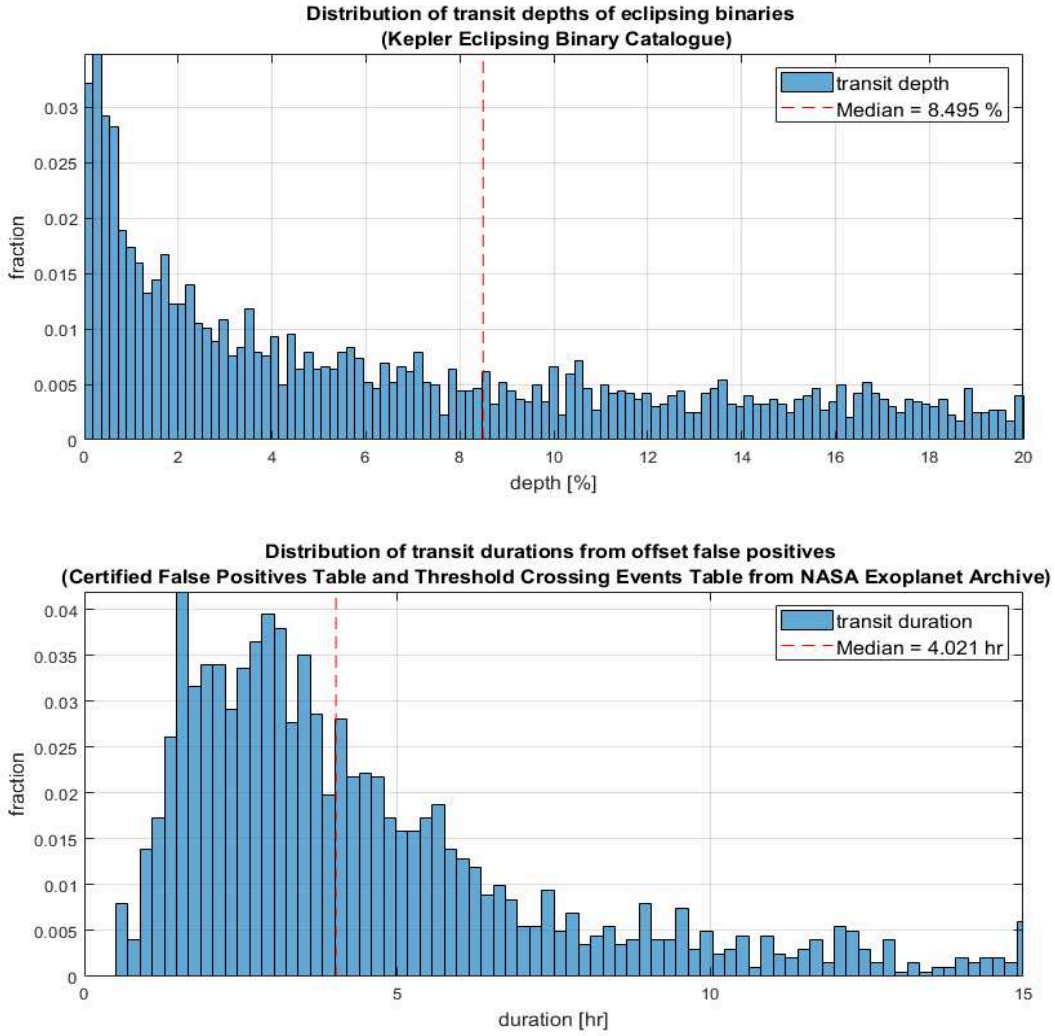


Source: author.

mentioning that the median values of $\delta_{back,k}$ and t_d adopted herein are also consistent with those determined from CoRoT observations (see Deleuil *et al.* (2018)).

Looking at the obtained results for N_{TCE}^{bad}, which are presented in Table 13 and Table 14, the important thing to notice at first glance is the fact that all tested aperture models have, fortunately, an intrinsically very low (less than 5%) overall sensitivity to detect mimicked planet transits caused by background eclipsing objects. In other words, these models are all insensitive to most of the potential false planet transits that may be produced by the contaminant sources in regions IV and VIII of Figure 50. This is

Figure 49 – Statistics on transit depth and transit duration of eclipsing binaries.



Source: *Kepler* Eclipsing Binary Catalogue (Third Revision) and NASA Exoplanet Archive.

surely mostly because of the high enclosure energy of PLATO PSFs, but the optimization scheme applied to each aperture model, privileging low NSR, is also key in this context. Nevertheless, the results also clearly show that compared to the binary mask employing weighted masks substantially increases the predicted occurrence of events mimicking planet transits. The Gaussian mask is expected to deliver up to $\sim 40\%$ higher N_{TCE}^{bad} than the binary mask, which is notably a huge discrepancy. The differences between gradient and binary masks are smaller, but still very significant: N_{TCE}^{bad} is up to $\sim 20\%$ higher for the gradient mask. Either correcting for the background or not, these differences rest roughly the same, so background correction has no significant impact in the comparative basis analysis between the different aperture models. In absolute terms though, the results

indicate that fully removing the background leads to an overall increase of more than 10% in N_{TCE}^{bad}, which is consistent with the analysis presented in subsection 4.7.

Overall, the obtained results for N_{TCE}^{bad}, in comparison to those of N_{TCE}^{good} presented in the previous section, makes the scenario of choosing weighted masks become highly unfavoured even though that kind of mask provides better overall performance in terms of NSR. Still, it would be legitimate to ask whether the obtained values for N_{TCE}^{bad} are indeed significant in an absolute sense, since they represent less than 5% of our full set of contaminant stars composed of ∼3.25 million sources. Properly answering this question requires carefully modelling the parameters $\delta_{back,k}$ and t_d for the PLATO target fields, which is though beyond the scope of this work. However, it is possible to obtain a rough idea of the occurrence of EBs (N_{beb}) that could potentially result from the weighted values shown in Table 13 and Table 14. First, we need to consider that these values refer to about 20% of the minimum number of expected targets for the P5 sample. Second, we may assume that the frequency of EBs (F_{eb}) for the PLATO mission might be of the order of 1%[10]. Accordingly, the expected occurrence of EBs at 7.1σ, for the P5 sample could be approximately estimated with $N_{\mathrm{beb}} \sim 5 \times N_{\mathrm{TCE}}^{bad} \times 1\%$. From the weighted values presented in Table 13 and Table 14, that gives $1,600 \lesssim N_{\mathrm{beb}} \lesssim 2,500$ (all three tested aperture models comprised). This allows us to conclude that N_{TCE}^{bad} is thus not negligible. Moreover, considering that the total number of targets in the P5 sample is comparable to the total number of observed targets by the *Kepler* mission, we verified that our approximative estimate on the expected N_{beb}, for the P5 sample, is very consistent to the statistics of background false positives of the *Kepler* mission. Indeed, the Certified False Positive Table on the NASA exoplanet archive gives at the present date 1,287 offset false positives out of 9,564 *Kepler* objects of interest. Such a consistency attests that our study is satisfactorily realistic. We stress however that accurate false positive estimates for the P5 sample cannot be provided by our study alone, in particular because it needs to be consolidated with PLATO's science exoplanet pipeline.

As a complement to the results presented in this section, Figure 51 shows, for each aperture model, a two-dimensional histogram containing the distribution of contaminant stars having $\mathrm{SPR}_k^{app} \geqslant \mathrm{SPR}_k^{crit}$, as a function of the differential P magnitude and the Euclidean distance between these sources and the corresponding targets. The parameters used to calculate SPR_k^{crit} were $\delta_{back,k} = 0.8$ mag, $t_d = 4$h, $N_T = 24$, $\eta = 7.1$, and $n_{tr} = 3$. This plot is of particular interest since it illustrates that the contaminant stars having sufficiently high average flux to produce background false positives are typically less than ∼ 10 mag fainter and located at less than ∼ 4 pixels away from the targets. Consequently, from the point of view of the distances, we verified that our approach of considering

[10] Fressin *et al.* (2013) give $F_{\mathrm{eb}} = 0.79\%$ for the *Kepler* mission. They defined it as being the fraction of EBs found by *Kepler*, including detached, semi-detached, and unclassified systems, divided by the number of *Kepler* targets.

contaminant sources located at up to 10 pixels distant from the targets was largely enough for the purposes of this work. From the point of view of the differential magnitude, three important aspects need to be considered when interpreting the results.

First, we note that stars in our input catalogue are limited in magnitude to $P \sim 21.1$. This means that for the faintest (and most numerous) P5 targets, for which P magnitude is as high as 12.66, the maximum differential magnitude from their contaminants is therefore as small as $21.1 - 12.66 = 8.44$ mag, i.e. smaller than the limit of ~ 10 mag suggested by the histograms. In contrast, P5 has targets as bright as 7.66 mag, so that the differential magnitude may be as high as $21.1 - 7.66 = 13.44$ mag. Hence, well above that limit.

Second, Figure 51 suggests an absence of stars at distances near zero, in particular at differential magnitudes above 5 mag. Such an anomaly is understood to be related to what was already pointed out in subsection 3.6 concerning bad estimates of the fluxes of stars fainter than $G \sim 17$ in the DR2 catalogue. This issue is reported in Evans *et al.* (2018) and assumed to be caused by factors such as poor background estimation, observation taken in the proximity of bright sources, binarity, and crowding. In these conditions, the capability to isolate stars is therefore compromised. Taking into account that the most problematic cases were removed from the DR2 release according to the authors, the lack of stars in the above mentioned areas of Figure 51 is justified. Yet scenarios of differential magnitude higher than 10 mag, at the same time that $\mathrm{SPR}_k^{app} \geqslant \mathrm{SPR}_k^{crit}$, should mostly occur at distances shorter than ~ 0.5 pixel, where the occurrence of contaminant stars is substantially smaller than that at longer distances (see Figure 40).

Third, the parameters used to build the histograms of Figure 51 correspond in practice to a near worst case scenario in terms of the expected occurrences of false transits caused by background eclipsing objects. Indeed, it considers photometry perfectly corrected for the background; contaminants stars being observed by 24 cameras (maximum sensitivity to transit signatures); and contaminant stars generating background transit depths of 0.8 mag ($\sim 52\%$), which is significantly high for a binary system. This means that the maximum differential magnitude is typically much smaller than 10 mag.

Taking into account all the above considerations, it is possible to state that Figure 51 gives a sufficiently realistic and unbiased representation of distances and differential magnitudes of contaminant stars that are likely to cause background false planet transits, regardless of the limitation in maximum magnitude of the adopted input catalogue. Furthermore, it is important to clarify that the missing fraction ($\sim$0.01%) of PSF energy in the images of Figure 21 entails no significant impact in our analysis. This small fractional energy may be non-negligible uniquely in cases in which the differential magnitude between target and contaminant stars is $\lesssim -4$ mag. These are however extremely rare scenarios in our input stellar field, and thus statistically insignificant to our analysis. Indeed, less than 0.5% of the contaminant sources in Figure 51 have differential magnitude smaller than

−2.6 mag.

Ultimately, the unique set of contaminant stars, from all three histograms presented in Figure 51, was extracted to build a histogram of the fractional distribution of contaminant stars having $\mathrm{SPR}_k^{app} \geqslant \mathrm{SPR}_k^{crit}$ (i.e. the fractional distribution of N_{TCE}^{bad}) as a function of Galactic latitude. The resulting plot is shown in Figure 52. It suggests that the occurrences of false positives caused by background eclipsing stars might increase exponentially towards the Galactic plane, which is consistent with the distributions of offset transit signals presented in Bryson *et al.* (2013). It is worthy mentioning that since the distribution of stars within IF privileges certain latitudes, because of its circular shape, such bias is prevented from being propagated to Figure 52 by considering contaminant sources within a sufficiently narrow Galactic longitude range $\mathrm{l}_{\mathrm{LoS}} \pm 1.5\,[\mathrm{deg}]$ (see Table 6).

Table 13 – Number N_{TCE}^{bad} of contaminant stars for which $\mathrm{SPR}_k^{raw} \geqslant \mathrm{SPR}_k^{crit}$, which supposes photometry with no background correction, as a function of the number N_T of telescopes observing the host star and the aperture model. The roman numerals correspond to the areas indicated in Figure 50. The percentiles indicate the amount of deviation of the values from weighted masks with respect to those from binary mask. The weighted values in the lower row correspond to the effective N_{TCE}^{bad}, obtained by assuming uniform star distribution and a fractional field of view as given in Table 2.

N_T	Binary Mask (I + II) = (V + VI)	Gradient Mask (I + III)	Gaussian Mask (V + VII)
24	40,135	48,005	55,520
18	36,835	43,690	50,785
12	32,830	38,485	44,565
6	26,545	31,050	35,995
weighted	**31,591**	**37,178 (+17.7%)**	**43,073 (+36.3%)**
N_{BEB}	1,580	1,859	2,154

Source: author.

Note: The presented values were determined from our dataset of ∼ 3.25 million contaminant stars. The SPR_k^{crit} was computed with $\delta_{back,k} \sim 0.1$ mag, $t_d = 4$h, $\eta = 7.1$, and $n_{tr} = 3$.

4.9.6 Long-term drift

As explained earlier in subsubsection 2.9.2, the pixels of PLATO detectors are relatively broad compared to the size of the PSFs. During observations, this will cause aperture photometry to be quite sensitive to the long-term star position drift occurring on the focal plane. To verify the corresponding impact on performance, this subsection presents the results of a long-term star position drift simulation applied to a given target star of IF that has $P = 11$ (see corresponding PSF and imagette in Figure 53).

Table 14 – Same as Table 13, but now representing the contaminant stars for which $\mathrm{SPR}_k^{app} \geqslant \mathrm{SPR}_k^{crit}$, which supposes photometry with perfect background correction.

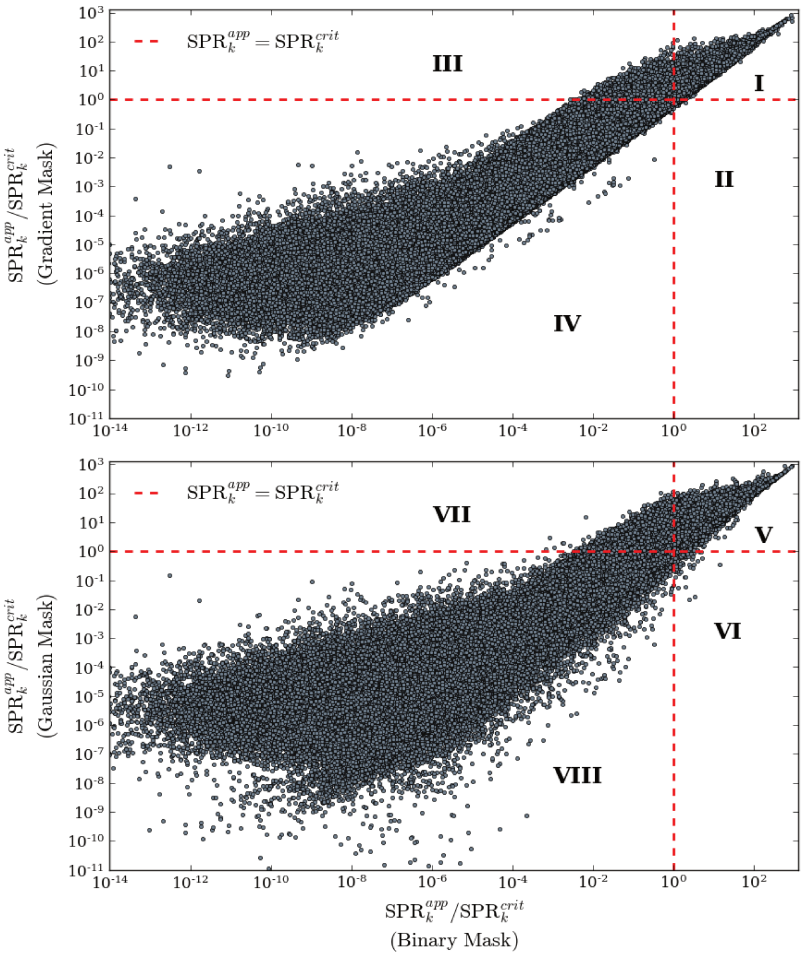
N_T	Binary Mask (I + II) = (V + VI)	Gradient Mask (I + III)	Gaussian Mask (V + VII)
24	45,180	54,720	63,555
18	41,575	50,185	58,540
12	37,055	44,380	51,885
6	30,185	35,820	41,570
weighted	**35,731**	**42,774 (+19.7%)**	**49,814 (+39.4%)**
N_{BEB}	1,787	2,139	2,491

Source: author.

To apply such a simulation, the first step consisted in determining how the parameters flux, NSR, and SPR of the photometry vary as a function of the intrapixel location of the target's barycentre. As pointed out in subsection 3.7, by convention the **x** and **y** CCD coordinates of the target barycentre are (each) located at no more than an absolute euclidean distance of 0.5 pixel from the imagette centre. As a consequence, there is a well defined region of the imagette within which the target barycentre can be located. This region is delimited by the white solid square represented in Figure 53. The three parameters (flux, NSR, and SPR) were therefore computed within this region for all intrapixel positions defined by a uniform grid of 1/100 pixel resolution on both **x** and **y** axis. It is important to mention that an exclusive aperture was calculated, according to the computation schemes described in subsection 4.8, at each individual intrapixel position in this grid. By doing so, the 2D maps of Figure 54 were obtained. These maps clearly show that both flux and SPR parameters may significantly vary depending on the target barycentre location within the area delimited by the white square, no matter the aperture model considered. In terms of NSR, all three aperture models provide stable performance throughout the different intrapixel positions. Readers might notice that, by symmetry, the pattern given by each map repeats itself beyond the square area.

The second step of the simulation consisted in drawing a light curve from the data of the flux maps, in order to compare the photometric flux evolution over the time among the different aperture models. In principle, stars can move following any trajectory across the CCDs. For the sake of simplicity, it is assumed herein an arbitrary motion of the target star along $\mathbf{y} = 0.15$ [px] of the flux maps of Figure 54, starting at $\mathbf{x} = -0.50$ [px] and going all the way to $\mathbf{x} = +0.50$ [px], which corresponds to a time scale of about three months (90 days). This resulted in the light curves presented in Figure 55. Although the chosen trajectory is not perfectly realistic, since stars will not move perfectly straight from one point to another, this is still well representative of the flux variations that are likely

Figure 50 – Scatter plot of SPR_k^{app} normalized by SPR_k^{crit}, computed for ~ 3.25 million contaminant stars. This illustration represents the particular case where SPR_k^{crit} is computed with $\delta_{back,k} \sim 0.1$ mag; $N_T = 12$; $n_{tr} = 3$; $t_d = 4$h. Values of N_{TCE}^{bad} are provided in Table 13 and Table 14. **Top**: Comparison between the values given by the gradient mask (vertical axis) and by the binary mask (horizontal axis). Region I: both masks exceed SPR_k^{crit}. Region II: only the binary mask exceeds SPR_k^{crit}. Region III: only the gradient mask exceeds SPR_k^{crit}. Region IV: no mask exceeds SPR_k^{crit}. **Bottom**: Comparison between the values given by the Gaussian mask (vertical axis) and the binary mask (horizontal axis). Regions V to VIII are analogous to I, II, III, and IV, respectively.

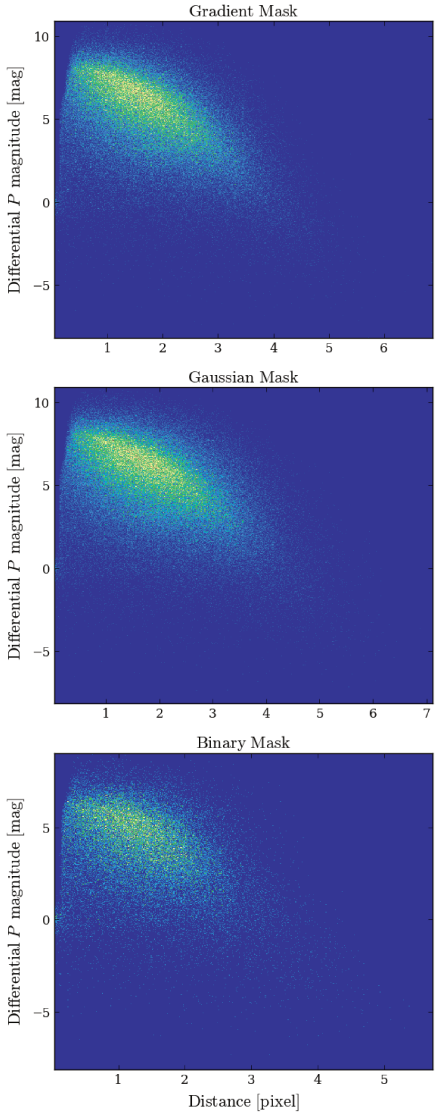


Source: author.

to manifest because of the long-term drift.

Observing the results presented in Figure 55, it is noted that the binary mask delivered a piecewise relatively stable light curve, but with eventual strong discontinuities. In fact, looking at the ensemble of 100 binary masks (one per sample) of the light curve, it was noticed that the discontinuities only occurred when the shape of the binary mask

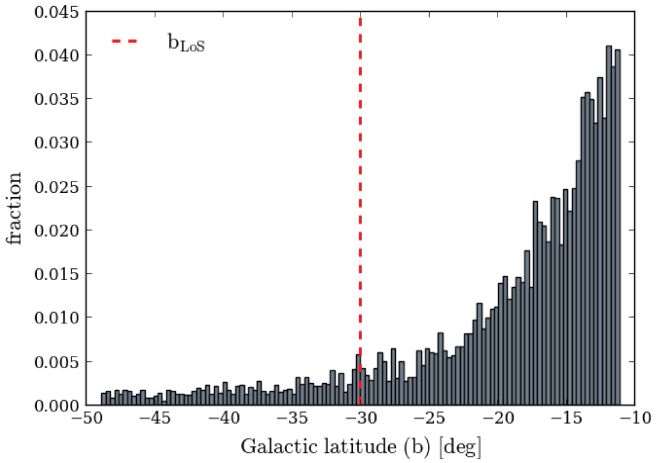
Figure 51 – Two-dimensional histograms of the distribution of contaminant stars with $\mathrm{SPR}_k^{app} \geqslant \mathrm{SPR}_k^{crit}$, for gradient (top), Gaussian (centre), and binary (bottom) masks. The vertical axis indicates the differential P magnitude between the contaminants and their respective targets, whereas the horizontal axis indicates the corresponding Euclidean distances. The parameters used to calculate SPR_k^{crit} are $\delta_{back,k} = 0.8$ mag, $N_T = 24$, $n_{tr} = 3$, and $t_d = 4$h.



Source: author.

changed from one sample to another. In other words, the regular parts of the light curve have the exactly same binary mask shape giving the best NSR at the corresponding samples. Indeed, in the example of Figure 55 there are eight unique binary mask shapes and seven effective mask updates (i.e. changes in shape). Concerning the weighted masks, the gradient mask presented a light curve with a prominent long-term drift but with remarkable

Figure 52 – Fractional distribution of N_{TCE}^{bad}, as a function of Galactic latitude (all three mask models comprised). This histogram (c.f. Bryson *et al.* (2013)) was built with contaminant sources that have Galactic longitude within the range $\mathrm{l_{LoS}} \pm 1.5\,[\mathrm{deg}]$. The red vertical line indicates the Galactic latitude $\mathrm{b_{LoS}}$ of $\mathrm{IF_{LoS}}$ (see IF coordinates in Table 6).



Source: author.

continuity, as only one (removable[11]) discontinuity occurred along the considered star motion's trajectory. The Gaussian mask delivered stronger long-term drift levels compared to that of the gradient mask, and several discontinuities that appeared at roughly regular intervals. Comparing both noiseless and noisy versions of the light curves from the Gaussian mask presented in Figure 55, it is noted that the amplitude of the discontinuities is of the order of the noise level for this particular test (target star with $P = 11$).

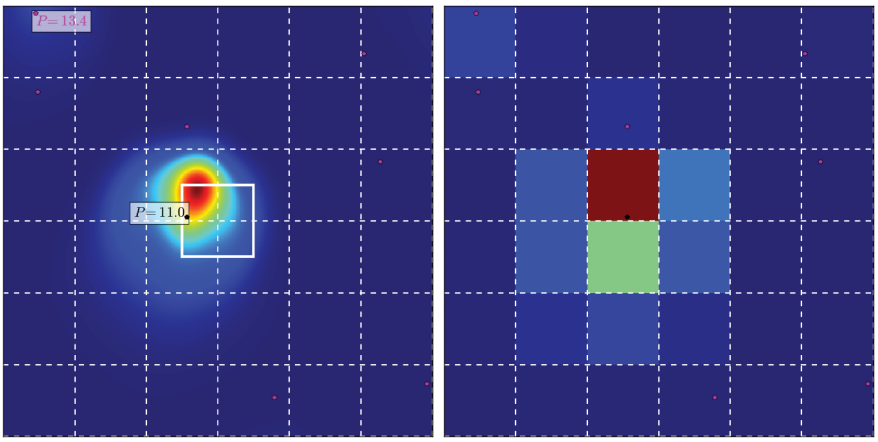
4.10 Discussions

4.10.1 Sensitivity to detect true and false planet transits

In this chapter, a novel approach for determining optimal photometric apertures was presented. The motivation behind developing this approach raised from the fact that most of the targets from the P5 sample will not have pixel data (images) available on ground for detecting false planet transits, which potentially limits the scientific exploitability of the light curves produced (on board) for these stars. This approach was founded in the hypothesis that choosing apertures following the common sense adopted in the literature, i.e. exclusively based on how well an aperture can detect legitimate planet transits, without properly taking into account the impact of false positives, is not necessarily the best strategy. Hence, the key objective in the development presented in this chapter was to verify whether apertures providing overall best NSR, that is best overall efficiency in detecting true

[11] <http://mathworld.wolfram.com/RemovableDiscontinuity.html>

Figure 53 – Input image used for simulating long-term star position drift. The target star in the frame has $P = 11$ (barycentre designated by the black dot). **Left**: high resolution PSF. The solid white square delimits the area within which the target barycentre can be located (see subsection 3.7). White dashed lines represents pixel borders. **Right**: corresponding low resolution PSF (imagette).

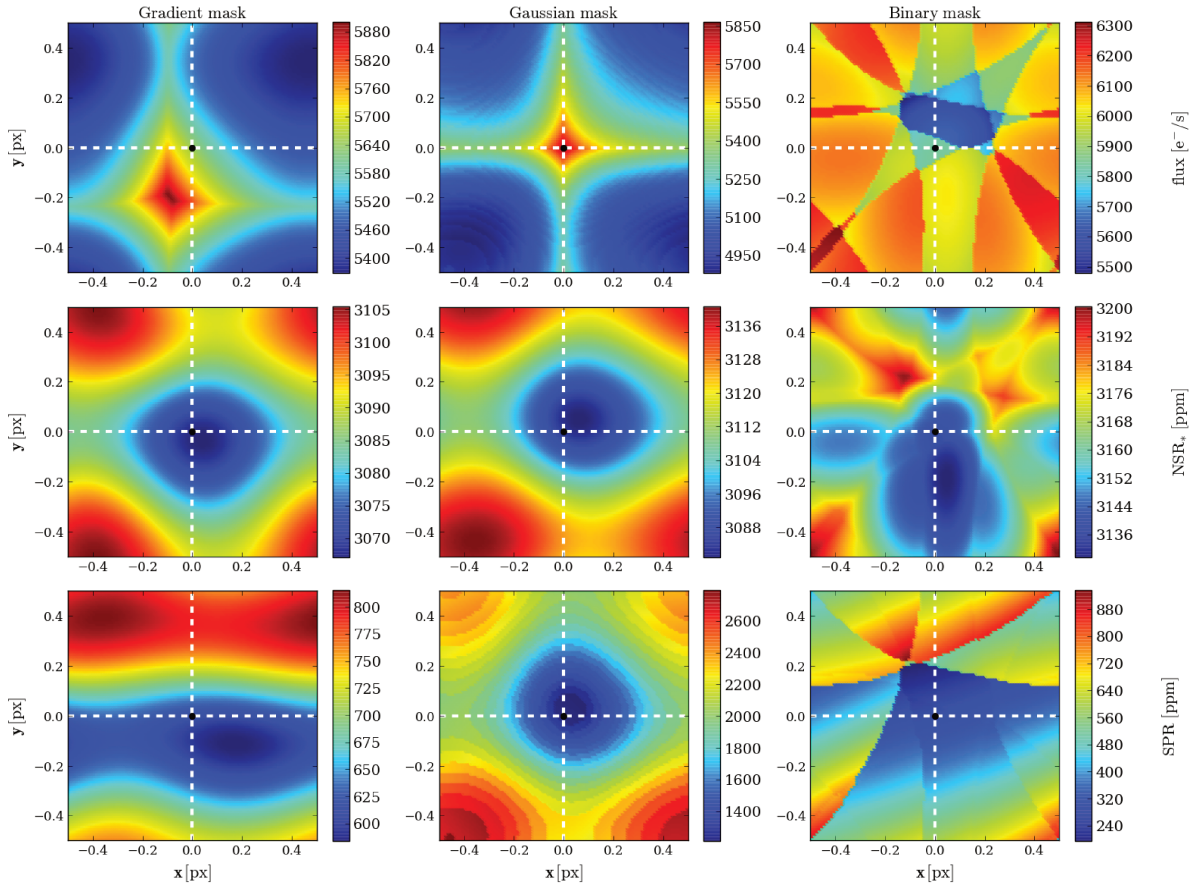


Source: author.

planet transits, also present satisfactory performance in terms of sensitivity in detecting background false planet transits. In this context, the two scientific metrics introduced in this work, N_{TCE}^{good} and N_{TCE}^{bad}, allowed us establishing an objective comparison between the sensitivity of an aperture in detecting true and false planet transits, respectively. With these two metrics, it was verified that the ideal aperture – i.e. that which simultaneously maximizes N_{TCE}^{good} and minimizes N_{TCE}^{bad} – cannot be obtained, since both metrics are conflicting: the former requires maximizing the NSR, whereas the latter requires minimizing the NSR. As a consequence, in order to find an optimal mask, one needs to find a compromise between these two aspects.

Both N_{TCE}^{good} and N_{TCE}^{bad} metrics were computed based on a large and realistic stellar population containing millions of stars from the *Gaia* catalogue; and for three distinct aperture models: a weighted mask providing global minimum noise-to-signal ratio (gradient mask); a weighted Gaussian mask giving sub-optimal noise-to-signal ratio; and a narrower binary mask to reduce the impact of contamination (crowing). As a result, this chapter also provided an extensive characterization of the photometric performance expected from the P5 targets, notably in terms of noise-to-signal ratio (NSR) and contamination (SPR). In Appendix A, a photometric performance breakdown (Table 15) of the P5 stellar sample is provided containing the individual noise contributions from each of the noise sources listed in Table 9.

Figure 54 – 2D maps of aperture flux (top), NSR (centre), and SPR (bottom) as a function of intrapixel location of target barycentre. The black dots in the centre of the maps represent the common corner of four adjacent pixels (pixel corner). The four edges of each map represent the centres of these pixels (pixel centre). All maps have spatial resolution of 1/100 pixel on both **x** and **y** axis. **Left**: gradient mask. **Centre**: Gaussian mask. **Right**: binary mask.

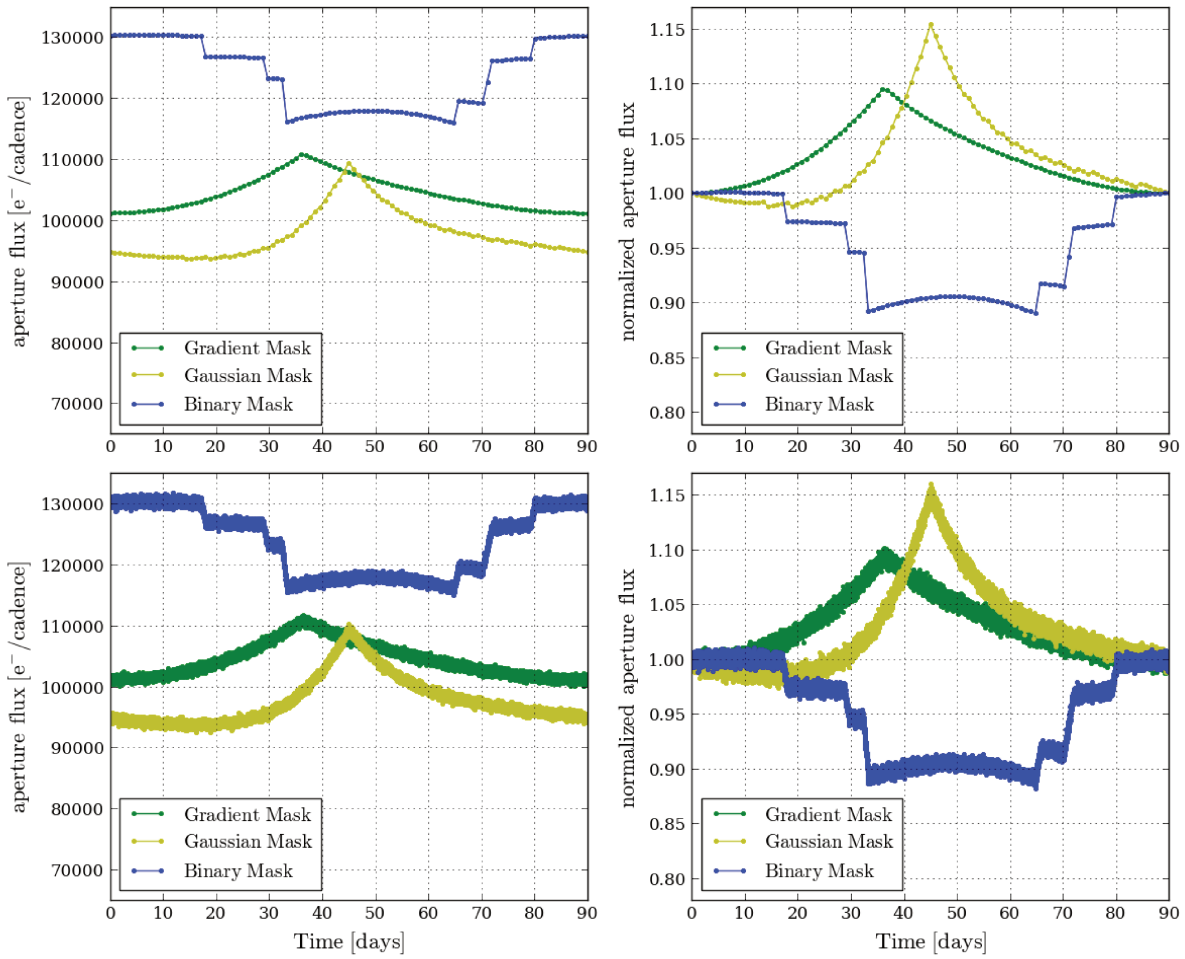


Source: author.

4.10.2 Sensitivity to long-term star position drift

To investigate how the non-homogenous energy distribution of PLATO PSFs (Figure 22) – as a function of intrapixel star position – impacts on photometry, 2D maps (Figure 54) of the photometric parameters flux, NSR and SPR were created. To built these maps, the three aperture models described in subsection 4.8 were calculated at the points defined by a uniform square grid centred at the common corner of four adjacent pixels. This square grid has 1-pixel size, 1/100 pixel resolution, and covers all the possible locations of a star barycentre within a pixel, following the convention – defined in subsection 3.7 – that the pixels of an imagette are selected such that the centre of the resulting imagette is located at no more than an absolute euclidean distance of 0.5 pixel from any of the x and

Figure 55 – Example of photometric flux variation expected from a star with $P \sim 11$, caused by long-term drift. **Top**: Average flux only, representing the flux variation occurring when the target moves along $\mathbf{y} = 0.15$ in Figure 54. The left panel shows fluxes in units of electron per exposure [e^-/cadence], while the right panel shows the corresponding normalized fluxes. **Bottom**: Curves of the top panel linearly interpolated, such as to have samples every 600 seconds and with added noise (photon, background, detector, and quantization). Noise amplitude corresponds to that observed in a single telescope.



Source: author.

y coordinates of the target barycentre. From the obtained results of long-term drift, it was possible to verify that mask-target assignments performed during each calibration phase become, soon or later, no longer optimal, since the flux distributions of the targets significantly change as these move across the pixels.

4.10.3 Sensitivity to satellite jitter

This chapter also presented some results showing at which extent the spacecraft jitter might impact on the noise performance of apertures photometry. These results allowed verifying that the effect of jitter does not required being taken into account in the calculation schemes (subsection 4.8) used to compute the apertures, since the noise performance degradation caused by the pointing error is almost negligible at nominal conditions. In contrast, in the eventual scenario of degraded pointing performance, the jitter contribution to the total noise might become not negligible at all, which could then justify taking it into consideration on the computation scheme that defines the aperture shape.

4.10.4 Detector saturation

It is important to clarify that the results presented in this chapter are valid for non-saturated stars. It implies that only one mask is attributed to each target, and each mask is limited in size by the (6×6) shape of an imagette. This is a fundamental assumption for the study presented herein. In the current instrument design, PLATO detectors are expected to exhibit saturation at pixels observing stars brighter than $P \sim 8.16 \pm 0.5$ (i.e. $V \sim 8.5 \pm 0.5$ @6,000K) after a 25s exposure (normal cameras). The exact saturation limit depends on the location of the star in the CCD and where its barycentre falls within a pixel. The brightest stars used in this study are thus at the very lower bound of this broad saturation threshold.

5 Conclusions and perspectives

The ESA PLATO space mission is devoted to unveiling and characterizing new extrasolar planets and their host stars. Compared to other missions of the same category, PLATO has unique characteristics that makes it much more than a simple mission dedicated to provide a large statistical sample of planetary systems. Indeed, PLATO was conceived to not only detect, but also to precisely characterize – in terms of radii, mass, mean density, age, atmosphere etc. – tens of small and low-mass planets orbiting within the habitable zone of bright Sun-like stars. With the observational accuracy envisaged for the mission, for example, it is expected a decrease on the uncertainties of planetary density from the current 30 to 50% to about 10% only (ESA (2017)). Such an improvement shall represent a compelling step forward towards the exploration of planetary diversity. Furthermore, the PLATO mission is concretely in position to confirm whether planets like the Earth indeed exist, including how many of them and what type of star they orbit.

PLATO satellite will encompass a very large (>2,100 deg^2) field of view, granting it the potential to survey up to one million stars depending on the final observation strategy. The telemetry budget of the spacecraft cannot handle transmitting individual images for such a huge stellar sample at the right cadence, so the development of an appropriate strategy to perform on-board data reduction is mandatory. To achieve it, mask-based (aperture) photometry was selected as the method to produce stellar light curves in flight because of its excellent compromise between scientific performance and implementation complexity. That is, such a method delivers noise performance comparable to that of more complex solutions such as PSF fitting, but runs much faster and consumes much less computational resources.

Aiming to maximize the scientific exploitability of the reduced data produced on-board of the PLATO spacecraft, the present research work has primarily focused on finding the optimal photometric apertures for extracting photometry of more than 250,000 PLATO targets (the P5 stellar sample). Accordingly, the development of this work considered three distinct aperture models: a weighted mask providing global minimum noise-to-signal ratio (gradient mask), obtained through a novel direct calculation method; a weighted Gaussian mask giving sub-optimal noise-to-signal ratio; and a narrower binary mask to reduce the impact of contamination. Taking into account the massive number of targets in the P5 sample, a statistical approach was adopted for characterizing the photometric performance of all three aperture models: each one was tested on tenths of thousands of synthetic images (imagettes) containing realistic stellar distribution, which overall comprises 50,000 potential PLATO targets surrounded by $\sim$ 3.25 million neighbouring contaminant stars. The stellar population used for building the imagettes was extracted from the *Gaia* DR2

catalogue and is representative of the provisional long-pointing southern PLATO field (SPF).

As a result, this work represents a realistic, comprehensive and unique characterization of the photometric performance expected from PLATO targets, in fact as never done before for any other mission of the same category. Overall, the ensemble of results and discussions derived from this work constitutes a significant contribution to the design of both on-board and on-ground science data processing pipelines of the PLATO mission.

5.1 Conclusions

5.1.1 Background light

Characterizing the intensities of background light is key for obtaining realistic estimates on the NSR of stellar photometry, in particular for faint stars. Yet even more crucial than that, accurate knowledge on the level of background light is mandatory for obtaining reliable estimates on the true planetary transit depths (and consequently on planetary radius) from the diluted observed depths (Equation 4.10). This becomes particularly relevant in the context of PLATO, taking into account that this mission is being conceived to not only detect, but also to accurately determine planetary radius at down to 3% accuracy.

Within PLATO's Lagrange-L2 orbit, scattered sky background light will be dominated by the zodiacal light. In subsection 3.1, an expression was provided for estimating the per camera intensities of zodiacal light on PLATO detectors. From that, it was possible to characterize, both spatially (across PLATO SPF) and temporally (over the course of a one-year orbit), the expected level of zodiacal light for the PLATO mission. The obtained results show that the spatial gradient of zodiacal light is expected to be considerably strong (up to $\sim 260\%$; Figure 32) between the minimum and maximum ecliptic latitudes of SPF, which is not surprising considering its large sky coverage ($\sim 2132\,\mathrm{deg}^2$). In addition, the results show that the temporal gradient is also expected to be quite significant in this field: up to $\sim 160\%$ variation over a time scale of three months for stars located at the lowest ecliptic latitudes; and up to $\sim 40\%$ variation over the same time scale for stars located around $\mathrm{IF}_{\mathrm{LoS}}$. Besides the zodiacal light, another important source of background light impacting the accuracy at which planetary radius can be determined is the flux originating from contaminant stars. The fractional flux contribution of an individual contaminant star, SPR_k (Figure 46), might be particularly significant when such a star is located at less than ~ 2.5 pixels from a target. The total fractional flux contribution from several contaminant stars, SPR_{tot} (Figure 47), is more likely in a statistical sense to cause significant transit dilution on faint ($P \gtrsim 12$) stars.

It is unambiguous to conclude, therefore, that important variations in background

light will occur over the course of the PLATO mission. As a consequence, these variations will unavoidably cause meaningful discrepancies between the measured depths of distinct transit events produced by the same planet, notably in cases when the corresponding orbital period is of the order of a few months. Moreover, background changes over the time create spurious signatures in the power spectrum of light curves, which limit the accuracy at which stellar oscillation frequencies can be determined. In the long run hence, a proper follow-up of the background light will need to be performed throughout the observations, so that light curves can be properly corrected for the (average) background.

5.1.2 Long-term drift

From the results presented in subsubsection 4.9.6, it is possible to verify that the long-term star position drift is expected to impact the photometry of PLATO targets in quite different ways depending on the considered aperture model. The gradient mask delivers the smallest number of discontinuities in the light curves, which translates into a better stability of the photometric flux. The Gaussian mask generates quite numerous (> 20) yet very small ($< 1\%$) jump discontinuities in the photometry. The binary mask is the method producing the largest ($\sim 3-10\,\%$) jump discontinuities, but these occur much less frequently than those produced by the Gaussian mask. Indeed, typically only a few (~ 5) events are expected to be produced by the binary mask over a time scale of three months. Before entering in any kind of comparison between the results obtained from each mask model, a few aspects need to be considered first as discussed in the next paragraph.

Correcting for discontinuities in light curves, including those produced by sudden pixel sensitivity dropout (SPSD) (Aigrain *et al.* (2017)), requires them to be properly detectable and characterizable. In a broad sense, both detection and characterization steps consist of identifying respectively the existence and the nature of the discontinuity. From the point of view of the mask update problematic in PLATO (Samadi *et al.* (2019)), characterizing its discontinuities should be in theory a relatively simple task, considering that the timestamps of all updates will be known on ground, as well as the PSF that is used to correct them. In contrast, the ease in detecting unexpected discontinuities depends basically on their aspect and amplitude, which is quite distinct for each aperture model as seen in the results illustrated in Figure 55.

From the above considerations, it is possible to state that, in terms of flux, the gradient mask is the method providing the best results in terms of flux stability, since this mask delivers the least number of discontinuities and their type (removable discontinuity) make them relatively easy to identify. The Gaussian mask presents several jump discontinuities which are typically small enough for being of the same order of the noise amplitude in the light curves. As a consequence, these discontinuities might be particularly difficult to identify and to correct from an algorithmic point of view. The Gaussian mask also present

removable discontinuities which are sharper – and thus easier to identify – than those produced by the gradient mask. The binary mask produces jump discontinuities which are typically strong and hence not difficult to identify. In a direct comparison between Gaussian and binary masks therefore, the discontinuities produced by the latter should be more efficiently corrected overall, notably because these can be more easily identified.

Since the masks were calculated to maximize the NSR at each intrapixel position of the grid, the NSR does not vary significantly across the pixel for any of the three masks. The 2D maps (Figure 54) show however significantly variations of the SPR parameter, in particular for the Gaussian and the binary masks. The fact that the gradient mask presents significantly less variations of the SPR across the pixel can be explained by its highly adaptable weights to the level of signal and contamination at each pixel individually. In contrast, the weights of the Gaussian mask are limited to the Gaussian function, and the binary mask weights are either zero or one, thereby explaining why these two masks cannot deliver the same stability of the SPR parameter as that of the gradient mask. Overall, the SPR sensitivity to intrapixel location of target stars reinforces the statement of the previous section that points for the need of establishing a proper follow-up of the background light to account for event-varying transit depths caused by instrumental (i.e. non-astrophysical) effects.

5.1.3 Optimal aperture

To determine the optimal aperture model for extracting photometry from the P5 targets, an innovative criterion was proposed based on two new science metrics: simulated number of target stars for which a planet orbiting it would be detected, denoted as N_{TCE}^{good} (to be maximized); and simulated number of contaminant stars that are sufficiently bright to generate false positives when eclipsed, denoted as N_{TCE}^{bad} (to be minimized). Both metrics depend on noise-to-signal ratio, stellar pollution ratio and simulated frequency of threshold crossing events at 7.1σ; they allow us to direct evaluate the sensitivity of apertures in detecting true and false planet transit signatures. Accordingly, we designated as optimal the model providing the best compromise between sensitivity to detect true and false planet transits. Such an approach distinguishes itself from previous works in the sense that it provides science metrics from which the performance between different aperture models can be compared, being therefore not limited to an instrument level comparison only, and thus more consistent to the scientific needs of the mission. Indeed, *Kepler* and TESS missions adopt, analogous to the stellar pollution (SPR), the crowding metric r (Batalha *et al.* (2010)) and the dilution parameter D (Sullivan *et al.* (2015)), respectively, to quantitatively distinguish photometric fluxes originating from targets and other sources. However, these are instrumental level parameters that are not taken into account for choosing their apertures. From the results presented in subsection 4.9, we verify that compared to the binary mask, weighted masks (gradient and Gaussian) best fit the

instrumental PSF at pixel resolution, thus providing lower NSR in general; however their larger wings inevitably encompass more fractional flux from contaminant stars. From a science perspective, all three mask models present comparable overall efficiency in detecting legitimate planet transits, but the binary mask is substantially (up to ∼30%) less likely to produce background false positives when compared to weighted masks. These results led us to choose the binary mask as the optimal solution for extracting photometry in flight from P5 targets, since it provides the best compromise between maximizing N_{TCE}^{good} and minimizing N_{TCE}^{bad}.

The present approach currently represents a consistent contribution to the science of exoplanet searches. It confirms that the ordinary concept adopted in the literature for finding apertures, which typically relies on noise minimization for maximum transit detection, without directly taking into account the impact from false positives, is not necessarily the best strategy. Indeed, this statement holds for PLATO's P5 sample, as the conventional approach would suggest us the use of weighted masks instead of the binary mask. Furthermore, the approach for choosing apertures proposed in this thesis has been proven to be decisive for the determination of a mask model capable of providing near maximum planet yield and substantially reduced occurrence of false positives for the PLATO mission, thereby significantly reducing the amount of (useless) radial velocity follow-up of false planet transits. Overall, this work constitutes an important step in the design of both on-board and on-ground science data processing pipelines.

5.1.4 Flight software budget

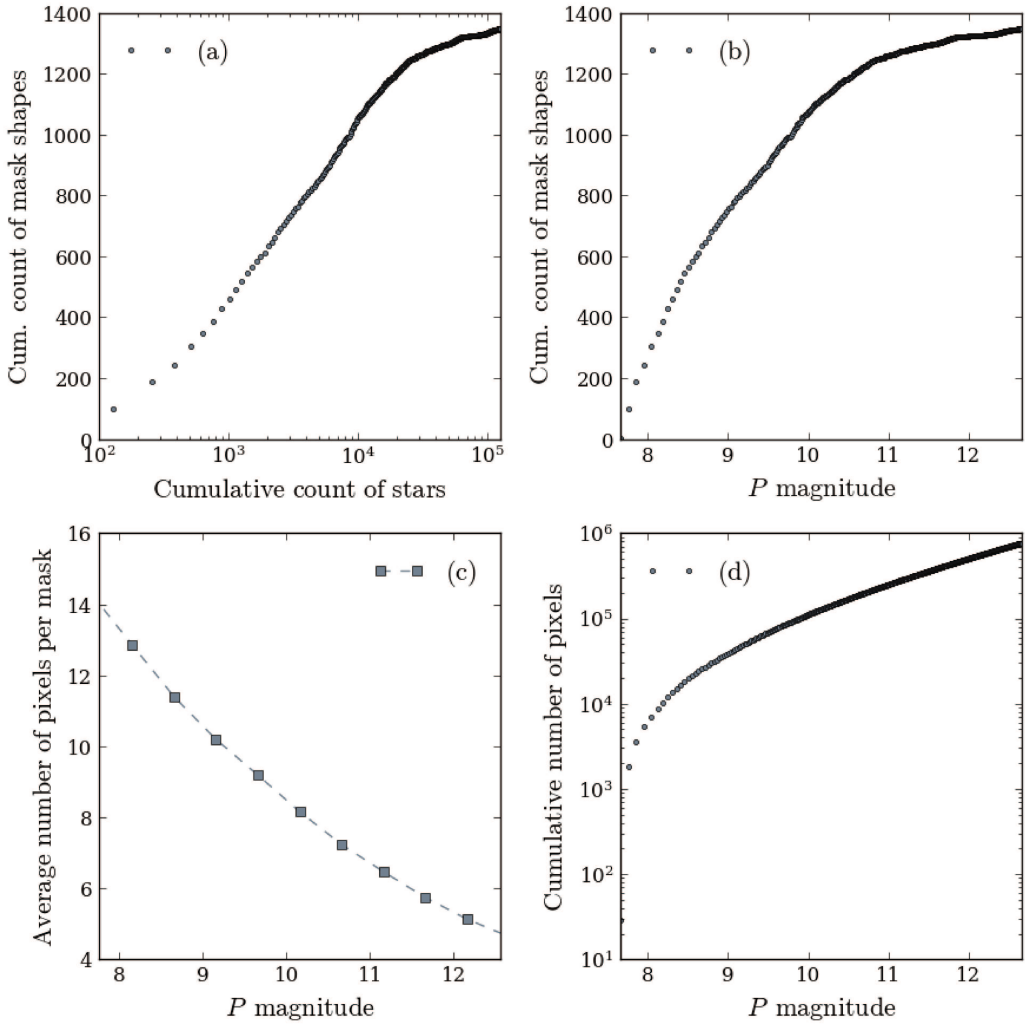
PLATO data processing pipeline requires implementing dedicated algorithms in the flight software, including an aperture photometry method for producing light curves for the P5 targets. Following the computation schemes described in subsection 4.8, the performance of three different aperture methods were compared by assigning them individually for 50,000 potential PLATO targets within SPF.

Considering the case of the gradient mask, which provides the lowest values of NSR, having such performance on board requires a unique mask shape per target to be uploaded to the flight software. This demands, in turn, prohibitive telemetry and time resources, especially considering that the total telemetry budget needs to be shared between 26 cameras. In addition, as explained in subsection 4.10, the masks will have to be regularly updated in flight to compensate for long-term star position drift, thus making the employment of gradient masks unfeasible.

For the Gaussian mask the outlook is not much more favourable, as this solution would require a massive set of widths (practically one per target) to guarantee the NSR performance results presented in Figure 45. Otherwise, one could in principle take advantage of the fact that the Gaussian mask has an analytical form – with small number

of parameters – to apply simplification schemes to avoid the need for having one particular mask per target. For instance, a feasible approach would consist of employing polynomial surfaces or fixed widths to cover the multiple combination scenarios in terms of magnitude and intra-pixel location of the targets. Nevertheless, that would inevitably reduce the overall NSR performance, which is the major benefit of using weighted masks, and potentially increase even more the number of false planet transit detections compared to the binary mask.

Figure 56 – Statistics on the morphology of binary masks. These results are based on all $\sim$ 127 thousand target stars within IF (see subsection 3.6). **(a)**: Cumulative count of unique binary mask shapes as a function of the cumulative count of target stars. **(b)**: Cumulative count of unique binary mask shapes as a function of target P magnitude. **(c)**: Average number of pixels composing the binary masks as a function of target P magnitude. **(d)**: Cumulative number of pixels composing the binary masks as a function of target P magnitude.



Source: author.

The binary mask, in turn, provides a virtually unbeatable capacity for compressing combinations of mask shapes without loss of performance. It can be visualized by looking at the data concerning the morphology of binary masks provided in Figure 56a. These data give the accumulated unique combinations of binary mask shapes computed from the set of binary masks used to extract photometry from all $\sim$ 127 thousand target stars within IF. It shows that the unique set saturates to about only 1,350 mask shapes, thereby giving a compression factor of almost 99%. This represents another significant advantage of employing the binary mask, since no weighted mask is actually capable of providing such compression capabilities – resulting in less on-board memory usage – while keeping the original performance of the full set of masks unchanged.

Furthermore, binary masks offers the advantage of permitting the flight software to perform calculations – involving the apertures – with integer numbers rather than float ones (as would be the case with weighted masks). This translates into less CPU usage on board. Indeed, in the context of the PLATO flight software, the Gaussian mask is almost 50% more CPU consuming than the binary mask. That represents a significant difference considering the typically tight CPU budgets involved in spatial missions.

5.1.5 Number of background eclipsing binaries

The work developed in this thesis allowed us to obtain a first estimate on the expected occurrence of background eclipsing binaries for the PLATO P5 stellar sample. From our results, this number is expected to be of the order of 80 occurrences per ten thousand target stars observed. Accordingly, for an observation scenario comprising two long-pointing fields (2 + 2 years), this ratio leads to a total of $\sim 2,000$ eclipsing binary occurrences over approximately 245,000 observed targets. Such number is notably large if compared to the expected planet yield under the same observation scenario, which is $\sim 4,600$ planets considering both P1 and P5 stellar samples (ESA (2017)). In that sense, the fact that the binary mask is expected to deliver up to 30% less occurrence of eclipsing binaries than weighted masks represents a significantly gain in terms of the amount of time spent to confirm transit-like events with ground-based radial velocity follow-up. In other words, the use of binary mask allows, from a statistical point of view, RV telescopes to be used much more efficiently, i.e. dedicating much more of their time in confirming legitimate planet transits rather than identifying false ones.

5.1.6 Weighted masks

Beyond the P5 sample, the weighted masks may be exploited as additional photometry extraction methods for the targets whose light curves will be produced on the ground from imagettes. Compared to more complex methods based on PSF fitting photometry (e.g. Silva *et al.* (2006), Deheuvels & Ballot (2019), Libralato *et al.* (2015), Nardiello *et*

al. (2016)), the gradient and Gaussian masks are much simpler and faster to calculate. They might be suitable for not too crowded fields or in situations where the existence of contaminants may not be too critic (e.g. for asteroseismology targets). We note however that these masks adapt their size to the presence of contaminant stars. This is possible since the expression for the NSR (Equation 4.1) takes into account the fluxes coming from contaminant companions, so whenever their signal are sufficiently strong, compared to that of the targets, the mask are reduced in width to keep NSR as low as possible. Moreover, the weighted masks can be implemented with ease in both *Kepler* and TESS data processing pipelines, so their usage is not limited to PLATO targets.

5.1.7 PLATO P photometric passband

The PLATO P photometric passband presented in subsection 3.5 allowed evidencing that stars specified in Johnson's V band have color-dependant fluxes on PLATO detectors. Quantitatively, the differences between V and P magnitudes are particularly large ($\gtrsim 0.8$ mag) for cold ($\mathcal{T}_\star \lesssim 4000$K) stars (see Figure 39). This is consistent with the fact that the spectral response of PLATO telescopes collects substantially more flux beyond $\lambda \sim 600$nm when compared to the V filter (see Figure 23). These results are important because they reveal that using the V band for characterizing the photometric performance of PLATO targets may lead to strong bias, since stars with identical V magnitudes may present significantly distinct fluxes at instrument level. Such bias does not occur when using the P band, making it therefore a more appropriate choice for e.g. selecting targets. Besides, stellar fluxes can be straightforwardly estimated from their corresponding P magnitudes by applying Equation 3.18. Furthermore, the color relationship presented in Equation 3.20 provides a straightforward and accurate way to switch between *Gaia* G and PLATO P magnitude systems.

5.2 Perspectives

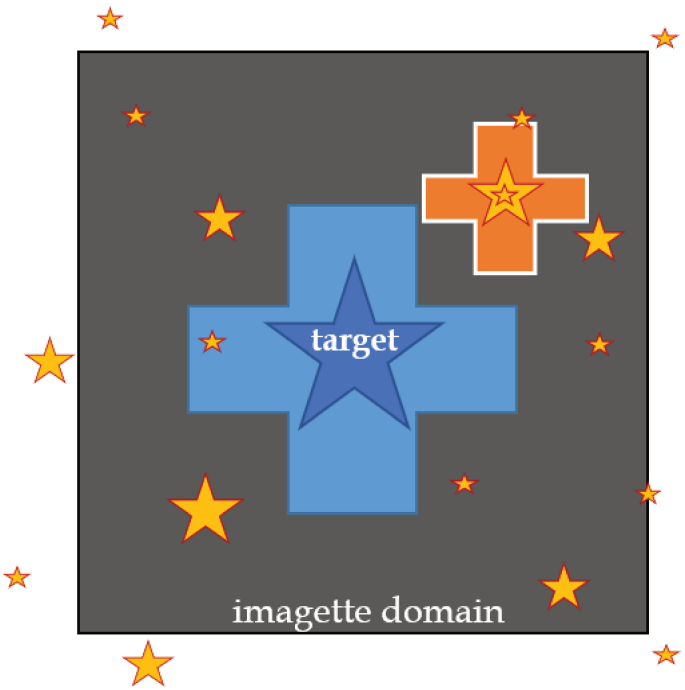
5.2.1 Detecting background false positives with double photometry

Despite the relevant contributions of the present study towards minimizing the frequency of background false positives in the P5 sample, a particular concern might still arise with regard to the potential difficulties in properly identifying, based on a single light curve per target, the false positives from the P5 detections. Indeed, detecting false positives under such condition is, when not impossible, a typically difficult task. Hence, it would be beneficial for the P5 photometry overall the development of an alternative (non-image-dependant) method for identifying false positives. With that in mind, the idea of applying double photometry in flight has emerged.

As already mentioned, only a limited number of targets of the P5 sample will have imagettes available on ground for extracting their photometry, which means that the vast

majority of these targets will have their photometry produced on board (Table 1). Yet, the apertures used to generate light curves on board will also be applied to imagettes of 6×6 pixel dimension. In parallel, the size of an imagette provides enough space for hosting an additional aperture (the secondary aperture) near that which is attributed to the target star (the primary aperture). Our hypothesis is that both primary and secondary apertures can, under certain conditions, bring altogether enough information allowing the identification of eventual background false positives in the target star's photometry, which otherwise would not be possible from the light curve produced by the primary aperture alone. More precisely, our hypothesis is that transit-like signals originating from a contaminant star will (typically) appear deeper in the secondary aperture than in the primary aperture, whereas transit-like signals originating from the target star will (typically) appear deeper in the primary aperture. Accordingly, such a method would allow us to distinguish between transit-like signals originating from the target star and transit-like signals originating from contaminant stars. While the former cannot be used to confirm whether it is a true or false planet transit, the latter can be used as an indicator for a potential background false positive. In order to such a method to be successful, two major challenges need to be overcome concerning the secondary mask: defining what shape (size) should it have and where to place it within the target imagette.

Figure 57 – Schematic of double aperture photometry on board. A secondary aperture (orange) is placed – within the 6×6 pixel area (dark grey) of the imagette – over a strategically selected contaminant star, in addition to the primary aperture (blue) of the target star.

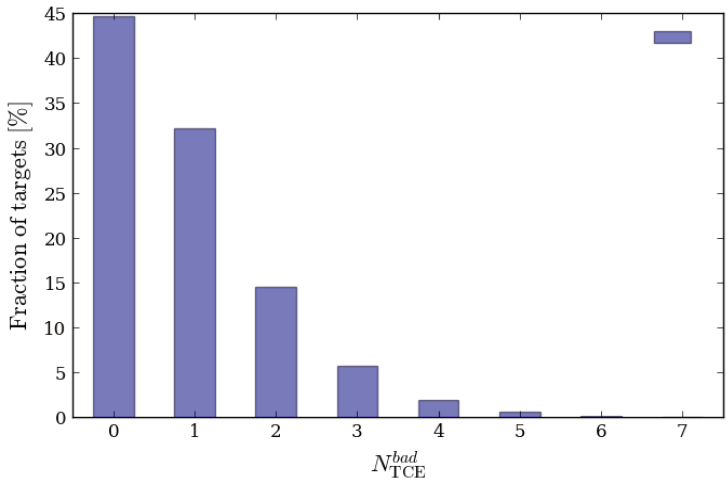


Source: author.

To answer the first question (what should be the size of the secondary aperture?), one could think of a strategy in which that aperture is set to be sufficiently large to cover as many contaminant stars nearby the target as possible, such as to detect as many background transit signals as possible. However, that strategy would not be effective because the larger the mask, the more diluted any transit-like signal within it will be, and the highest will also be its noise-to-signal ratio (notably owing to both background and readout noises). As a consequence, any transit-like signal originating from a contaminant star would most likely appear stronger in the target's photometry, suggesting therefore that it is originated from the target star when it is not. This constraint imposes therefore a limited size for the secondary aperture to limit the amount of dilution and noise in its photometry.

Finding an answer to the second question (where to place the secondary aperture?) might appear to be impossible, considering the high density of contaminant stars nearby a target star (Figure 40) and considering that any of these contaminants could in principle produce a background false positive. However, the results presented in Figure 51 show that the set of contaminants star with potential to produce background false positives are typically located at no more than a few pixels far from the target location, i.e. within the imagette domain. Moreover, the number of contaminant stars for which $\mathrm{SPR}_k \geqslant \mathrm{SPR}_k^{crit}$, i.e. N_{TCE}^{bad}, is on average one to two (rarely more than two) per target, as indicated in Figure 58 (computed for the binary mask). Also, this plot shows that about 80% of the targets have on average no more than one N_{TCE}^{bad}, which is quite encouraging for the idea of a secondary mask as conceived herein.

Figure 58 – Distribution of N_{TCE}^{bad} per target star (binary mask), based on the data of Figure 51.



Source: author.

Thus, a good strategy would be to place the secondary mask over the contaminant star presenting the greatest SPR_k among the set of contaminants given in Figure 51. Doing it increases the probability of placing an aperture over a contaminant star that will effectively produce a background false positive.

Finally, to statistically verify the hit ratio of such a method, it is necessary as a future work to compute secondary masks for the set of contaminants mentioned above and check for how many cases the noise-to-signal ratio and the level of dilution of the secondary photometry is adequately enough allowing us to state that if a transit depth in the secondary photometry is higher than in the primary one, then the transit signal observed in the primary photometry does not originate from a planet orbiting the target star. In fact, the frequency of targets for which such an statement is true corresponds to the expected frequency of background transit depths that the method is potentially capable of identifying. Fully characterizing the validity and the applicability of this method represents one of the major works in perspective.

5.2.2 Impact of PSF modelling and background correction uncertainties on planetary radius estimation

The SPR metric presented in this work is essential for quantifying the dilution level of observed planetary transit depths, and hence crucial for accurately determining the radii of new detected planets (Equation 4.8). In practice however, the SPR cannot be perfectly known owing (primarily) to imperfections associated to the PSF inversion process, background flux correction and uncertainties on stellar magnitudes. Accordingly, an important work to be performed in the future consists in quantifying the expected contributions of PSF modelling and background correction uncertainties on planetary radii estimation errors. This is particularly important in the context of the PLATO mission, considering its challenging science requirements on the accuracy of planetary raddi estimations (subsection 2.4).

In order to perform such work, one can simply replicate the study presented in this thesis by considering inverted PSFs (instead of the theoretical ones shown in Figure 21) and background flux values B (zodiacal light) given by the background modelling algorithm of WP 322 000 (Figure 25). Statistically speaking, that would give a realistic idea on how the PSF inversion and the background correction might impact on PLATO's exoplanet search science. Furthermore, it would also be possible to derive more scientifically meaningful metrics to assess the effectiveness of the PSF inversions. Today, the performance of inverted PSFs is evaluated exclusively in terms of L1 and L2 norms[1].

[1] The p-norm of an N-dimension vector v is defined as $||v_p|| = (v_1^p + v_2^p + v_3^p + ... + v_N^p)^{(1/p)}$.

5.2.3 Updating the binary masks used on board

Because of long-term star position drift (subsubsection 4.9.6), mask-target assignments performed during each calibration phase become, soon or later, no longer optimal since the flux distributions of the targets is significantly impacted by the intrapixel locations of their corresponding barycentres (Figure 22). To compensate for this effect, the proposed solution consists in tracking the targets by updating the placement of their apertures on board, as explained in Samadi *et al.* (2019). This will involve both ground and flight segments of the mission, as the apertures will first be computed on the ground and then transmitted to the spacecraft. Both criteria and timescale on which such actions shall be performed are yet to be defined.

Another challenge related to the mask update issue is the impracticability to upload all binary mask shapes (see Figure 56a) at once (e.g. during calibration phases) at a reasonable amount of time. Hence, instead of uploading all the mask shapes to stock them on board, they will rather be uploaded to the spacecraft at an on-demand basis, i.e. in conformity to the drift path being followed by the target stars during the observations. However, determining the required upload frequency is far from being a trivial problem to tackle, in particular owing to the large number of targets involved and the relatively limited telemetry resources available for transferring data between the spacecraft and the ground segment. Furthermore, the adopted frequency shall be consistent with the assumptions and constraints of on-ground operations.

Results from a preliminary study suggests the needed for a mask update frequency of once every 15 days per target star on average. However, such study still need to be refined to include e.g. the temporal gradient of zodiacal light (Figure 32) and the capabilities of the jitter correction algorithm, as this can also be used to correct sufficiently small drifts – contributing therefore to reduce the average frequency of mask updates.

5.2.4 Quantifying photometric performance of target stars fainter than $V \sim 13$

PLATO can take advantage of the many directions on the sky it will potentially observe throughout its lifetime, with high photometric precision, to perform studies of Galactic archaeology[2]. In that case, red giant (high luminosity) stars play a important role as these can be probed out to larger distances from our Solar System. Aperture photometry will also be widely employed to extract light curves in flight of (red giant) stars as faint as $V \sim 16$. Hence, an extended characterization of the aperture photometry performance (presented in Table 15) including stars down to $P = 15.66$ ($V = 16$ @ 6,000K) is necessary for determining what is the achievable seismic and planet detection performances from stars fainter than $V \sim 13$. In that particular context, Mosser *et al.* (2019) developped seismic performance metrics that can be applied to existing and future asteroseismic

[2] <http://www.ifa.hawaii.edu/~dhuber/archeology.html>

observations, including PLATO. However, the performance values derived for the PLATO mission are still based on the old photometric noise models published in Rauer *et al.* (2014). Therefore, the characterization of aperture photometry for fainter stars can be used to revisit the seismic performance of red giants to make it more consistent with the current status of the PLATO instrument performance.

Bibliography

Adams, F. C.; Graves, G. J. M.; Laughlin, G. Red Dwarfs and the End of the Main Sequence. In: Garcia-Segura, G. *et al.* (Ed.). *Revista Mexicana de Astronomia y Astrofisica Conference Series*. [S.l.: s.n.], 2004. v. 22, p. 46–49. Cited on page 21.

AERTS, C.; CHRISTENSEN-DALSGAARD, J.; KURTZ, D. W. *Asteroseismology*. Springer Netherlands, 2010. Disponível em: <https://doi.org/10.1007/978-1-4020-5803-5>. Cited on page 23.

AIGRAIN, S. *et al.* Precise time series photometry for the kepler-2.0 mission. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 447, n. 3, p. 2880–2893, jan 2015. Disponível em: <https://doi.org/10.1093/mnras/stu2638>. Cited on page 80.

AIGRAIN, S. *et al.* Robust, open-source removal of systematics in kepler data. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 471, n. 1, p. 759–769, jun 2017. Disponível em: <https://doi.org/10.1093/mnras/stx1422>. Cited on page 111.

ANDRAE, R. *et al.* Gaia data release 2. *Astronomy & Astrophysics*, EDP Sciences, v. 616, p. A8, aug 2018. Disponível em: <https://doi.org/10.1051/0004-6361/201732516>. Cited on page 76.

AUVERGNE, M. *et al.* The CoRoT satellite in flight: description and performance. *Astronomy & Astrophysics*, EDP Sciences, v. 506, n. 1, p. 411–424, mar 2009. Disponível em: <https://doi.org/10.1051/0004-6361/200810860>. Cited on page 16.

Baglin, A. *et al.* CoRoT: a high precision photometer for stellar ecolution and exoplanet finding. In: *36th COSPAR Scientific Assembly*. [S.l.: s.n.], 2006. v. 36, p. 3749. Cited 2 times on pages 16 and 23.

BAGLIN, A.; CHAINTREUIL, S.; VANDERMARCQ, O. II.1 the CoRoT observations. In: *The CoRoT Legacy Book*. EDP Sciences, 2016. p. 29. Disponível em: <https://doi.org/10.1051/978-2-7598-1876-1.c021>. Cited on page 42.

Bajaj, V.; Khandrika, H. *Comparing Aperture Photometry Software Packages*. [S.l.], 2017. 10 p. Cited on page 79.

BATALHA, N. M. *et al.* SELECTION, PRIORITIZATION, AND CHARACTERISTICS OF KEPLER TARGET STARS. *The Astrophysical Journal*, IOP Publishing, v. 713, n. 2, p. L109–L114, mar 2010. Disponível em: <https://doi.org/10.1088/2041-8205/713/2/l109>. Cited 2 times on pages 83 and 112.

BAUDIN, F.; DAMIANI, C. Des planétes terres partout? *L'Astronomie*, Société Astronomique de France, v. 133, p. 24–34, apr 2019. Disponível em: <http://boutique-saf-astronomie.com/acatalog/LAstronomie-Avril-2019-N-126-Astro_04_2019.html>. Cited on page 20.

BESSELL, M. S. UBVRI passbands. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 102, p. 1181, oct 1990. Disponível em: <https://doi.org/10.1086/132749>. Cited on page 47.

BIN, J.; TIAN, F.; LIU, L. New inner boundaries of the habitable zones around m dwarfs. *Earth and Planetary Science Letters*, Elsevier BV, v. 492, p. 121–129, jun. 2018. Disponível em: <https://doi.org/10.1016/j.epsl.2018.04.003>. Cited on page 15.

Bohlin, R. C. HST Stellar Standards with 1% Accuracy in Absolute Flux. In: Sterken, C. (Ed.). *The Future of Photometric, Spectrophotometric and Polarimetric Standardization*. [S.l.: s.n.], 2007. (Astronomical Society of the Pacific Conference Series, v. 364), p. 315. Cited on page 69.

BöRNER, A. *et al. Instrument Signal and Noise Budget. PLATO-DLR-PL-RP-001, issue 4.0*. [S.l.], 2018. 1–59 p. Cited 2 times on pages 71 and 135.

BORUCKI, W. J. *et al.* Kepler planet-detection mission: Introduction and first results. *Science*, American Association for the Advancement of Science (AAAS), v. 327, n. 5968, p. 977–980, jan 2010. Disponível em: <https://doi.org/10.1126/science.1185402>. Cited 2 times on pages 16 and 40.

BOZZA, V.; MANCINI, L.; SOZZETTI, A. (Ed.). *Methods of Detecting Exoplanets*. Springer International Publishing, 2016. Disponível em: <https://doi.org/10.1007/978-3-319-27458-4>. Cited 3 times on pages 25, 27, and 28.

BRYSON, S. T. *et al.* Identification of background false positives from Kepler data. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 125, n. 930, p. 889–923, aug 2013. Disponível em: <https://doi.org/10.1086/671767>. Cited 5 times on pages 28, 31, 80, 99, and 103.

BRYSON, S. T. *et al.* Selecting pixels for kepler downlink. In: RADZIWILL, N. M.; BRIDGER, A. (Ed.). *Software and Cyberinfrastructure for Astronomy*. SPIE, 2010. Disponível em: <https://doi.org/10.1117/12.857625>. Cited on page 79.

BUCHHAVE, L. A. *et al.* An abundance of small exoplanets around stars with a wide range of metallicities. *Nature*, Springer Science and Business Media LLC, v. 486, n. 7403, p. 375–377, jun. 2012. Disponível em: <https://doi.org/10.1038/nature11121>. Cited on page 22.

BULDGEN, G. *et al.* Stellar acoustic radii, mean densities, and ages from seismic inversion techniques. *Astronomy & Astrophysics*, EDP Sciences, v. 574, p. A42, jan. 2015. Disponível em: <https://doi.org/10.1051/0004-6361/201424613>. Cited on page 23.

Busso, G. *et al. Gaia DR2 documentation Chapter 5: Photometry*. [S.l.], 2018. 5 p. Cited 2 times on pages 72 and 76.

CAMPANTE, T. L.; SANTOS, N. C.; MONTEIRO, M. J. P. F. G. (Ed.). *Asteroseismology and Exoplanets: Listening to the Stars and Searching for New Worlds*. Springer International Publishing, 2018. Disponível em: <https://doi.org/10.1007/978-3-319-59315-9>. Cited 5 times on pages 20, 21, 23, 24, and 26.

CAPORAEL, L. R. Astrobiology as hybrid science: Introduction to the thematic issue. *Biological Theory*, Springer Nature, v. 13, n. 2, p. 69–75, abr. 2018. Disponível em: <https://doi.org/10.1007/s13752-018-0296-8>. Cited on page 15.

CASAGRANDE, L.; VANDENBERG, D. A. On the use of gaia magnitudes and new tables of bolometric corrections. *Monthly Notices of the Royal Astronomical Society: Letters*, Oxford University Press (OUP), v. 479, n. 1, p. L102–L107, jun 2018. Disponível em: <https://doi.org/10.1093/mnrasl/sly104>. Cited 2 times on pages 70 and 71.

CESSA, V. *et al.* CHEOPS: a space telescope for ultra-high precision photometry of exoplanet transits. In: CUGNY, B.; SODNIK, Z.; KARAFOLAS, N. (Ed.). *International Conference on Space Optics — ICSO 2014*. SPIE, 2017. Disponível em: <https://doi.org/10.1117/12.2304164>. Cited on page 17.

Chambers, K. C. *et al.* The Pan-STARRS1 Surveys. *arXiv e-prints*, p. arXiv:1612.05560, dez. 2016. Cited on page 60.

COX, A. N. (Ed.). *Allen's Astrophysical Quantities*. Springer New York, 2002. Disponível em: <https://doi.org/10.1007/978-1-4612-1186-0>. Cited on page 21.

DEHEUVELS, S.; BALLOT, J. *Photometry of Imagettes. PLATO-imagettes, issue 1, rev.0*. [S.l.], 2019. 1–27 p. Cited 3 times on pages 29, 115, and 116.

DELEUIL, M. *et al.* Planets, candidates, and binaries from the CoRoT/exoplanet programme. *Astronomy & Astrophysics*, EDP Sciences, v. 619, p. A97, nov 2018. Disponível em: <https://doi.org/10.1051/0004-6361/201731068>. Cited on page 95.

DELEUIL, M.; FRIDLUND, M. CoRoT: The first space-based transit survey to explore the close-in planet population. In: *Handbook of Exoplanets*. Springer International Publishing, 2018. p. 1135–1158. Disponível em: <https://doi.org/10.1007/978-3-319-55333-7_79>. Cited on page 16.

DRUMMOND, R. *et al.* Correcting for background changes in CoRoT exoplanet data. *Astronomy & Astrophysics*, EDP Sciences, v. 487, n. 3, p. 1209–1220, jul 2008. Disponível em: <https://doi.org/10.1051/0004-6361:200809639>. Cited on page 85.

ESA. *PLATO - Definition Study Report (RED BOOK). ESA-SCI(2017)1*. [S.l.], 2017. 1–139 p. Disponível em: <http://sci.esa.int/jump.cfm?oid=59252>. Cited 9 times on pages 19, 37, 38, 40, 41, 44, 46, 109, and 115.

EVANS, D. W. *et al.* Gaia data release 2. *Astronomy & Astrophysics*, EDP Sciences, v. 616, p. A4, aug 2018. Disponível em: <https://doi.org/10.1051/0004-6361/201832756>. Cited 4 times on pages 70, 71, 75, and 98.

FIALHO, F. D. O. *et al.* Jitter correction algorithms for the COROT satellite mission: Validation with test bench data and MOST on-orbit photometry. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 119, n. 853, p. 337–346, mar 2007. Disponível em: <https://doi.org/10.1086/513882>. Cited on page 81.

FISCHER, D. A.; VALENTI, J. The planet-metallicity correlation. *The Astrophysical Journal*, IOP Publishing, v. 622, n. 2, p. 1102–1117, abr. 2005. Disponível em: <https://doi.org/10.1086/428383>. Cited on page 21.

FORTIER, A. *et al.* Planet formation models: the interplay with the planetesimal disc. *Astronomy & Astrophysics*, EDP Sciences, v. 549, p. A44, dez. 2012. Disponível em: <https://doi.org/10.1051/0004-6361/201220241>. Cited on page 20.

FRESSIN, F. *et al.* THE FALSE POSITIVE RATE OF KEPLER AND THE OCCURRENCE OF PLANETS. *The Astrophysical Journal*, IOP Publishing, v. 766, n. 2, p. 81, mar 2013. Disponível em: <https://doi.org/10.1088/0004-637x/766/2/81>. Cited on page 97.

Gaia Collaboration *et al.* Gaia data release 2. *Astronomy & Astrophysics*, EDP Sciences, v. 616, p. A1, aug 2018b. Disponível em: <https://doi.org/10.1051/0004-6361/201833051>. Cited on page 60.

Gaia Collaboration *et al.* The Gaia mission. *Astronomy & Astrophysics*, EDP Sciences, v. 595, p. A1, nov 2016. Disponível em: <https://doi.org/10.1051/0004-6361/201629272>. Cited 2 times on pages 20 and 60.

GALANTE, D. *et al. Astrobiologia: uma ciência emergente*. Tikinet Edição Ltda, 2016. ISBN 978-85-66241-03-7. Disponível em: <http://tikinet.kinghost.net/default.asp>. Cited on page 15.

GEHAN, C. *Evolution of the core rotation of stars on the red giant branch: from large-scale measurements towards a characterisation of the angular momentum transport*. Tese (Theses) — Observatoire de Paris – LESIA, set. 2018. Disponível em: <https://hal.archives-ouvertes.fr/tel-01890184>. Cited on page 23.

GILLILAND, R. L. *et al.* KEPLERMISSION STELLAR AND INSTRUMENT NOISE PROPERTIES. *The Astrophysical Journal Supplement Series*, IOP Publishing, v. 197, n. 1, p. 6, out. 2011. Disponível em: <https://doi.org/10.1088/0067-0049/197/1/6>. Cited on page 93.

GOUPIL, M. Expected asteroseismic performances with the space project PLATO. *EPJ Web of Conferences*, EDP Sciences, v. 160, p. 01003, 2017. Disponível em: <https://doi.org/10.1051/epjconf/201716001003>. Cited on page 37.

GROLLEAU, E. *et al. Synopsis of on-board processing algorithms. PLATO-LESIA-PDC-RP-023, issue 1, rev.4*. [S.l.], 2018. 1–27 p. Disponível em: <https://projets-lesia.obspm.fr/projects/platoalgo/files>. Cited on page 49.

GRüN, E. *et al.* (Ed.). *Interplanetary Dust*. Springer Berlin Heidelberg, 2001. Disponível em: <https://doi.org/10.1007/978-3-642-56428-4>. Cited on page 59.

GUYONNET, A. *et al.* Evidence for self-interaction of charge distribution in charge-coupled devices. *Astronomy & Astrophysics*, EDP Sciences, v. 575, p. A41, feb 2015. Disponível em: <https://doi.org/10.1051/0004-6361/201424897>. Cited on page 45.

HANDBERG, R.; LUND, M. N. K2p2: Reduced data from campaigns 0–4 of the k2 mission. *Astronomy & Astrophysics*, EDP Sciences, v. 597, p. A36, dec 2016. Disponível em: <https://doi.org/10.1051/0004-6361/201527753>. Cited on page 80.

HELLER, R. Analytic solutions to the maximum and average exoplanet transit depth for common stellar limb darkening laws. *Astronomy & Astrophysics*, EDP Sciences, v. 623, p. A137, mar. 2019. Disponível em: <https://doi.org/10.1051/0004-6361/201834620>. Cited on page 83.

HENRY, T. J.; KIRKPATRICK, J. D.; SIMONS, D. A. The solar neighborhood, 1: Standard spectral types (k5-m8) for northern dwarfs within eight parsecs. *The Astronomical Journal*, IOP Publishing, v. 108, p. 1437, out. 1994. Disponível em: <https://doi.org/10.1086/117167>. Cited on page 21.

HOWELL, S. B. Two-dimensional aperture photometry - signal-to-noise ratio of point-source observations and optimal data-extraction techniques. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 101, p. 616, jun 1989. Disponível em: <https://doi.org/10.1086/132477>. Cited on page 79.

IVEZIĆ, Ž. *et al.* LSST: From science drivers to reference design and anticipated data products. *The Astrophysical Journal*, American Astronomical Society, v. 873, n. 2, p. 111, mar. 2019. Disponível em: <https://doi.org/10.3847/1538-4357/ab042c>. Cited on page 60.

JENKINS, J. M. *et al.* OVERVIEW OF THE KEPLER SCIENCE PROCESSING PIPELINE. *The Astrophysical Journal*, IOP Publishing, v. 713, n. 2, p. L87–L91, mar 2010. Disponível em: <https://doi.org/10.1088/2041-8205/713/2/l87>. Cited on page 83.

JENKINS, J. M. *et al.* Transiting planet search in the kepler pipeline. In: RADZIWILL, N. M.; BRIDGER, A. (Ed.). *Software and Cyberinfrastructure for Astronomy*. SPIE, 2010b. Disponível em: <https://doi.org/10.1117/12.856764>. Cited on page 80.

JOHNSON, J. A. *et al.* Giant planet occurrence in the stellar mass-metallicity plane. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 122, n. 894, p. 905–915, ago. 2010. Disponível em: <https://doi.org/10.1086/655775>. Cited on page 21.

JORDI, C. *et al.* Gaiabroad band photometry. *Astronomy & Astrophysics*, EDP Sciences, v. 523, p. A48, nov 2010. Disponível em: <https://doi.org/10.1051/0004-6361/201015441>. Cited on page 70.

KASTING, J. F.; WHITMIRE, D. P.; REYNOLDS, R. T. Habitable zones around main sequence stars. *Icarus*, Elsevier BV, v. 101, n. 1, p. 108–128, jan. 1993. Disponível em: <https://doi.org/10.1006/icar.1993.1010>. Cited on page 15.

KONACKI, M. *et al.* An extrasolar planet that transits the disk of its parent star. *Nature*, Springer Nature, v. 421, n. 6922, p. 507–509, jan. 2003. Disponível em: <https://doi.org/10.1038/nature01379>. Cited on page 25.

KOPPARAPU, R. K. *et al.* HABITABLE ZONES AROUND MAIN-SEQUENCE STARS: NEW ESTIMATES. *The Astrophysical Journal*, IOP Publishing, v. 765, n. 2, p. 131, fev. 2013. Disponível em: <https://doi.org/10.1088/0004-637x/765/2/131>. Cited on page 15.

LEBRETON, Y.; GOUPIL, M. J. Asteroseismology for "à la carte" stellar age-dating and weighing. *Astronomy & Astrophysics*, EDP Sciences, v. 569, p. A21, sep 2014. Disponível em: <https://doi.org/10.1051/0004-6361/201423797>. Cited on page 23.

LEINERT, C. *et al.* The 1997 reference of diffuse night sky brightness. *Astronomy and Astrophysics Supplement Series*, EDP Sciences, v. 127, n. 1, p. 1–99, jan 1998. Disponível em: <https://doi.org/10.1051/aas:1998105>. Cited 3 times on pages 57, 58, and 60.

LIBRALATO, M. *et al.* A PSF-based approach toKepler/k2data – i. variability within theK2campaign 0 star clusters m 35 and NGC 2158. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 456, n. 2, p. 1137–1162, dec 2015. Disponível em: <https://doi.org/10.1093/mnras/stv2628>. Cited 2 times on pages 115 and 116.

Llebaria, A.; Guterman, P. Building up Photometric Apertures for the Exoplanet Channel. In: Fridlund, M. *et al.* (Ed.). *The CoRoT Mission Pre-Launch Status - Stellar Seismology and Planet Finding*. [S.l.: s.n.], 2006. (ESA Special Publication, v. 1306), p. 293. Cited on page 79.

LUND, M. N. *et al.* K2p2—a PHOTOMETRY PIPELINE FOR THE k2 MISSION. *The Astrophysical Journal*, IOP Publishing, v. 806, n. 1, p. 30, jun 2015. Disponível em: <https://doi.org/10.1088/0004-637x/806/1/30>. Cited on page 80.

LUND, M. N. *et al.* Data preparation for asteroseismology with TESS. *EPJ Web of Conferences*, EDP Sciences, v. 160, p. 01005, 2017. Disponível em: <https://doi.org/10.1051/epjconf/201716001005>. Cited on page 80.

MARCHIORI, V. *et al.* In-flight photometry extraction of PLATO targets. *Astronomy & Astrophysics*, EDP Sciences, v. 627, p. A71, jul. 2019. Disponível em: <https://doi.org/10.1051/0004-6361/201935269>. Cited 3 times on pages 34, 57, and 79.

MASSEY, R. *et al.* An improved model of charge transfer inefficiency and correction algorithm for the hubble space telescope. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 439, n. 1, p. 887–907, feb 2014. Disponível em: <https://doi.org/10.1093/mnras/stu012>. Cited on page 45.

MATHÉ, C.; SAMADI, R. *PSF modelling: Research optimal parameters with different microscanning mode. PLATO-LESIA-PDC-TN-018, issue 1, rev.0*. [S.l.], 2015. 1–23 p. Disponível em: <https://projets-lesia.obspm.fr/projects/platoalgo/wiki/321000>. Cited on page 49.

MAYOR, M.; QUELOZ, D. A jupiter-mass companion to a solar-type star. *Nature*, Springer Nature, v. 378, n. 6555, p. 355–359, nov. 1995. Disponível em: <https://doi.org/10.1038/378355a0>. Cited on page 16.

MOLINA, J.; DALLAMICO, J. *Scattered Straylight Analysis Report. PLATO-OHB-PL-TN-012, issue 01D01*. [S.l.], 2016. 1–34 p. Cited on page 59.

MOSSER, B. *et al.* Seismic performance. *Astronomy & Astrophysics*, EDP Sciences, v. 622, p. A76, jan 2019. Disponível em: <https://doi.org/10.1051/0004-6361/201834607>. Cited on page 120.

MOSSER, B.; MIGLIO, A. IV.2 pulsating red giant stars. In: *The CoRoT Legacy Book*. EDP Sciences, 2016. p. 197. Disponível em: <https://doi.org/10.1051/978-2-7598-1876-1.c042>. Cited on page 23.

NARDIELLO, D. *et al.* A PSF-based approach toKepler/k2data – III. search for exoplanets and variable stars within the open cluster m 67 (NGC 2682). *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 463, n. 2, p. 1831–1843, aug 2016. Disponível em: <https://doi.org/10.1093/mnras/stw2169>. Cited 2 times on pages 115 and 116.

NASUDDIN, K. A.; ABDULLAH, M.; HAMID, N. S. A. Characterization of the south atlantic anomaly. *Nonlinear Processes in Geophysics*, Copernicus GmbH, v. 26, n. 1, p. 25–35, mar. 2019. Disponível em: <https://doi.org/10.5194/npg-26-25-2019>. Cited on page 42.

NAYLOR, T. An optimal extraction algorithm for imaging photometry. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 296, n. 2, p. 339–346, may 1998. Disponível em: <https://doi.org/10.1046/j.1365-8711.1998.01314.x>. Cited on page 79.

PALACIOS, A. *et al.* POLLUX: a database of synthetic stellar spectra. *Astronomy and Astrophysics*, EDP Sciences, v. 516, p. A13, jun 2010. Disponível em: <https://doi.org/10.1051/0004-6361/200913932>. Cited on page 69.

PERRYMAN, M. A. C. *The exoplanet handbook*. Cambridge: New York Cambridge University Press, 2018. ISBN 9781108304160. Cited on page 24.

PERTENAIS, M.; WOHLFEIL, J.; PETER, G. *Instrument Technical Requirement Document (TRD). PLATO-DLR-PL-RS-0001, issue 3.2*. [S.l.], 2018. 1–237 p. Cited on page 48.

PICKLES, A. J. A stellar spectral flux library: 1150–25000 Å. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 110, n. 749, p. 863–878, jul 1998. Disponível em: <https://doi.org/10.1086/316197>. Cited on page 47.

PLATO Study Team. *PLATO Science Requirements Document. ESA-PLATO-ESTEC-SCI-RS-001, issue 5, rev.3*. [S.l.], 2017. 1–54 p. Cited 4 times on pages 35, 37, 39, and 43.

RAUER, H. *et al.* The PLATO 2.0 mission. *Experimental Astronomy*, Springer Nature, v. 38, n. 1-2, p. 249–330, sep 2014. Disponível em: <https://doi.org/10.1007/s10686-014-9383-4>. Cited 2 times on pages 33 and 121.

REESE, D.; MARCHIORI, V. *PSF inversions in the presence of contaminating stars. PLATO-LESIA-PDC-TN-0039, issue 1, rev.0*. [S.l.], 2018. 1–23 p. Disponível em: <https://projets-lesia.obspm.fr/projects/platoalgo/wiki/321000>. Cited on page 50.

REESE, D. R. *et al.* Estimating stellar mean density through seismic inversions. *Astronomy & Astrophysics*, EDP Sciences, v. 539, p. A63, fev. 2012. Disponível em: <https://doi.org/10.1051/0004-6361/201118156>. Cited on page 23.

RICKER, G. R. *et al.* Transiting exoplanet survey satellite. *Journal of Astronomical Telescopes, Instruments, and Systems*, SPIE-Intl Soc Optical Eng, v. 1, n. 1, p. 014003, oct 2014. Disponível em: <https://doi.org/10.1117/1.jatis.1.1.014003>. Cited 2 times on pages 16 and 40.

SAMADI, R. *et al. Synopsis of on-board processing algorithms. PLATO-LESIA-PDC-RP-024, issue 2, rev.4*. [S.l.], 2018. 1–17 p. Disponível em: <https://projets-lesia.obspm.fr/projects/platoalgo/files>. Cited on page 49.

SAMADI, R. *et al.* The PLATO solar-like light-curve simulator. *Astronomy & Astrophysics*, EDP Sciences, v. 624, p. A117, abr. 2019. Disponível em: <https://doi.org/10.1051/0004-6361/201834822>. Cited 5 times on pages 30, 49, 55, 111, and 120.

SHIELDS, A. L.; BALLARD, S.; JOHNSON, J. A. The habitability of planets orbiting m-dwarf stars. *Physics Reports*, Elsevier BV, v. 663, p. 1–38, dez. 2016. Disponível em: <https://doi.org/10.1016/j.physrep.2016.10.003>. Cited on page 15.

SHORT, A. *et al.* An analytical model of radiation-induced charge transfer inefficiency for CCD detectors. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 430, n. 4, p. 3078–3085, feb 2013. Disponível em: <https://doi.org/10.1093/mnras/stt114>. Cited on page 45.

SILVA, L. P. da *et al.* Estimation of a super-resolved PSF for the data reduction of undersampled stellar observations. *Astronomy & Astrophysics*, EDP Sciences, v. 452, n. 1, p. 363–369, maio 2006. Disponível em: <https://doi.org/10.1051/0004-6361:20054614>. Cited 2 times on pages 115 and 116.

SMITH, J. C. *et al.* Finding optimal apertures in kepler data. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 128, n. 970, p. 124501, oct 2016. Disponível em: <https://doi.org/10.1088/1538-3873/128/970/124501>. Cited 2 times on pages 79 and 80.

STETSON, P. B. DAOPHOT - a computer program for crowded-field stellar photometry. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 99, p. 191, mar 1987. Disponível em: <https://doi.org/10.1086/131977>. Cited on page 79.

SULLIVAN, P. W. *et al.* THE TRANSITING EXOPLANET SURVEY SATELLITE: SIMULATIONS OF PLANET DETECTIONS AND ASTROPHYSICAL FALSE POSITIVES. *The Astrophysical Journal*, IOP Publishing, v. 809, n. 1, p. 77, aug 2015. Disponível em: <https://doi.org/10.1088/0004-637x/809/1/77>. Cited on page 112.

TWICKEN, J. D. *et al.* Kepler data validation i—architecture, diagnostic tests, and data products for vetting transiting planet candidates. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 130, n. 988, p. 064502, apr 2018. Disponível em: <https://doi.org/10.1088/1538-3873/aab694>. Cited on page 26.

TWICKEN, J. D. *et al.* Photometric analysis in the kepler science operations center pipeline. In: RADZIWILL, N. M.; BRIDGER, A. (Ed.). *Software and Cyberinfrastructure for Astronomy*. SPIE, 2010. Disponível em: <https://doi.org/10.1117/12.856790>. Cited on page 85.

VERAS, D. *et al.* Prospects for detecting decreasing exoplanet frequency with main-sequence age using PLATO. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 453, n. 1, p. 67–72, aug 2015. Disponível em: <https://doi.org/10.1093/mnras/stv1615>. Cited on page 20.

WEILER, M. Revised gaia data release 2 passbands. *Astronomy & Astrophysics*, EDP Sciences, v. 617, p. A138, sep 2018. Disponível em: <https://doi.org/10.1051/0004-6361/201833462>. Cited 2 times on pages 70 and 71.

WIDENHORN, R. Charge diffusion in the field-free region of charge-coupled devices. *Optical Engineering*, SPIE-Intl Soc Optical Eng, v. 49, n. 4, p. 044401, apr 2010. Disponível em: <https://doi.org/10.1117/1.3386514>. Cited on page 45.

WOLF, C. *et al.* SkyMapper southern survey: First data release (DR1). *Publications of the Astronomical Society of Australia*, Cambridge University Press (CUP), v. 35, 2018. Disponível em: <https://doi.org/10.1017/pasa.2018.5>. Cited on page 60.

WOLSZCZAN, A.; FRAIL, D. A. A planetary system around the millisecond pulsar psr1257 + 12. *Nature*, Springer Nature, v. 355, n. 6356, p. 145–147, jan. 1992. Disponível em: <https://doi.org/10.1038/355145a0>. Cited on page 16.

Appendix

APPENDIX A – Photometric performance breakdown of PLATO P5 stellar sample

Table 15 – Predicted binary mask photometric performance, relatively to PLATO's P5 stellar sample. Noise values in this table represent medians plus the corresponding upper and lower bounds at 90% confidence level, computed over the number of targets indicated at each row. These values are consistent with those presented in Börner *et al.* (2018). Values of the column $\mathrm{NSR}_{1\mathrm{h}}^{jitter}$ correspond to the plot of Figure 45, thereby valid for $N_T = 24$ telescopes. Noise (σ) values represent the fractional contributions to $\mathrm{NSR}_{1\mathrm{h}}^{jitter}$ due to noise originating from target stars (F_T), contaminant stars (F_C), background zodiacal light (B), detector (D), quantization (Q) and photometric jitter (J).

P magnitude (± 0.1)	Number of targets	Number of contaminants	Average contam. counts per target	σ_{F_T} [ppm hr$^{1/2}$]	σ_{F_C} [ppm hr$^{1/2}$]	σ_B [ppm hr$^{1/2}$]	σ_D [ppm hr$^{1/2}$]	σ_Q [ppm hr$^{1/2}$]	σ_J [ppm hr$^{1/2}$]	$\mathrm{NSR}_{1\mathrm{h}}^{jitter}$ [ppm hr$^{1/2}$]	SPR_{tot} [ppm]
7.66	82	3,251	39.6	$10.5^{+0.6}_{-0.5}$	$0.1^{+0.7}_{-0.1}$	$0.7^{+0.2}_{-0.1}$	$1.5^{+0.2}_{-0.2}$	$0.2^{+0.0}_{-0.0}$	$0.2^{+0.3}_{-0.1}$	$10.6^{+0.7}_{-0.6}$	$178^{+6,313}_{-177}$
8.16	127	5,132	40.4	$12.7^{+0.8}_{-0.8}$	$0.2^{+1.2}_{-0.2}$	$1.0^{+0.3}_{-0.2}$	$2.1^{+0.3}_{-0.4}$	$0.2^{+0.0}_{-0.0}$	$0.2^{+0.3}_{-0.1}$	$12.9^{+0.8}_{-0.8}$	$270^{+13,314}_{-267}$
8.66	224	8,594	38.4	$16.0^{+1.0}_{-0.9}$	$0.3^{+2.0}_{-0.3}$	$1.5^{+0.5}_{-0.3}$	$3.1^{+0.5}_{-0.5}$	$0.4^{+0.1}_{-0.1}$	$0.4^{+0.9}_{-0.3}$	$16.4^{+1.2}_{-1.0}$	$414^{+18,888}_{-413}$
9.16	380	15,023	39.5	$20.1^{+1.3}_{-1.2}$	$0.4^{+3.3}_{-0.4}$	$2.2^{+0.9}_{-0.4}$	$4.6^{+1.0}_{-0.6}$	$0.5^{+0.1}_{-0.1}$	$0.5^{+1.0}_{-0.4}$	$20.8^{+1.7}_{-1.4}$	$488^{+32,134}_{-487}$
9.66	648	27,489	42.4	$25.4^{+1.7}_{-1.5}$	$0.7^{+3.8}_{-0.7}$	$3.3^{+1.2}_{-0.6}$	$6.9^{+1.3}_{-1.0}$	$0.8^{+0.2}_{-0.1}$	$0.7^{+1.6}_{-0.6}$	$26.7^{+2.1}_{-1.6}$	$819^{+29,566}_{-818}$
10.16	1045	44,341	42.4	$32.2^{+2.2}_{-1.9}$	$1.0^{+6.7}_{-1.0}$	$5.0^{+1.9}_{-0.9}$	$10.6^{+1.7}_{-1.8}$	$1.2^{+0.2}_{-0.2}$	$1.0^{+2.5}_{-0.8}$	$34.5^{+3.1}_{-2.3}$	$964^{+50,807}_{-962}$
10.66	1596	67,728	42.4	$40.8^{+2.8}_{-2.4}$	$1.5^{+8.0}_{-1.4}$	$7.6^{+3.0}_{-1.5}$	$16.0^{+3.0}_{-3.0}$	$1.9^{+0.4}_{-0.4}$	$1.4^{+3.0}_{-1.0}$	$44.8^{+4.1}_{-3.1}$	$1,238^{+45,401}_{-1237}$
11.16	2472	108,392	43.8	$52.0^{+3.6}_{-3.2}$	$2.2^{+12.9}_{-2.2}$	$11.7^{+4.7}_{-2.4}$	$24.6^{+4.8}_{-4.6}$	$2.9^{+0.6}_{-0.5}$	$1.9^{+4.1}_{-1.4}$	$59.2^{+6.2}_{-4.4}$	$1,745^{+70,036}_{-1743}$
11.66	3729	160,505	43.0	$66.1^{+5.2}_{-4.3}$	$3.3^{+20.7}_{-3.2}$	$17.8^{+7.5}_{-3.7}$	$37.6^{+7.4}_{-6.9}$	$4.4^{+0.9}_{-0.8}$	$2.6^{+6.4}_{-2.0}$	$79.1^{+9.7}_{-6.6}$	$2,262^{+106,434}_{-2260}$
12.16	5705	254,262	44.6	$84.3^{+7.1}_{-5.8}$	$4.7^{+28.6}_{-4.6}$	$27.2^{+11.7}_{-5.5}$	$57.3^{+12.1}_{-10.5}$	$6.7^{+1.4}_{-1.2}$	$3.6^{+7.8}_{-2.7}$	$106.8^{+14.8}_{-10.1}$	$2,838^{+119,402}_{-2835}$
12.66	7733	334,293	43.2	$102.6^{+9.2}_{-7.4}$	$6.4^{+37.8}_{-6.2}$	$38.5^{+16.7}_{-7.7}$	$81.1^{+17.0}_{-15.0}$	$9.6^{+2.0}_{-1.8}$	$4.3^{+8.6}_{-3.2}$	$138.5^{+20.1}_{-14.7}$	$3,362^{+133,603}_{-3358}$

Source: author.

Annex

ANNEX A – Published papers

Astronomy & Astrophysics

Worldwide astronomical and astrophysical research

Editor-in-Chief: Thierry Forveille
Letters Editor-in-Chief: João Alves
Managing Editor: David Elbaz

2018 Impact Factor: 6.209

PUBLISH WITH A&A

Astronomy & Astrophysics is a leading international journal that publishes papers on all aspects of astronomy and astrophysics (theoretical, observational, and instrumental).

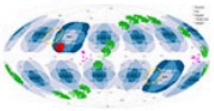
Find out more

LATEST HIGHLIGHTS | All

- A population of high-velocity absorption-line systems residing in the Local Group (Bouma et al.)
- No missing baryons in massive spirals (Posti et al.)
- Optical polarised phase function of the HR4796A dust ring (Milli et al.)
- CIELO-RGS: a catalog of soft X-ray ionized emission lines (Mao et al.)

LATEST VOLUME | All

Volume 627 (July 2019)
(open volume)



Sky fields for ESA's planet-hunter satellite PLATO (Marchiori, V., et al., 627, A71)

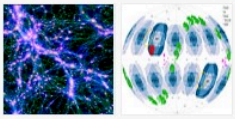
LATEST SPECIAL ISSUES | All

Rosetta 2 (open issue)

The Solar Orbiter mission (open issue)

Planck 2018 (open issue)

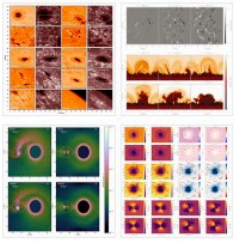
Latest articles | Latest volumes | Most downloaded articles



Volume 627 (July 2019) (open volume)

(1) Simulated magnetic fields (green), X-ray (orange) and radio (blue) emission from the cosmic web (Vazza, F., et al., 627, A5)

(2) Sky fields for ESA's planet-hunter satellite PLATO (Marchiori, V., et al., 627, A71)



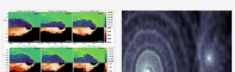
Volume 626 (June 2019)

(1) Centre-to-limb Ellerman bomb visibility in UV continua (Vissers, G. J. M.,626, A4)

(2) Magnetic field and temperature during flux emergence (Hansteen, V., et al., 626, A33)

(3) How rotation and binarity affect AGB star outflows (Saladino, M. I., et al., 626, A68)

(4) Three-dimensional neural tomography of a sunspot (Asensio Ramos, A., & Díaz Baso, C. J., 626, A102)



Volume 625 (May 2019)

A&A 627, A71 (2019)
https://doi.org/10.1051/0004-6361/201935269

Astronomy & Astrophysics

In-flight photometry extraction of PLATO targets

Optimal apertures for detecting extrasolar planets

V. Marchiori[1,2], R. Samadi[1], F. Fialho[2], C. Paproth[3], A. Santerne[4], M. Pertenais[3], A. Börner[3], J. Cabrera[5], A. Monsky[6], and N. Kutrowski[7]

[1] Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique, Observatoire de Paris, Université PSL, CNRS, 5 Pl. Jules Janssen, 92190 Meudon, France
e-mail: victor.marchiori@obspm.fr
[2] Escola Politécnica – Departamento de Engenharia de Telecomunicações e Controle, Universidade de São Paulo, Av. Prof. Luciano Gualberto, trav. 3, n. 158, 05508-010 São Paulo, Brazil
[3] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Optische Sensorsysteme, Rutherfordstraße 2, 12489 Berlin-Adlershof, Germany
[4] Aix Marseille Université, CNRS, CNES, LAM, Marseille, France
[5] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Planetenforschung, Rutherfordstraße 2, 12489 Berlin-Adlershof, Germany
[6] OHB System AG, Universitätsallee 27-29, 28359 Bremen, Germany
[7] Thales Alenia Space, 5 Allée des Gabians, 06150 Cannes, France

Received 14 February 2019 / Accepted 14 May 2019

ABSTRACT

Context. The ESA PLATO space mission is devoted to unveiling and characterizing new extrasolar planets and their host stars. This mission will encompass a very large (>2100 deg^2) field of view, granting it the potential to survey up to one million stars depending on the final observation strategy. The telemetry budget of the spacecraft cannot handle transmitting individual images for such a huge stellar sample at the right cadence, so the development of an appropriate strategy to perform on-board data reduction is mandatory.
Aims. We employ mask-based (aperture) photometry to produce stellar light curves in flight. Our aim is thus to find the mask model that optimizes the scientific performance of the reduced data.
Methods. We considered three distinct aperture models: binary mask, weighted Gaussian mask, and weighted gradient mask giving lowest noise-to-signal ratio, computed through a novel direct method. Each model was tested on synthetic images generated for 50 000 potential PLATO targets. We extracted the stellar population from the *Gaia* DR2 catalogue. An innovative criterion was adopted for choosing between different mask models. We designated as optimal the model providing the best compromise between sensitivity to detect true and false planet transits. We determined the optimal model based on simulated noise-to-signal ratio and frequency of threshold crossing events.
Results. Our results show that, although the binary mask statistically presents a few percent higher noise-to-signal ratio compared to weighted masks, both strategies have very similar efficiency in detecting legitimate planet transits. When it comes to avoiding spurious signals from contaminant stars however the binary mask statistically collects considerably less contaminant flux than weighted masks, thereby allowing the former to deliver up to ~30% less false transit signatures at 7.1σ detection threshold.
Conclusions. Our proposed approach for choosing apertures has been proven to be decisive for the determination of a mask model capable to provide near maximum planet yield and substantially reduced occurrence of false positives for the PLATO mission. Overall, this work constitutes an important step in the design of both on-board and on-ground science data processing pipelines.

Key words. instrumentation: photometers – planets and satellites: detection – techniques: photometric – methods: numerical – catalogs – zodiacal dust

1. Introduction

PLAnetary Transits and Oscillations of stars (PLATO)[1] Rauer et al. (2014) is a space mission from the European Space Agency (ESA) whose science objective is to discover and characterize new extrasolar planets and their host stars. Expected to be launched by end 2026, this mission will focus on finding photometric transit signatures of Earth-like planets orbiting the habitable zone of main-sequence Sun-like stars. Thanks to its very large field of view (~2132 deg^2) covered by multiple (6 to 24) telescopes, PLATO will be able to extract long duration (few months to several years) photometry from a significantly large sample of bright targets ($V < 11$) at very high photometric precision (~50 ppm h$^{1/2}$). The resulting scientific data are expected to provide stellar ages with accuracy as low as 10% and radii of Earth-like planets with accuracy as low as 3% (ESA 2017, see also Goupil 2017).

The PLATO data processing pipeline is a critical component of the payload, which is composed of multiple ground- and flight-based algorithms. These are necessary to convert the raw data acquired by the instrument, which inevitably carries unwanted systematic disturbances, into scientifically exploitable light curves. Typical examples of systematic errors are the long-term star position drift, pointing error due to satellite jitter,

[1] https://www.cosmos.esa.int/web/plato

charge transfer inefficiency (CTI) from the detectors, pixel saturation, outliers, and sky background. To work around these errors, extensive studies have been carried out focussed on the definition of data processing algorithms. These studies include the development of photometry extraction methods which are key for the success of the mission and motivate the present work.

The PLATO photometer will be capable to produce light curves for up to one million stars, depending on the final observation strategy. In contrast, transmitting individual images for each target, at sufficiently short cadence[2] for further ground-based processing requires prohibitive telemetry resources. Hence, for a substantial fraction of the targets, an appropriate data reduction strategy (prior to data compression) needs to be executed. In that case, the most suitable encountered solution consists in producing their light curves on board, in a similar way as performed for the targets of COnvection ROtation and planetary Transits (CoRoT; Auvergne et al. 2009) Space Telescope. By doing so, the spacecraft transmits data packages to the ground segment containing a single flux value per cadence for each star rather than multiple flux values from several pixels. Within the mission design of PLATO, the group of targets whose light curves will be produced on board are part of a stellar sample called P5. Considering a scenario of two long pointing observations, this set represents more than 245 000 F5 to late-K spectral class dwarf and sub-giant stars with V magnitude ranging from 8 to 13; it was idealized to generate large statistical information on planet occurrence rate and systems evolution. For all other stellar samples, which are primarily composed of the brightest targets (more details in ESA 2017), the photometry will be extracted on the ground from individual images, thereby following the same principle as that of *Kepler* Space Telescope (Borucki et al. 2010) and Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2014) targets.

In view of its acknowledged high performance and straightforward implementation, mask-based (aperture) was adopted as in-flight photometry extraction method to be implemented in the PLATO data processing pipeline. In such technique, each light curve sample is generated by integrating the target flux over a limited number of pixels, which shall be appropriately selected to maximize the scientific exploitability of the resulting time-series light curve. In this context, the present work unfolds the development carried out for defining the optimal collection of pixels for extracting photometry from non-saturated stars in the P5 sample.

There is a noteworthy number of publications on the theme of photometric masks. Among the oldest, we put some emphasis on the work of Howell (1989), in which the idea of a growth curve (signal-to-noise ratio as a function of aperture radius) for point-source observations is presented; on the stellar photometry package DAOPHOT[3] from Stetson (1987), which is still widely used today; and on the solution proposed by Naylor (1998), which consists of employing weighted masks for imaging photometry, providing improved noise-to-signal ratio (NSR) performance compared to binary masks. Later on, and orientated to planet transit finding and asteroseismology, Llebaria & Guterman (2006) and Bryson et al. (2010) developed strategies to compute optimized binary masks[4] for extracting light curves from CoRoT and *Kepler* targets, respectively. More recently, Smith et al. (2016) proposed a new method to assign apertures for *Kepler* targets, focussed on planet detection and mitigation of systematic errors, through an optimization scheme based on NSR and Combined Differential Photometric Precision (CDPP)[5] (Jenkins et al. 2010a). As described in *Kepler's* Data Processing Handbook[6], this method is implemented within the photometry analysis component of *Kepler's* science pipeline. Alongside, Aigrain et al. (2015) and Lund et al. (2015) provided techniques for mask pixel selection for *Kepler* K2 targets. The former proposes circular apertures, which has satisfactory performance for sufficiently bright targets and is relatively robust to systematic errors. The latter uses clustering of pixels, which best fits the flux distribution of the targets, being therefore more suitable for dense fields. A modified version of this method is employed in Handberg & Lund (2016) for reducing the data of *Kepler* K2 targets from campaigns 0 to 4. Besides, it is also considered as one of the possible solutions for extracting light curves from TESS targets (Lund et al. 2017).

In this paper, we are evidently interested in solutions that are better suited for both exoplanet search and asteroseismology, which brings thus our attention to those that were developed for the space missions CoRoT, *Kepler,* and TESS. Considering these three examples, we notice that the notion of optimal aperture is employed to distinguish apertures that minimize NSR or some noise-related metric such as CDPP. That is, of course, a reasonable way to proceed because the sensitivity at which a planet transit can be found in a light curve, for instance, is strongly correlated to its noise level. On the other hand, the higher the ease in identifying a transit-like signal, either because of sufficiently low NSR or CDPP, the higher the probability that a background object in the scene generates a threshold crossing event (TCE)[7]. This background object could be, for example, a stellar eclipsing binary (EB) mimicking a true planet transit. Background false positives may be efficiently identified in certain cases when, besides the light curves, the corresponding pixel data is also available, as demonstrated by Bryson et al. (2013); however, most of the stars in P5 unfortunately lack that extra information[8] because of telemetry constraints already mentioned. Under such an unfavourable scenario, conceiving photometric masks based uniquely on how well a transit-like signal can be detected, paying no attention to potential false positives may not be the best strategy. To verify the consistence of this hypothesis, we introduce in this paper two science metrics that allow us to directly quantify the sensitivity of an aperture in detecting true and false[9] planet transits. Then we determine whether or not the best compromise between these two parameters is obtained from apertures having overall lower NSR.

This paper is organized as follows (see Fig. 1). Section 2 describes the main payload characteristics, including instrument point spread function (PSF), spectral response, and noise. Also,

[2] Based on mission science requirements, PLATO light curves will be sampled at either 25, 50, or 600 s (see ESA 2017).

[3] http://www.star.bris.ac.uk/~mbt/daophot/.

[4] In the context of *Kepler's* data processing pipeline, such an aperture is referred to as simple aperture photometry. It was primarily designed to minimize noise for maximum transit detection sensitivity and as input for determining a halo of pixels to be downlinked along with the aperture pixels.

[5] Roughly speaking, CDPP is an estimate of how well a transit-like signal can be detected (Smith et al. 2016).

[6] https://archive.stsci.edu/kepler/manuals/

[7] This concept was created in the context of the *Kepler* science pipeline and designates a statistically significant transit-like signature marked for further data validation (e.g. see Twicken et al. 2018).

[8] For an observation scenario covering two long pointing fields, the telemetry budget dedicated to the P5 sample includes, in addition to the light curves, more than 9000 imagettes – at 25 s cadence – and centre of brightnesses (COB) for 5% of the targets (ESA 2017).

[9] In this paper, we address the occurrence of false planet transits caused by background eclipsing objects, in particular EBs.

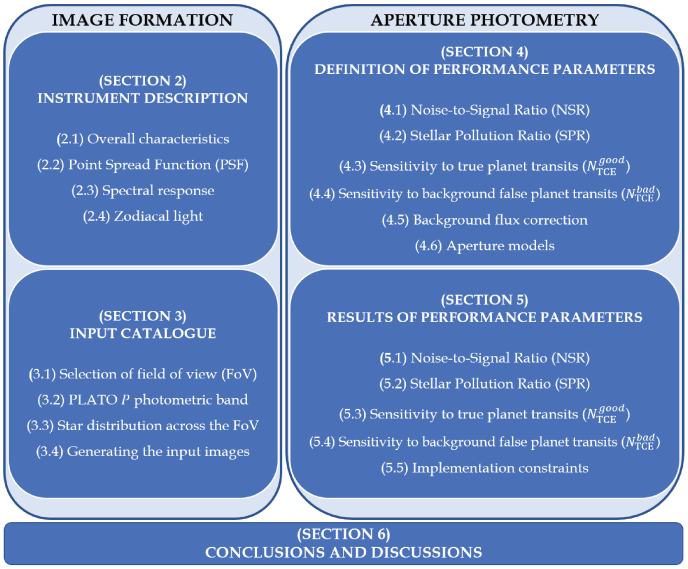


Fig. 1. Overview of paper content.

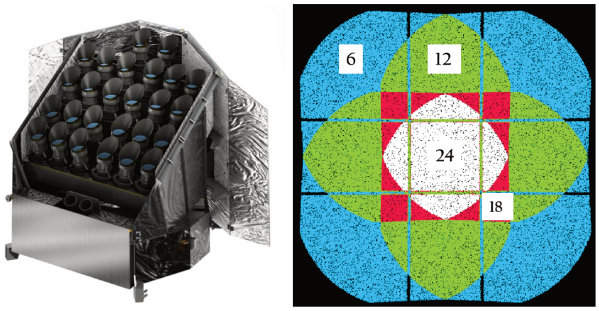


Fig. 2. *Left*: representation of the PLATO spacecraft with 24+2 telescopes. Credit: OHB-System AG. *Right*: layout of the resulting field of view obtained by grouping the normal telescopes into a 4×6 overlapping configuration. The colour code indicates the number of telescopes covering the corresponding fractional areas (Table 1): 24 (white), 18 (red), 12 (green), and 6 (blue). Credit: The PLATO Mission Consortium.

Table 1. Summary of main payload characteristics.

Description	Value
Optics	(24+2) telescopes with axisymmetric dioptric design
TOU spectral range	500–1000 nm
Pupil diameter (per telescope)	12 cm
Detector	Back-illuminated Teledyne-e2v CCD 270
N-CAM focal plane	4 full-frame CCDs (4510 × 4510 pixels each)
F-CAM focal plane	4 frame-transfer CCDs (4510 × 2255 pixels each)
Pixel length	18 μm
On-axis plate scale (pixel field of view)	15 arcsec
Quantization noise	~7.2 e^- rms px^{-1}
Readout noise (CCD+FEE) at beginning of life	~50.2 e^- rms px^{-1}
Detector smearing noise	~45 e^- px^{-1} s^{-1}
Detector dark current noise	~4.5 e^- px^{-1} s^{-1}
N-CAM cadence	25 s
N-CAM exposure time	21 s
N-CAM readout time	4 s
F-CAM cadence	2.5 s
N-CAM field of view	~1037 deg^2 (circular)
F-CAM field of view	~619 deg^2
Full field of view	~2132 deg^2
Fractional field of view	294 deg^2 (24 telescopes)
	171 deg^2 (18 telescopes)
	796 deg^2 (12 telescopes)
	871 deg^2 (6 telescopes)

an expression is derived to provide an estimation on the intensities of zodiacal light entering each PLATO telescope. Section 3 gives details on the extracted data from the adopted input catalogue (*Gaia* DR2). That information is used to build synthetic input images, called imagettes, to characterize the performance of aperture photometry. A synthetic PLATO *P* photometric passband, calibrated on the VEGAMAG system, is derived to avoid the inconvenience of having colour dependency when estimating stellar fluxes – at detector level – from visual magnitudes. Colour relationships with Johnson's *V* and *Gaia G* magnitudes are provided. Section 4 describes the methodology applied to find the optimal aperture model to extract photometry from stars in P5 sample. Three models are tested, including a novel direct method for computing a weighted aperture providing global lowest NSR. Section 5 shows comparative results between all aperture models with respect to their sensitivity in detecting true and false planet transits. Lastly, Sect. 6 concludes with discussions on the presented results.

2. The instrument

2.1. Overall characteristics

The PLATO payload relies on an innovative multi-telescope concept consisting of 26 small aperture (12 cm pupil diameter) and wide circular field of view (~1037 deg^2) telescopes mounted in a single optical bench. Each telescope is composed of an optical unit (TOU), a focal plane assembly holding the detectors, and a front-end electronics (FEE) unit. The whole set is divided into 4 groups of 6 telescopes (herein called normal telescopes or N-CAM) dedicated to the core science and 1 group of 2 telescopes (herein called fast telescopes or F-CAM) used as fine guidance sensors by the attitude and orbit control system. The normal telescope assembly results in a overlapped field of view arrangement (see Fig. 2), allowing them to cover a total sky extent of about 2132 deg^2, which represents almost 20 times the active field of the *Kepler* instrument. The N-CAM and F-CAM designs are essentially the same, except for their distinct readout cadence (25 and 2.5 s, respectively) and operating mode (full-frame and frame-transfer, respectively). In addition, each of the two F-CAM includes a bandpass filter (one bluish and the other reddish) for measuring stellar flux in two distinct wavelength bands. Table 1 gives an overview of the main payload characteristics based on ESA (2017).

2.2. Point spread function

Starlight reaching the focal plane of PLATO cameras will inevitably suffer from distortions caused by both optics and detectors, causing this signal to be non-homogeneously spread

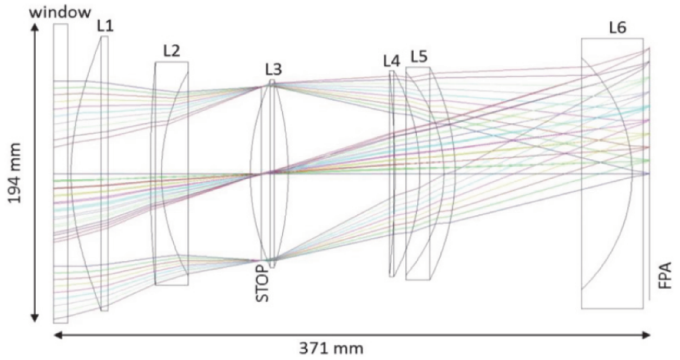


Fig. 3. Baseline optical layout of each PLATO telescope. Credit: The PLATO Mission Consortium.

out over several pixels. The physical model describing such effects is the PSF, from which we can determine – at sub-pixel level – how stellar signals are distributed over the pixels of the detector. In this work, we used synthetic optical PSF models obtained from the baseline telescope optical layout (Fig. 3) simulated on ZEMAX® software. Estimated assembly errors such as lens misalignment and focal plane defocus are included.

Beyond optics, the detectors also degrade the spatial resolution of stellar images through charge disturbances processes such as CTI (Short et al. 2013; Massey et al. 2014), "brighter-fatter" (Guyonnet et al. 2015), and diffusion (Widenhorn 2010). Several tests are being carried out by ESA to characterize such effects for the charge coupled devices (CCD) of PLATO cameras, so at the present date no formal specifications for the corresponding parameters are available. However, the optical PSFs alone are known to be a non-realistic final representation of the star signals. Therefore, to obtain a first order approximation of the real physics behind the PSF enlargement taking place at PLATO detectors with respect to the diffusion, the optical PSFs are convolved to a Gaussian kernel with a standard deviation of 0.2 pixel. The resulting simulated PSF models are shown in Fig. 4 for 15 angular positions, α, within the field of view of one camera. In this paper, PSF shape variations due to target colour are assumed to be of second order and are thus ignored.

To reduce the overlap of multiple stellar signals and increase photometric precision, PLATO cameras are primarily designed to ensure that about 77% of the PSF flux is enclosed, on average, within $\sim 2.5 \times 2.5$ pixels across the field of view, or 99% within $\sim 5 \times 5$ pixels. As a consequence, the size of the pixels are relatively large compared to that of the PSF, making the distribution of energy from stars very sensitive to their barycentre location within a pixel (see Fig. 5).

During and after launch, the space environment unavoidably causes overall changes in the instrument response that cannot always be accurately predicted, including variations in the PSF model. Nevertheless, accurate knowledge of the PSFs is imperative for proper correction of systematic errors in the light curves and computing the photometric apertures, so a strategy for reconstructing the PSFs is needed. As the individual raw images downlinked from the spacecraft cannot describe the distribution of stellar flux on the detectors with sufficient resolution, high resolution PSFs such as those in Fig. 4 will be reconstructed on the ground from micro-scanning sessions (Samadi et al. 2019). This process, which will occur every three months during instrument calibration phases, basically consists of acquiring

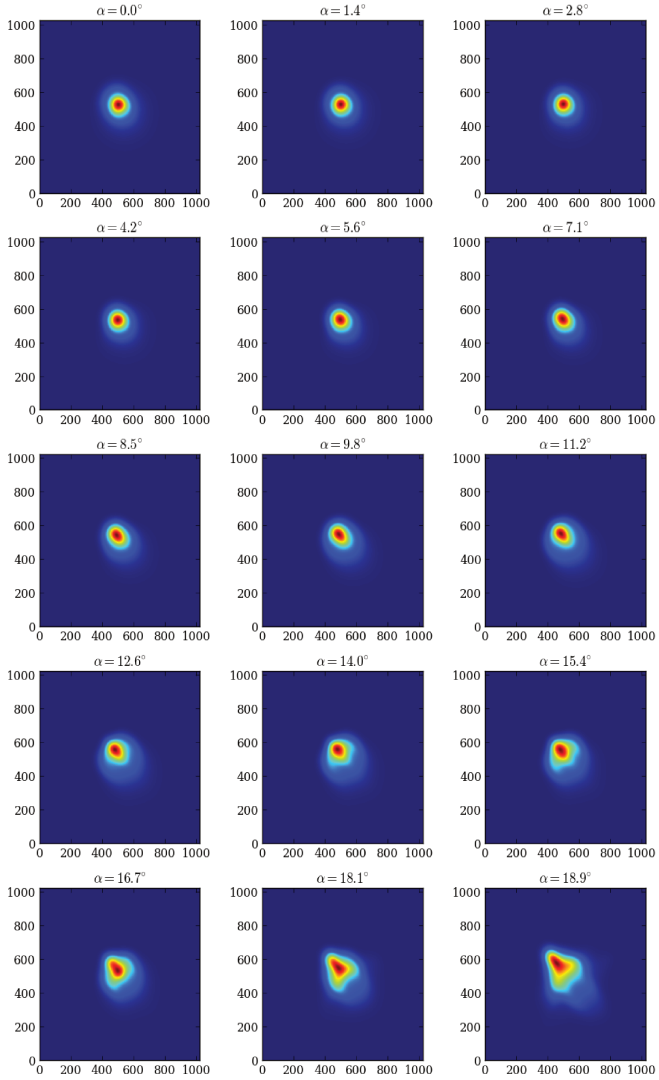


Fig. 4. Simulated PSF shapes of PLATO telescopes (1/128 pixel resolution) as a function of the angular position, α, in the sky of a source at $-45°$ azimuth. Angular positions range from $\alpha = 0°$ (centre) to $\alpha = 18.9°$ (edge) of the camera field of view. Each optical PSF is convolved to a Gaussian diffusion kernel with standard deviation of 0.2 pixel to simulate the CCD behaviour. Each image above corresponds to a CCD surface of 8×8 pixels, which is enough to encompass ~99.99% of the total PSF energy.

a series of raw images from subpixel displacements following an Archimedean spiral. Then, inverse methods are employed to reconstruct the PSFs from the lower resolution micro-scanning images.

2.3. Spectral response and vignetting

The spectral response of a photometer represents its efficiency in converting incident photons into effective counts of electrons at detector level. It is derived from the combined effect of optical transmission and CCD quantum efficiency. In parallel, another parameter impacting instrument efficiency is vignetting, which consists of an inherent optical feature that causes attenuation of image brightness. Such an effect increases non-linearly as the angular position, α, of the source augments with respect to the optical axis ($\alpha = 0$) of the instrument. A preliminary spectral response curve of the PLATO cameras is presented in Fig. 6,

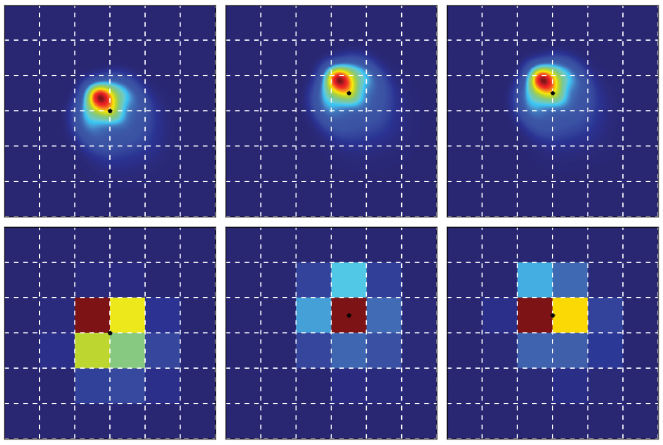


Fig. 5. Energy distribution of PSF across the pixels for three distinct intra-pixel target barycentre locations (black dots): at pixel corner (*left*), at pixel centre (*middle*), and at the border of two adjacent pixels (*right*). Dashed white lines represent pixel borders. *Top*: high resolution PSF at $\alpha = 14°$. *Bottom*: corresponding low resolution PSF.

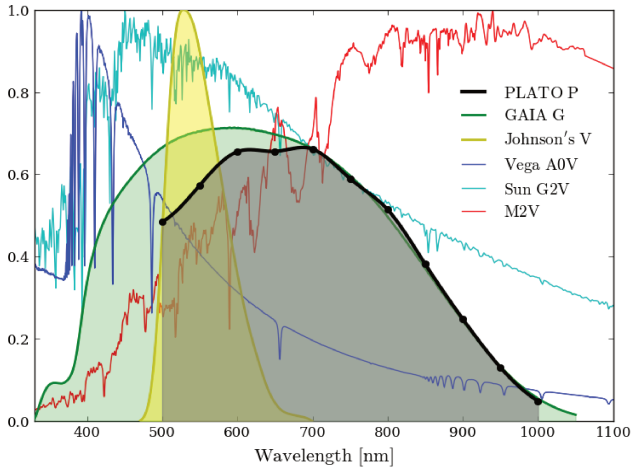


Fig. 6. Black: preliminary spectral response of PLATO N-CAM at beginning of life. Values are currently known at the black dots; these are crossed by a cubic spline interpolation curve. Green: *Gaia G* band. Yellow: Johnson's *V* filter. Blue: Vega (A0V) normalized spectrum. Cyan: Sun (G2V) normalized spectrum. Red: normalized spectrum of M2V-type star.

alongside the *Gaia G* passband[10], Johnson's *V* filter from Bessell (1990), Vega A0V Kurucz template (`alpha_lyr_stis_008`) from the CALSPEC[11] database, E-490[12] reference solar spectrum from the American Society for Testing and Materials (ASTM), and a M2V-type star synthetic spectrum from the Pickles atlas[13] (Pickles 1998). Vignetting intensities, $\mathrm{f_{vig}}$, are given in Table 2 as a function of the off-axis angle, α, of the target.

2.4. Zodiacal light

Scattered sky background light account for significant noise contribution, impacting photometry performance. Hence, we are interested in estimating the amount of diffuse background flux captured by the detectors of each PLATO camera. As the spacecraft will be positioned in L2 orbit (located at approximately 1.01 au from the Sun), sky background flux entering its cameras will be dominated by the zodiacal light, i.e. sunlight scattered by interplanetary dust particles agglomerated across the ecliptic plane. Zodiacal light brightness is conventionally expressed in counts of 10th visual magnitude solar-type stars per square degree, also known as $\mathrm{S_{10}}$ unit. By denoting $F_\odot(\lambda)$ as the solar spectrum at 1 au and adopting a corresponding apparent visual magnitude $V_{\mathrm{Sun}} = -26.74$ mag, the $\mathrm{S_{10}}$ unit is formally defined as

$$\mathrm{S_{10}} = 10^{-0.4(10-V_{\mathrm{Sun}})} F_\odot(\lambda)\ \mathrm{deg^{-2}} = 6.61 \times 10^{-12} F_\odot(\lambda)\ \mathrm{sr^{-1}}. \quad (1)$$

Tabular data containing zodiacal light measurements in $\mathrm{S_{10}}$ units are available in Leinert et al. (1998). The published values are valid for an observer located in the vicinity of Earth and at monochromatic wavelength (500 nm). Outside these conditions, a semi-analytical model containing a few correction factors shall be applied. Based on that model we have built an expression (see Table 3 for parameters description) to estimate the amount of zodiacal light flux f_{ZL}^P on one N-CAM

$$f_{\mathrm{ZL}}^P = f_{\mathrm{ZL}} \left(6.61 \times 10^{-12}\ \mathrm{sr^{-1}}\right) (h\,c)^{-1}\, \Omega\, \Theta\, \mathrm{f_{L2}} \left(1 - \mathrm{f_{vig}}\right) \times \times \int_{\lambda_1}^{\lambda_2} F_\odot(\lambda)\, \mathrm{f_{red}}(\lambda)\, S(\lambda)\, \lambda\, \mathrm{d}\lambda. \quad (2)$$

Modelling $F_\odot(\lambda)$ with ASTM's E-490 reference solar spectrum (see Fig. 6) results (expressed in units of $\mathrm{e^-\ px^{-1}\ s^{-1}}$) in

$$f_{\mathrm{ZL}}^P \sim 0.39\, f_{\mathrm{ZL}} \left(1 - \mathrm{f_{vig}}\right). \quad (3)$$

Excluding vignetting ($f_{\mathrm{vig}} = 0$), 1 $\mathrm{S_{10}}$ of zodiacal light ($f_{\mathrm{ZL}} = 1$) corresponds thus to about $0.39\ \mathrm{e^-\ px^{-1}\ s^{-1}}$ being generated at the detectors of one PLATO camera. We note that this value might be updated in the future, depending on the final spectral response of the instrument.

Table 2. Combined natural and mechanical obscuration vignetting, f_{vig}, as a function of the off-axis angle, α, of the target.

α (deg)	f_{vig} (%)	α (deg)	f_{vig} (%)	α (deg)	f_{vig} (%)
0.0	0.0	7.053	1.51	14.001	5.85
1.414	0.06	8.454	2.16	15.370	7.03
2.827	0.24	9.850	2.93	16.730	8.53
4.238	0.55	11.241	3.80	18.081	11.58
5.647	0.97	12.625	4.78	18.887	13.69

Table 3. Description of the parameters of Eq. (2).

Description	Symbol	Value	Unit
Zodiacal light tabulated data	f_{ZL}	see Leinert et al. (1998)	$\mathrm{S_{10}}$
Planck's constant	h	6.63×10^{-34}	J s
Speed of light in vacuum	c	2.99×10^{8}	$\mathrm{m\,s^{-1}}$
Field of view solid angle (per pixel)	Ω	4.2×10^{-9}	sr
Entrance pupil surface (per camera)	Θ	113.1	$\mathrm{cm^2}$
Spectral range of PLATO telescopes	$[\lambda_1, \lambda_2]$	[500, 1000]	nm
Sun's spectral irradiance	$F_\odot(\lambda)$	E-490 spectrum	$\mathrm{W\,cm^{-2}\,nm^{-1}}$
Correction factor for L2 point	$\mathrm{f_{L2}}$	0.975	adim
Instrument vignetting	$\mathrm{f_{vig}}$	see Table 2	adim
Reddening correction factor	$\mathrm{f_{red}}(\lambda)$	see Leinert et al. (1998)	adim
Spectral response of PLATO telescopes	$S(\lambda)$	see Fig. 6	adim

[10] https://www.cosmos.esa.int/web/gaia/auxiliary-data
[11] http://www.stsci.edu/hst/observatory/crds/calspec.html
[12] https://www.nrel.gov/grid/data-tools.html
[13] http://www.stsci.edu/hst/observatory/crds/pickles_atlas.html

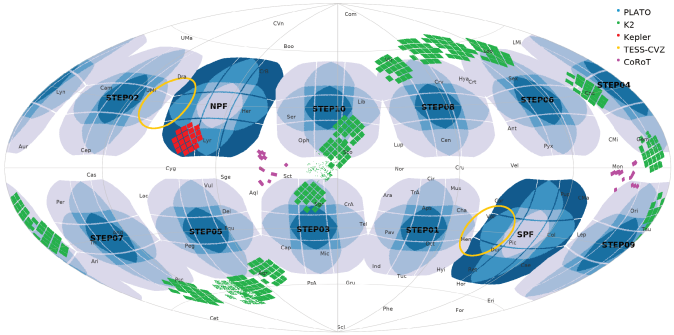


Fig. 7. Sky coverage in Galactic coordinates of PLATO's provisional SPF and NPF long-duration LOP fields, including the possible locations of the short-duration SOP fields (STEP 01-10), to be defined two years before launch. The illustration also shows some sky areas covered by the surveys: *Kepler* (red), *Kepler*-K2 (green), TESS (Continuous Viewing Zones-CVZ; yellow) and CoRoT (magenta). Courtesy of Valerio Nascimbeni (INAF-OAPD, Italy), on behalf of the PLATO Mission Consortium.

3. Input stellar catalogue

An input stellar catalogue is an essential tool for space missions dedicated to asteroseismology and exoplanet searches. Besides its crucial role in field and target selection, it is also noticeably useful for estimating and characterizing the performance of photometry extraction methods prior to mission launch. Indeed, an input catalogue allows us to produce synthetic sky images containing realistic stellar distributions, including their relative positions, apparent magnitudes, effective temperatures, gravities, metallicities, and more. At the present date, a PLATO Input Catalogue (PIC) is being developed based on the ultra-high precision astrometric data from the *Gaia* mission (Gaia Collaboration 2016). In the future, the PIC might also include information available from other sky surveys such as the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008), Panoramic Survey Telescope and Rapid Response System (PanSTARSS; Chambers et al. 2016), and SkyMapper (Wolf et al. 2018). The PIC will provide abundant and detailed stellar information for optimized target selection vis-à-vis mission science goals. As the PIC was not yet available[14] by the time that the present work was started, we have adopted the *Gaia* data release 2 (DR2; Gaia Collaboration 2018) as our input catalogue, which provides all the information needed for the present work.

3.1. Observing strategy and input field selection

With an nominal mission duration of four years, two observation scenarios are considered for PLATO. The first consists of two long-duration (2+2 years) observation phases (LOP) with distinct sky fields. The second consists of a single LOP of three years plus one step-and-stare operation phase (SOP) covering multiple fields lasting a few months each. Mission design constraints require the LOP fields to have absolute ecliptic latitude and declination above 63° and 40°, respectively. Under such conditions, two LOP fields are actually envisaged: a southern PLATO field (SPF) centred at Galactic coordinates $l = 253°$ and $b = -30°$ (towards the Pictor constellation) and a northern PLATO field (NPF) centred at $l = 65°$ and $b = 30°$ (towards the Lyra and Hercules constellations and also including the *Kepler* target field). An illustration containing the locations of both SPF and NPF is shown in Fig. 7, as well as the possible locations for the SOP fields, which will be defined two years before launch (ESA 2017). For the purposes of this paper, we adopted as input field (IF) the fraction of SPF that is equivalent to the area covered by a single PLATO camera centred at SPF centre. That represents roughly half of the SPF area in the sky and encompasses about 12.8 million stars listed in the *Gaia* DR2 catalogue with G magnitude comprised between 2.45 and 21. Table 4 presents, in different reference systems, the coordinates of IF centre, hereafter referred to as IF line of sight ($IF_{\rm LoS}$). The sky area covered by IF in ecliptic coordinates with zero point $\phi_\odot$ in the Sun is illustrated in Fig. 8. We also present in the this figure the zodiacal light (see model description in Sect. 2.4) perceived by an observer located in L2 orbit and pointing towards our coordinate zero point.

[14] An early version of the PIC is currently available for PLATO consortium members upon request to the PLATO Data Center Board.

Table 4. Coordinates of the input field line of sight ($IF_{\rm LoS}$) in different reference systems.

Reference system	Longitude	Latitude
Galactic	$l_{\rm LoS} = 253°$	$b_{\rm LoS} = -30°$
Equatorial	$\alpha_{\rm LoS} = 86.80°$	$\delta_{\rm LoS} = -46.40°$
Ecliptic	$\phi_{\rm LoS} = 83.62°$	$\beta_{\rm LoS} = -69.77°$

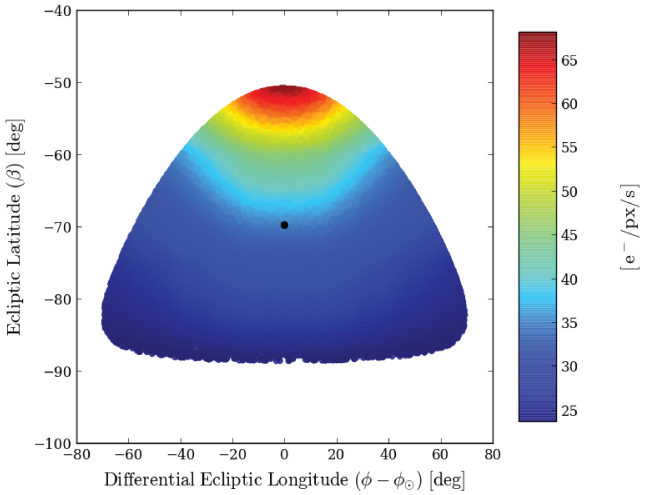


Fig. 8. Scatter plot of zodiacal light across IF in differential ecliptic coordinates centred on the Sun. Values are valid for an observer in L2 orbit having the Sun's ecliptic longitude $\phi_\odot$ aligned with $\phi_{\rm LoS} = 83.62°$. Instrument vignetting is not included ($f_{\rm vig} = 0$).

3.2. Synthetic PLATO P photometric passband

3.2.1. Definition and relationship with V band

The PLATO mission was designed based on stellar magnitudes specified in the visible band. Nevertheless, to avoid the inconvenience of having colour dependency when estimating stellar fluxes at detector level, from the visual magnitudes, it is more appropriate to work in a proper instrument photometric band. Therefore, we build in this paper a synthetic P magnitude calibrated in the VEGAMAG system

$$P = -2.5 \log_{10} \left(\frac{\int_{\lambda_1}^{\lambda_2} f(\lambda)\, S(\lambda)\, \lambda\, {\rm d}\lambda}{\int_{\lambda_1}^{\lambda_2} f_{\rm Vega}(\lambda)\, S(\lambda)\, \lambda\, {\rm d}\lambda} \right) + P_{\rm Vega}, \quad (4)$$

Table 5. Normal camera (N-CAM) predicted flux, $f_P^{\rm ref}$, for a reference 6000 K G0V star as a function of its V and P magnitudes.

V (mag)	P (mag)	$f_P^{\rm ref}$ (per camera) (10^3 e$^-$ s^{-1})	$f_P^{\rm ref}$ (24 cameras) (10^3 e$^-$ s^{-1})
8.0	7.66	143.820	3451.7
8.5	8.16	90.745	2177.9
9.0	8.66	57.256	1,374.1
9.5	9.16	36.126	867.0
10.0	9.66	22.794	547.1
10.5	10.16	14.382	345.2
11.0	10.66	9.074	217.8
11.5	11.16	5.726	137.4
12.0	11.66	3.613	86.7
12.5	12.16	2.279	54.7
13.0	12.66	1.438	34.5

Notes. Values include vignetting for a source at $\alpha = 14°$ (Table 2) and are consistent with the current instrument design.

where $f(\lambda)$ is the spectral flux of a given star, $f_{\rm Vega}(\lambda)$ is the spectral flux of the Vega A0V type star (see Fig. 6), and $P_{\rm Vega}$ is its magnitude in the P band, the latter assumed to be equal to $V_{\rm Vega} = 0.023$ mag (Bohlin 2007). The P band zero point is given by

$$zp = 2.5 \log_{10}\left((h\,c)^{-1}\,\Theta \int_{\lambda_1}^{\lambda_2} f_{\rm Vega}(\lambda)\,S(\lambda)\,\lambda\,{\rm d}\lambda\right) + P_{\rm Vega}. \quad (5)$$

This constant (see Table 6) provides a straightforward way for switching between stellar flux and magnitudes using

$$P = -2.5 \log_{10}\left((h\,c)^{-1}\,\Theta \int_{\lambda_1}^{\lambda_2} f(\lambda)\,S(\lambda)\,\lambda\,{\rm d}\lambda\right) + zp. \quad (6)$$

Thus, having the zero point zp and the magnitude P of a given star, its respective total flux f_P (per camera and expressed in units of e$^-$ s^{-1}) can be estimated with

$$f_P = 10^{-0.4(P-zp)}. \quad (7)$$

For switching between P and V magnitudes, we determine the $V-P$ relationship using the Johnson-Cousins V filter (Fig. 6) and modelling $f(\lambda)$ with synthetic stellar spectra extracted from the POLLUX database (Palacios et al. 2010). As for the calibration star Vega, we adopted the template `alpha_lyr_stis_008` (Fig. 6) from CALSPEC. The resulting $V-P$ samples are shown in Fig. 9 as a function of the effective temperature $T_{\rm eff}$, the latter ranging from 4000 K to 15 000 K in steps of 500 K. The corresponding fitted polynomial is

$$V - P = -1.184 \times 10^{-12}(T_{\rm eff})^3 + 4.526 \times 10^{-8}(T_{\rm eff})^2 - 5.805 \times 10^{-4} T_{\rm eff} + 2.449. \quad (8)$$

Therefore, for a star with specified visual magnitude and $T_{\rm eff}$, we can determine its P magnitude with Eq. (8) and then applied Eq. (7) to estimate the respective flux at detector level. Table 5 shows the predicted flux $f_P^{\rm ref}$ for a reference PLATO target, i.e. a 6000 K G0V spectral type star, as a function of its V and P magnitudes. The values include brightness attenuation due to vignetting for a source at $\alpha = 14°$. In this scenario, a reference PLATO star with $V = 11$ has $P = 10.66$ and $f_P^{\rm ref} = 9.074$ ke$^-$ s^{-1} per camera, or ∼218 ke$^-$ s^{-1} when cumulating over 24 cameras.

Table 6. Zero points zp of our synthetic P, G, $G_{\rm BP}$, and $G_{\rm RP}$ photometric passbands calibrated with Vega `alpha_lyr_stis_008` model.

Synthetic passband	Vega (mag)	zp (mag)	zp dev. (A) (mag)	zp dev. (B) (mag)
P	0.023	20.62		
G	0.029	25.6879	4.6×10^{-4}	4.70×10^{-2}
$G_{\rm BP}$	0.039	25.3510	4.3×10^{-4}	1.10×10^{-2}
$G_{\rm RP}$	0.023	24.7450	1.69×10^{-2}	1.50×10^{-2}

Notes. Vega magnitudes for *Gaia* passbands are extracted from Casagrande & VandenBerg (2018). Absolute deviations (zp dev.) of G, $G_{\rm BP}$, and $G_{\rm RP}$ zero points are computed with respect to the reference DR2 magnitudes presented in Evans et al. (2018) (A) and the revised versions in Weiler (2018) (B).

3.2.2. Obtaining *P* and *V* from *Gaia* magnitudes

We also need to determine expressions for converting from the magnitude scales available in our adopted input catalogue (*Gaia* DR2) to our synthetic V and P magnitudes. *Gaia* collects data in three photometric systems: G, $G_{\rm BP}$, and $G_{\rm RP}$. As defined in Jordi et al. (2010), all of these systems are calibrated in the VEGAMAG system, following therefore the same philosophy as Eqs. (4)–(6). To keep consistency with our previously adopted V and P bands, we applied the same Vega model to derive synthetic calibrations for the three *Gaia* bands. Consequently, we imposed to the latter the corresponding Vega magnitudes listed in Casagrande & VandenBerg (2018). Table 6 summarizes the obtained zero points for our synthetic G, $G_{\rm BP}$, and $G_{\rm RP}$ bands from this approach. They present satisfactorily low deviations with respect to the reference DR2 magnitudes published in Evans et al. (2018) and the later improved versions in Weiler (2018). Then, to obtain both P and V magnitudes from the *Gaia* G band, we determined $G-P$ and $V-P$ relationships by means of the $G_{\rm BP}-G_{\rm RP}$ colour index, resulting in the plots shown in Fig. 9. The corresponding fitted polynomials, within the range $-0.227 \le G_{\rm BP}-G_{\rm RP} \le 4.524$, are

$$G - P = 0.00652\,(G_{\rm BP}-G_{\rm RP})^3 - 0.08863\,(G_{\rm BP}-G_{\rm RP})^2 + 0.37112\,(G_{\rm BP}-G_{\rm RP}) + 0.00895; \quad (9)$$

$$V - P = -0.00292\,(G_{\rm BP}-G_{\rm RP})^3 + 0.10027\,(G_{\rm BP}-G_{\rm RP})^2 + 0.37919\,(G_{\rm BP}-G_{\rm RP}) + 0.00267. \quad (10)$$

Unlike Eq. (8), the expressions in Eqs. (9) and (10) are described as a function of the $G_{\rm BP}-G_{\rm RP}$ colour index, rather than the $T_{\rm eff}$. The reason for that is the low availability of effective temperatures in DR2 (less than 10% of the sources). In contrast, $G_{\rm BP}$ and $G_{\rm RP}$ magnitudes are simultaneously available for more than 80% of the sources. To verify the consistency of our synthetic calibrations derived from synthetic stellar spectra, we compared our $V-G = (V-P)-(G-P)$ relationship with the $V-G$ polynomial fit (Busso et al. 2018) derived from Landolt[15] standard stars (398 sources) observed with *Gaia*. As shown in Fig. 9, our synthetic $V-G$ curve exhibits satisfactory agreement with the $V-G$ polynomial fit obtained from the true *Gaia* observations. The maximum absolute error between both curves is 9.8×10^{-2} mag at $G_{\rm BP}-G_{\rm RP} = 2.75$ mag. Hence, for the purposes of this paper, we consider that the polynomials of Eqs. (9) and (10) give sufficiently accurate estimates of P and V magnitudes from the G magnitude of the DR2 catalogue.

[15] https://www.eso.org/sci/observing/tools/standards/Landolt.html

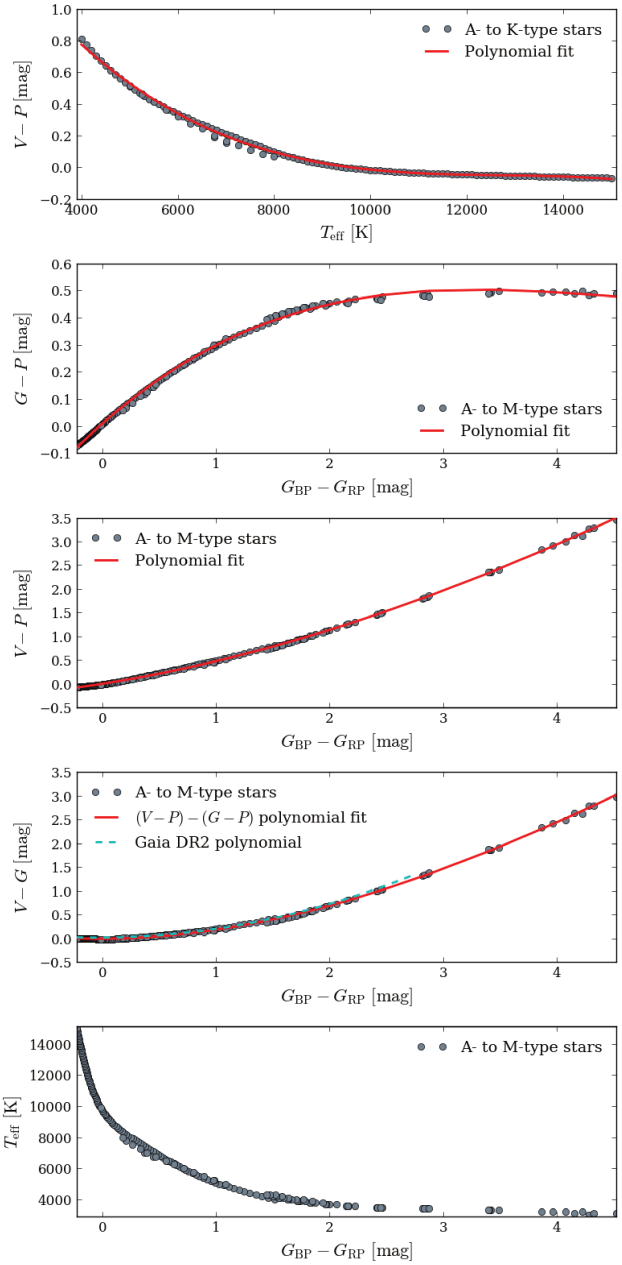


Fig. 9. Relationships between the photometric passbands V, P, and G, obtained by modelling $f(\lambda)$ with A- to M-type synthetic stellar spectra extracted from the POLLUX database. Red polynomials are derived from Eqs. (8)–(10), and are applicable within the range $4000 \leq T_{\rm eff} \leq 15000$ (top frame) and $-0.227 \leq G_{\rm BP} - G_{\rm RP} \leq 4.524$. *Gaia* DR2 polynomial is based on Landolt stars observed with *Gaia* and is applicable within the range $-0.5 \leq G_{\rm BP} - G_{\rm RP} \leq 2.75$. It has a scatter of 4.6×10^{-2} mag (see Busso et al. 2018). The relationship between $T_{\rm eff}$ and our synthetic $G_{\rm BP} - G_{\rm RP}$ color index (bottom frame) is consistent with the color–temperature relations published in Andrae et al. (2018).

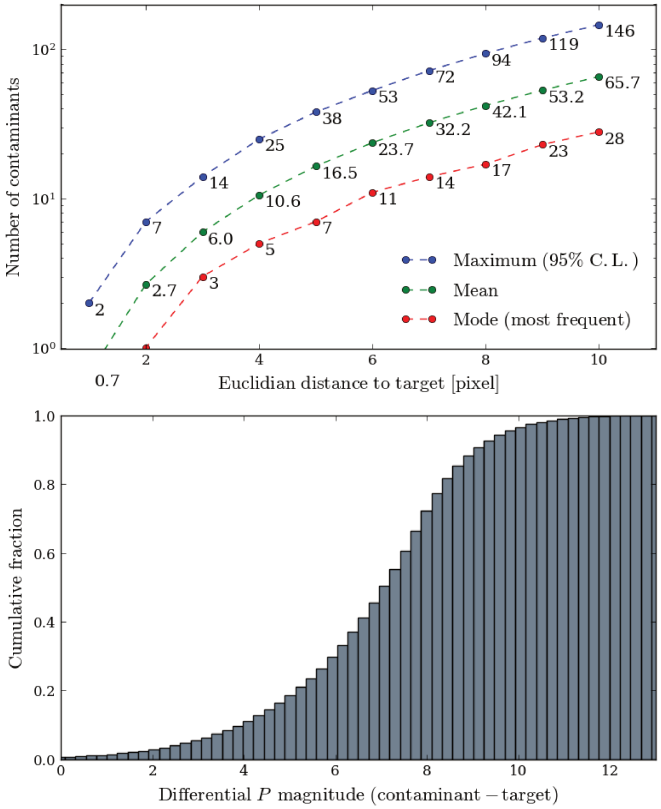


Fig. 10. *Top*: number of contaminants brighter than $P \sim 21.1$ as a function of the Euclidean distance from the target stars ($7.66 \leq P \leq 12.66$). Maximum values have 95% confidence level. *Bottom*: cumulative fraction of the differential P magnitude between contaminant and target stars, the former located at up to 10 pixels in distance from the latter.

3.3. Identifying target and contaminant stars

We define in this section the ensemble of target and contaminant stars from the input catalogue that will be used to build input images for simulating aperture photometry. First, we determined the position of each star within IF at the focal plane array of one PLATO camera (as explained in Sect. 3.1, IF covers exactly the field of a single camera). Next, following the definition of the P5 sample, we assigned as targets those stars located within IF that have $0.57 \leq G_{\rm BP} - G_{\rm RP} \leq 1.84$ (F5 to late-K spectral types) and P magnitude in the range $7.66 \leq P \leq 12.66$, the latter corresponding to $8.0 \leq V \leq 13.0$ for a reference PLATO target, i.e. a 6000 K G0V star. This accounts for about 127 000 sources. Target selection based on the P band is more convenient than the V band, as it allows us to overcome the colour dependency of the latter. In other words, this approach ensures that all targets assume flux values within a fixed range (that of Table 5), regardless of their effective temperature. This is thus consistent with a target selection strategy driven by noise performance, magnitude, and spectral type. As for the contaminant stars, they correspond to all existing sources in the input catalogue located within 10 pixel radius around all targets. This accounts for about 8.3 million stars with P magnitude comprised in the range $2.1 \lesssim P \lesssim 21.1$. It is important to mention that only sources satisfying $-0.227 \leq G_{\rm BP} - G_{\rm RP} \leq 4.524$ are included in the ensemble of contaminant stars to conform with the range of applicability of the polynomials described in Eqs. (9) and (10). According to the above description, Fig. 10 presents some statistics (distances and differential magnitudes) on the distribution of contaminant stars relative to their corresponding targets.

A few considerations are necessary concerning the sources in *Gaia* DR2. Evans et al. (2018) reported some very likely saturation and imperfect background subtraction issues affecting sources with $G \lesssim 3.5$ and $G \gtrsim 17$, respectively. Since the central point in this work is to establish a relative performance comparison between different photometric aperture models – particularly in scenarios of high stellar crowding – we decided to not remove those sources from our working subset of stars. The inaccuracies resulting from the mentioned issues will ultimately be evenly

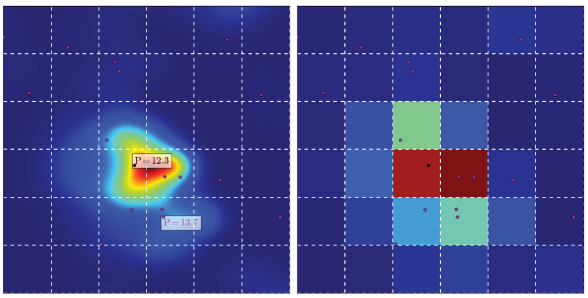


Fig. 11. Example of input image. *Left*: high resolution PSF ($\alpha = 18°$) for a target with $P = 12.3$ (barycentre designated by the black dot) surrounded by several contaminants (respective barycentres designated by the magenta dots). The brightest contaminant in the frame (tagged below the target) has $P = 13.7$. All other contaminants are at least 2 mag fainter than the target. *Right*: corresponding low resolution PSF (imagette) at pixel level. Zodiacal light is not shown in the scene. Dashed white lines represent pixel borders.

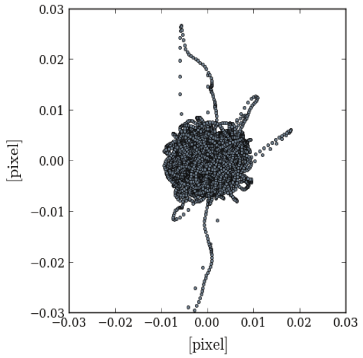


Fig. 12. Illustration of a possible star position motion on the PLATO focal plane. We use as input a simulated time series of 1 h duration sampled at 8 Hz based on the current status of the pointing requirements. The Euclidian distance scatter is 2.25 mpixel rms with respect to the nominal (zero) position. Credit of the time series simulation: PLATO Industrial Core Team (OHB-System AG, TAS, RUAG Space).

propagated to all tested mask models, having therefore no potential to significantly impact the comparative basis analysis.

3.4. Setting up the input imagettes

During science observations, in-flight photometry extraction will be performed independently for each target by integrating its flux over a set of selected pixels (aperture or mask). Such pixel collection is to be chosen from a 6×6 pixels square window called imagette, assigned uniquely to each target. An imagette is sufficiently large to encompass the near totality (~99.99%) of the energy from the corresponding target. Characterizing the expected performance of mask-based photometry therefore requires building up such imagettes, which shall be composed of realistic stellar content (targets and respective contaminants). To do so, we applied the following procedure:

1. Use the input catalogue and derived properties to obtain the magnitudes, fluxes, and locations of target and contaminant stars at intrapixel level. From that, we consider a random stellar subset composed of 50 000 targets (from the total of ~127k potential targets within IF). These are neighboured by ~3.25 million contaminant stars.
2. Employ the PSFs presented in Sect. 2.2 as models of both instrument optical and detector responses to stellar flux.
3. By convention, the pixels of an imagette are selected such that the centre of the resulting imagette is located at no more than an absolute Euclidean distance of 0.5 pixel from the target barycentre (see examples in Fig. 5). This is done to maximize the amount of target energy falling within its imagette.
4. Translate each sample of the satellite pointing time series from Fig. 12 into a corresponding shifted imagette with respect to the nominal position (zero). These are used as input to compute the jitter noise in the photometry.

Following this process, an input image (reference frame) like that illustrated in Fig. 11 was generated for each target (including respective contaminants). Shifted images to account for satellite motion are therefore computed target by target with respect to their respective reference frames. To increase simulation speed, the jitter time series was down-sampled by a factor of 10, resulting in a 0.8 Hz signal keeping the same statistical properties (mean, variance, and spectral energy distribution) as the original signal. Based on that, a total of 2880 shifted images (1 h duration signal) were produced per reference image, resulting in a total of $50\,000 \times (2880 + 1) \sim 144 \times 10^6$ synthetic imagettes. These are used as input for a detailed and realistic characterization of the performance expected from aperture photometry.

4. Aperture photometry

In order to find the optimal aperture model for extracting photometry from PLATO P5 targets, we applied the following steps:

i. Formalize an expression for the NSR.
ii. Define an expression for estimating the fractional flux from contaminant stars entering the apertures. This parameter is referred to as stellar pollution ratio (SPR).
iii. Build different aperture models based on NSR and width.
iv. Apply each mask model to the input images generated as described in Sect. 3.4.
v. Compute, for each mask model, the number of target stars for which an Earth-like planet orbiting it would be detected, i.e., trigger a TCE. This parameter is referred to as $N_{\rm TCE}^{\rm good}$.
vi. Compute, for each mask model, the number of contaminant stars that are likely to produce, whenever they are eclipsed, background false positives. This parameter is referred to as $N_{\rm TCE}^{\rm bad}$.

The above steps are detailed throughout the next sections in this chapter. Next, the performance of the different mask models are compared and commented on detail in Sect. 5.

4.1. Noise-to-signal ratio

The NSR is the principal performance indicator for evaluating the exploitability of photometry signals. For PLATO stellar light curves derived from aperture photometry applied to imagettes, we used the following metric to compute the per cadence NSR (NSR_*; see parameters description in Table 7):

$$\mathrm{NSR}_* = \frac{\sqrt{\sum_{n=1}^{36}\left(\sigma^2_{F_{\mathrm{T_n}}} + \sum_{k=1}^{N_{\mathrm{C}}}\sigma^2_{F_{\mathrm{C_{n,k}}}} + \sigma^2_{B_{\mathrm{n}}} + \sigma^2_{D_{\mathrm{n}}} + \sigma^2_{Q_{\mathrm{n}}}\right) w_{\mathrm{n}}^2}}{\sum_{n=1}^{36} F_{\mathrm{T_n}}\, w_{\mathrm{n}}}. \quad (11)$$

A per cadence light curve sample corresponds to the integrated mask flux over one exposure interval of the detectors, which corresponds to 21 s (Table 1) for PLATO N-CAM. In the context of PLATO, NSR scales over multiple independent samples and measurements,

Table 7. Description of the parameters of Eq. (11).

Description	Symbol
Photon noise from the target star	$\sigma^2_{F_T}$
Photon noise from a contaminant star	$\sigma^2_{F_C}$
Background noise from the zodiacal light	σ^2_B
Overall detector noise (including readout, smearing, and dark current)	σ^2_D
Quantization noise	σ^2_Q
Average flux from the target star	F_T
Average flux from a contaminant star	F_C
Mask weight in the interval [0, 1]	w
imagette pixel index = $\{1, 2, 3, \ldots, 36\}$	n
Contaminant star index = $\{1, 2, 3, \ldots, N_C\}$	k
Number of contaminant stars within 10 pixel radius around the target	N_C

$$\mathrm{NSR} = \frac{10^6}{12\sqrt{t_d\, N_T}}\,\mathrm{NSR}_*, \tag{12}$$

where t_d is the observation duration in hours and N_T is the number of telescopes observing the star. The constant in the denominator of the above expression stands for the square root of the number of samples in one hour, i.e. $\sqrt{3600\,\mathrm{s}/25\,\mathrm{s}} = 12$, based on the 25 s cadence (Table 1) of the PLATO N-CAMs. For a signal with duration of one hour we use (expressed in units of ppm $\mathrm{h}^{1/2}$)

$$\mathrm{NSR}_{1\,\mathrm{h}} = \frac{10^6}{12\sqrt{N_T}}\,\mathrm{NSR}_*. \tag{13}$$

We note that flux noise induced by satellite jitter is not included in Eq. (12) at this stage. To do so would be a fairly complicated task because jitter contribution depends on the final shape of the aperture (see Fialho et al. 2007). Later in this paper we explain how to include jitter noise in the NSR expressions, subsequent to the determination of the apertures.

4.2. Stellar pollution ratio

We present herein the SPR. This factor permits us to quantify the average fractional contaminant flux from background stars captured by an aperture. We let $F_{C,k}$ be the photometric flux contribution from a single contaminant star k and F_{tot} the total flux. We have

$$F_{C,k} = \sum_{n=1}^{36} F_{C_{n,k}}\, w_n, \tag{14}$$

$$F_{tot} = \sum_{n=1}^{36}\left(F_{T_n} + B_n + \sum_{k=1}^{N_C} F_{C_{n,k}}\right) w_n, \tag{15}$$

where B_n is the average background flux at pixel n from the zodiacal light. We denote SPR_k as the fractional flux from the contaminant star k with respect to the total photometric flux (target plus contaminants and zodiacal light), i.e.

$$\mathrm{SPR}_k = \frac{F_{C,k}}{F_{tot}}. \tag{16}$$

Accordingly, the fractional flux from all contaminant stars is

$$\mathrm{SPR}_{tot} = \sum_{k=1}^{N_C} \mathrm{SPR}_k. \tag{17}$$

We note that SPR_{tot} is complementary to the crowding metric r defined in Batalha et al. (2010), i.e. $\mathrm{SPR}_{tot} = 1 - r$.

4.3. Detectability of planet transits

When a planet eclipses its host star, it produces a maximum transit depth δ_p which is, at first order approximation, equal to the square of the ratio between the planet radius and the star radius

$$\delta_p = \left(R_p/R_\star\right)^2. \tag{18}$$

In practice, δ_p is always diluted by the contaminant flux from surrounding stars and background light, such that the observed transit depth δ_{obs} is a fraction of the original transit depth δ_p

$$\delta_{obs} = (1 - \mathrm{SPR}_{tot})\,\delta_p. \tag{19}$$

Traditionally, a planet detection is not considered scientifically exploitable unless it has been observed at least three times. Furthermore, observed transits must reach a certain level of statistical significance, η, of the total noise, σ. In this paper, we adopted the threshold[16] of 7.1σ ($\eta_{min} = 7.1$) as a minimum condition for characterizing a TCE with three transits. It yields

$$\delta_{obs} \geq \eta_{min}\,\sigma = 7.1\sigma. \tag{20}$$

The total noise σ scales with the signal (transit) duration t_d and with the number of transit events n_{tr}, resulting

$$\sigma = \mathrm{NSR}_{1\,\mathrm{h}}/\sqrt{t_d\, n_{tr}}. \tag{21}$$

By combining the above expressions we can determine the range of detectable planet radius (cf. Batalha et al. 2010)

$$R_p \geq R_\star\sqrt{\frac{\eta}{(1 - \mathrm{SPR}_{tot})}\,\frac{\mathrm{NSR}_{1h}}{\sqrt{t_d\, n_{tr}}}}. \tag{22}$$

Earth-like planets located at about 1au from Sun-like stars have $\delta_p \sim 84$ ppm and $t_d \sim 13$ h. Consequently, it is required that $\mathrm{NSR}_{1h} \lesssim 74\,\mathrm{ppm\,h}^{1/2}$ for that type of planet to be detected at $\eta = \eta_{min} = 7.1$, $n_{tr} = 3$ and $\mathrm{SPR}_{tot} = 0$. From Eq. (22), we can obtain the statistical significance η at which a planet can be detected

$$\eta = \delta_p\sqrt{t_d\, n_{tr}}\,(1 - \mathrm{SPR}_{tot})\,/\mathrm{NSR}_{1h}. \tag{23}$$

Therefore, an aperture model providing the highest number of targets stars with $\eta \geq \eta_{min}$ (i.e. highest N^{good}_{TCE}), for $n_{tr} \geq 3$, is that being more likely in a statistical sense to detect true planet transits.

4.4. Sensitivity to background false transits

In this section, we derive a metric to evaluate the sensitivity of an aperture in detecting false planet transits originating from astrophysical eclipses of contaminant stars. Such events may occur, in particular, when the contaminant star in question is part of an EB system and is sufficiently bright and sufficiently close to a target star. False planet transits caused by grazing EBs are thus not addressed herein.

When a given contaminant star k is eventually eclipsed, we observe in the raw photometry a within aperture fractional flux

[16] This criterion was established to ensure that no more than one false positive due to random statistical fluctuations occurs over the course of the *Kepler* mission (Jenkins et al. 2010b).

decrease $\Delta F^{\rm raw}_{\rm C,k}$ and a corresponding within aperture fractional magnitude increase $\Delta m^{\rm raw}_{\rm C,k}$, such that

$$\Delta m^{\rm raw}_{\rm C,k} = -2.5\log_{10}\left(\frac{F^{\rm raw}_{\rm C,k} - \Delta F^{\rm raw}_{\rm C,k}}{F^{\rm raw}_{\rm C,k}}\right). \quad (24)$$

By denoting $\Delta m^{\rm raw}_{\rm C,k}$ as the background transit depth $\delta_{\rm back,k}$ in mag units and $\Delta F^{\rm raw}_{\rm C,k}/F_{\rm tot}$ as the resulting observed transit depth $\delta_{\rm obs,k}$ in the raw light curve, relative to the contaminant star k, we obtain

$$\delta_{\rm obs,k} = {\rm SPR}^{\rm raw}_{\rm k}\left(1 - 10^{-0.4\delta_{\rm back,k}}\right), \quad (25)$$

with

$${\rm SPR}^{\rm raw}_{\rm k} = \frac{F^{\rm raw}_{\rm C,k}}{F_{\rm tot}}. \quad (26)$$

This expression shows that the background transit depth $\delta_{\rm back,k}$ affects the light curve as an observed transit depth $\delta_{\rm obs,k}$, which is proportional to ${\rm SPR}^{\rm raw}_{\rm k}$, i.e. the SPR of the contaminant star k in the raw photometry. Because $\delta_{\rm obs,k}$ is the result of a false planet transit, we want it to be sufficiently small to prevent it triggering a TCE, i.e.

$$\delta_{\rm obs,k} < \eta_{\rm min}\,\sigma. \quad (27)$$

Although the above statement holds if, and only if, the ${\rm SPR}_{\rm k}$ is below a certain level for given $\delta_{\rm back,k}$, η, $t_{\rm d}$, and $n_{\rm tr}$. We denote such a threshold as the critical SPR (${\rm SPR}^{\rm crit}_{\rm k}$) of the contaminant star k. It can be determined with

$${\rm SPR}^{\rm crit}_{\rm k} = \frac{\eta}{\left(1 - 10^{-0.4\delta_{\rm back,k}}\right)}\frac{{\rm NSR}_{\rm 1h}}{\sqrt{t_{\rm d}\,n_{\rm tr}}}. \quad (28)$$

Therefore, an aperture model providing the lowest number of contaminant stars for which ${\rm SPR}^{\rm raw}_{\rm k} \geq {\rm SPR}^{\rm crit}_{\rm k}$ (i.e. lowest $N^{\rm bad}_{\rm TCE}$), for $\eta \geq \eta_{\rm min} = 7.1$ and $n_{\rm tr} \geq 3$, is that more likely in a statistical sense to naturally reject false planet transits caused by background eclipsing objects.

4.5. Background flux correction

Background correction refers to subtracting, from the raw photometry, flux contributions from contaminant sources and scattered stray light (e.g. zodiacal and Galactic lights). The spatial distribution of background light is commonly describe using polynomial models, whose coefficients are determined based on flux measurements taken at strategically selected pixels (see e.g. Drummond et al. 2008; Twicken et al. 2010). For PLATO, the strategy for background correction is not yet characterized at the present date, thus no accurate information on this subject is available for inclusion in our study. Notwithstanding, we investigate in this section what would be the impact of an ideally perfect background correction on the science metrics $N^{\rm good}_{\rm TCE}$ and $N^{\rm bad}_{\rm TCE}$. We assume therefore a hypothetical scenario in which $B_{\rm n} = F_{\rm C,k} = {\rm SPR}_{\rm k} = {\rm SPR}_{\rm tot} = 0$.

In this case, the observed depth of a legitimate planet transit simply converges to its true depth, i.e. $\delta_{\rm obs} = \delta_{\rm p}$ (the transit dilution is completely cancelled). In parallel, the parameter η (Eq. (23)) increases, meaning that the apertures become more sensitive to detect true planet transits, which ultimately implies an increase in $N^{\rm good}_{\rm TCE}$ as well.

Analysing the impact on $N^{\rm bad}_{\rm TCE}$ is not as straightforward as it is for $N^{\rm good}_{\rm TCE}$. First, we denote hereafter $F^{\rm corr}_{\rm tot}$ as the total photometric flux resulted after the background correction, which only contains signal from the target

$$F^{\rm corr}_{\rm tot} = \sum_{n=1}^{36} F_{\rm T_n}\,w_{\rm n}. \quad (29)$$

Next, we denote $\Delta F^{\rm raw}_{\rm C,k}/F^{\rm corr}_{\rm tot}$ as the resulting observed transit depth $\delta^{\rm corr}_{\rm obs,k}$, after background correction, caused by an eclipse of the contaminant star k. This leads us, using Eq. (24), to an expression for $\delta^{\rm corr}_{\rm obs,k}$ which is similar to that of Eq. (25), except that the term $(F^{\rm raw}_{\rm C,k}/F^{\rm corr}_{\rm tot})$ appears in place of ${\rm SPR}^{\rm raw}_{\rm k}$, resulting in

$$\delta^{\rm corr}_{\rm obs,k} = \left(F^{\rm raw}_{\rm C,k}/F^{\rm corr}_{\rm tot}\right)\left(1 - 10^{-0.4\delta_{\rm back,k}}\right). \quad (30)$$

The above identity shows that removing the background flux from the photometry does not suppress the false transit caused by a background EB. Indeed, although the average flux from the eclipsing contaminant star goes to zero ($F_{\rm C,k} = 0$) in the corrected photometry, the transit depth $\delta^{\rm corr}_{\rm obs,k}$ depends on the intrinsic (raw) contaminant flux $F^{\rm raw}_{\rm C,k}$ that is present in the scene, which is thus independent of any further processing applied in the photometry. Besides, this result is consistent with the fact that the background correction only removes the nominal (out-of-transit) average flux of the contaminant source from the photometry, therefore becoming no longer effective if such signal changes after the correction (e.g. owing to an eclipse, i.e. when $\delta_{\rm back,k} \neq 0$).

For convenience, we define herein the apparent SPR (${\rm SPR}^{\rm app}_{\rm k}$), which is manifested during the eclipse of a contaminant star k in a light curve with flux fully corrected for the background

$${\rm SPR}^{\rm app}_{\rm k} = \frac{F^{\rm raw}_{\rm C,k}}{F^{\rm corr}_{\rm tot}}. \quad (31)$$

This yields

$$\delta^{\rm corr}_{\rm obs,k} = {\rm SPR}^{\rm app}_{\rm k}\left(1 - 10^{-0.4\delta_{\rm back,k}}\right). \quad (32)$$

Comparing Eqs. (25) and (32), we note that $\delta^{\rm corr}_{\rm obs,k}$ is greater than $\delta_{\rm obs,k}$, since ${\rm SPR}^{\rm app}_{\rm k} > {\rm SPR}^{\rm raw}_{\rm k}$. This means that the apertures become more sensitive to detect false planet transits from background eclipsing objects when the corresponding photometry is corrected for the average background flux. This happens because the background correction reduces the dilution of such transits. From all the above considerations, it is possible to state therefore that the background correction is expected to increase both $N^{\rm good}_{\rm TCE}$ and $N^{\rm bad}_{\rm TCE}$ metrics.

4.6. Aperture models

From a purely scientific point of view on planet detection, an ideal aperture is that which is fully sensitive to all true, and fully insensitive to all false, planet transits. However, apertures cannot perfectly disentangle the flux of targets from that of their contaminant sources, so the ideal mask is physically impossible to achieve. Indeed, Eqs. (23) and (28) show us that maximizing the yield of true planet transits and minimizing the occurrences of false planet transits are conflicting objectives: the former requires minimizing NSR and the latter maximizing it. Therefore, the concept of optimal aperture, in the context of this work, is defined as offering the best compromise regarding these two facets, even if the priority is of course to maximize the probability of finding true planet transits. With that in mind, we present

in this section three mask models, each having a different shape and thus supplying distinct performance in terms of NSR and SPR. This gives us elements to check whether a solution giving overall best NSR also has satisfactory performance in terms of SPR and vice versa.

4.6.1. Gradient mask

As NSR is the main performance parameter to be evaluated, a logical mask model to experiment with is that having weights $w_{\rm n}$ providing the best NSR_* for each target. Since the masks have by definition the same dimension of the imagettes, i.e. modest 6×6 pixels, it would be suitable to compute the collection of pixels providing minimum NSR by exhaustive search, i.e. by simple trials of several $w_{\rm n}$ combinations, keeping that with lowest NSR_*. Naturally, that kind of approach is far from efficient, especially considering that this procedure must be executed for tens of thousands of target stars. To avoid this inconvenience, we developed a direct method for calculating $w_{\rm n}$ giving the best NSR. To determine such a mask, we rely on the fact that NSR_*, at its minimum, should have a gradient identically equal to zero ($\nabla \mathrm{NSR}_* = 0$) with respect to the weights. From this, we obtain 36 non-linear equations of the form

$$w_{\rm n}\,\sigma_{\rm n}^2 \sum_{i=1}^{36} w_i\,F_{T_i} = F_{\rm T_n} \sum_{i=1}^{36} w_i^2\,\sigma_i^2, \tag{33}$$

where i is the imagette pixel index $= \{1, 2, 3, \ldots, 36\}$. One simple solution beyond the trivial with $w_{\rm n}$ satisfying the above equality can be calculated directly with

$$w_{\rm n} = \frac{F_{\rm T_n}}{\sigma_{\rm n}^2}. \tag{34}$$

Conventionally, all $w_{\rm n}$ are then normalized by $\max_{[w_{\rm n}]}$ to satisfy $0 \leq w_{\rm n} \leq 1$, so that each weight $w_{\rm n}$ directly represents the fraction of the imagette flux being caught by the aperture at the corresponding pixel n. For illustration, Fig. 14a shows the resulting gradient mask for the input image example of Fig. 11.

In order to simplify our terminology, the masks $w_{\rm n}$ obtained from Eq. (34) are hereafter referred to as gradient masks based on the fact that they are determined from the mathematical gradient of NSR_* expression. Each time they are mentioned however we should keep in mind that they correspond to the masks providing the global minimum NSR from all the possible combinations of mask weights $w_{\rm n}$ in Eq. (11).

4.6.2. Gaussian mask

Having examined the shape of gradient masks applied to several stars, we noticed that they look very similar to a bell shaped curve. Therefore, we decided to test Gaussian-like masks to verify whether they could provide near-best NSRs when compared to gradient masks. Depending on the performance difference, the advantage of having an analytical mask that requires fewer parameters to be computed could justify its choice over the gradient mask. On these terms, we calculate the weights $w_{\rm n}$ of a Gaussian mask using the conventional symmetric Gaussian function expression

$$w_{x,y} = \exp\left(-\frac{(x - x_\star)^2 + (y - y_\star)^2}{2\sigma_w^2}\right), \tag{35}$$

where (x, y) are Cartesian coordinates of the imagette pixels with shape 6×6; $(x_\star, y_\star)$ are the coordinates of the target barycentre within the imagette; $\sigma_{\rm w}$ is the mask width in pixels on both x and y dimensions; $w_{x,y}$ is the mask weight in the interval [0, 1] at (x, y).

As the imagette dimension is fixed and the target position within it is well known thanks to the input catalogue, choosing a Gaussian mask for a given target reduces to finding a proper width. For that, we simply iterate over different values of $\sigma_{\rm w}$ and keep that giving the lowest NSR_*, as shown in Fig. 13. For illustration, Fig. 14b shows the resulting best NSR Gaussian mask for the input image example of Fig. 11.

4.6.3. Binary mask

Binary masks are non-weighted apertures, meaning that the photometry is extracted by fully integrating pixel fluxes within the mask domain and discarding those which are outside it. This type of aperture was extensively employed to produce light curves of CoRoT and *Kepler* targets, so it is well known for delivering satisfactory performance. In the context of PLATO, we applied the following routine to compute a binary mask for each target imagette.

1. Arrange all pixels n from the target imagette in increasing order of $\mathrm{NSR}_{\rm n}$

$$\mathrm{NSR}_{\rm n} = \frac{\sqrt{\sigma_{F_{\rm T_n}}^2 + \sum_{k=1}^{N_{\rm C}} \sigma_{F_{\rm C_{n,k}}}^2 + \sigma_{B_{\rm n}}^2 + \sigma_{D_{\rm n}}^2 + \sigma_{Q_{\rm n}}^2}}{F_{\rm T_n}}. \tag{36}$$

2. Compute the aggregate noise-to-signal $\mathrm{NSR}_{\rm agg}(m)$, as a function of the increasing number of pixels $m = \{1, 2, 3, \ldots, 36\}$, stacking them to conform to the arrangement in the previous step and starting with the pixel owning the smallest $\mathrm{NSR}_{\rm n}$

$$\mathrm{NSR}_{\rm agg}(m) = \frac{\sqrt{\sum_{n=1}^{m} \left(\sigma_{F_{\rm T_n}}^2 + \sum_{k=1}^{N_{\rm C}} \sigma_{F_{\rm C_{n,k}}}^2 + \sigma_{B_{\rm n}}^2 + \sigma_{D_{\rm n}}^2 + \sigma_{Q_{\rm n}}^2\right)}}{\sum_{n=1}^{m} F_{\rm T_n}}. \tag{37}$$

3. Define as the aperture the collection of pixels m providing minimum $\mathrm{NSR}_{\rm agg}(m)$.

As the binary mask gets larger following the above routine, the NSR typically evolves as illustrated in Fig. 13. Accordingly, the resulting best NSR binary mask for the input image example of Fig. 11 is shown in Fig. 14c.

5. Performance assessment

We present in this section the photometric performance of the three aperture models defined in Sect. 4.6. The results are presented in terms of NSR, SPR, number $N_{\rm TCE}^{\rm good}$ of target stars with sufficiently low NSR permitting the detection of planets orbiting them, and number $N_{\rm TCE}^{\rm bad}$ of contaminant stars with sufficiently high SPR to produce, should they be eclipsed, false positives. The results were obtained by applying each aperture model to all 50 000 input imagettes from Sect. 3.4.

5.1. Noise-to-signal ratio

As already pointed out in Sect. 4.1, the per cadence NSR_* from Eq. (12) does not include photometric flux noise induced by spacecraft jitter because of its dependency on aperture weights. Once the apertures are computed however, we can include jitter noise in the photometry using the shifted imagettes described

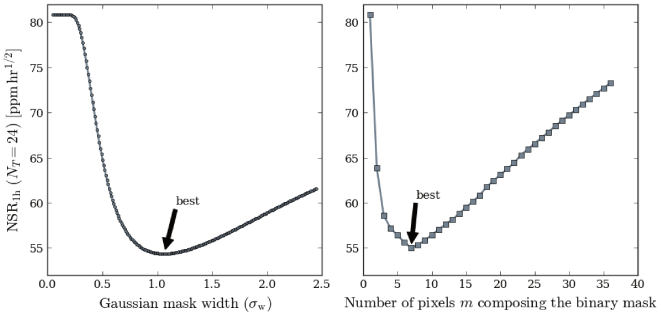


Fig. 13. Example of NSR evolution curve as a function of the increasing aperture size for a target star with $P = 11$. *Left*: Gaussian mask. *Right*: binary mask.

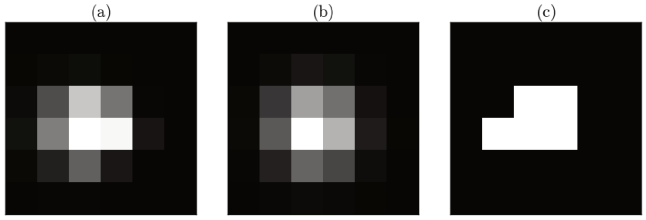


Fig. 14. Aperture shapes computed as described in Sect. 4.6, for the input image example of Fig. 11. *Left*: gradient mask. *Centre*: Gaussian mask. *Right*: binary mask.

in Sect. 3.4. We denote $\mathrm{NSR}_*^{\mathrm{jitter}}$ the per cadence NSR, which includes star motion due to satellite jitter, i.e.

$$\mathrm{NSR}_*^{\mathrm{jitter}} = \mathrm{NSR}_* \sqrt{1 + \left(\frac{\sigma_\mathrm{J}^2}{\sigma_*^2}\right)}, \tag{38}$$

where σ_J is the photometric jitter noise obtained from the shifted imagettes and σ_* corresponds to the numerator of the expression in Eq. (11). The above expression considers stationary random noise for both photometric flux and satellite jitter. Table 8 shows the impact of spacecraft jitter on the photometry under nominal and degraded scenarios of pointing performance. We verified that in nominal conditions the impact of jitter on the photometry is negligible, showing that including jitter in the calculation scheme of the apertures would not only represent a complicated procedure, but also a useless effort in that particular case.

The performance parameter $\mathrm{NSR}_{1\,\mathrm{h}}^{\mathrm{jitter}}$ was computed for our subset of input images assuming a nominal satellite jitter. The results are shown in Fig. 15. Overall, the three aperture models present comparable results for targets brighter than $P \sim 10.5$, with differences of less than 2% on average. The Gaussian mask has consistent suboptimal NSR performance over the entire P5 magnitude range, that is only ~1% higher on average than the gradient mask. The binary mask has better performance on average than the Gaussian mask for targets brighter than $P \sim 9$, but its performance degrades rapidly with increasing magnitude. For the faintest P5 targets, the binary mask presents NSR values about 6% higher on average and ~8% higher in the worst scenarios with respect to the gradient mask. Therefore, looking exclusively in terms of NSR, weighted masks are clearly the best choice.

5.2. Stellar pollution ratio

We present in Figs. 16 and 17 the results of SPR_k and $\mathrm{SPR}_\mathrm{tot}$, respectively. The total SPR ($\mathrm{SPR}_\mathrm{tot}$) was computed for all 50 000 sources of our working subset of targets, while the per

Table 8. Maximum noise-to-signal degradation at 95% confidence level as a function of satellite jitter amplitude, computed from a sample of 10 000 targets.

Aperture model	Nominal	3 × Nominal	5 × Nominal	7 × Nominal
Gradient	0.31%	2.9%	8.1%	16.2%
Gaussian	0.41%	3.6%	10.0%	19.43%
Binary	0.50%	4.7%	12.3%	23.2%

Notes. Four scenarios are considered: nominal (Fig. 12), three times (3×) nominal, five times (5×) nominal, and seven times (7×) nominal jitter.

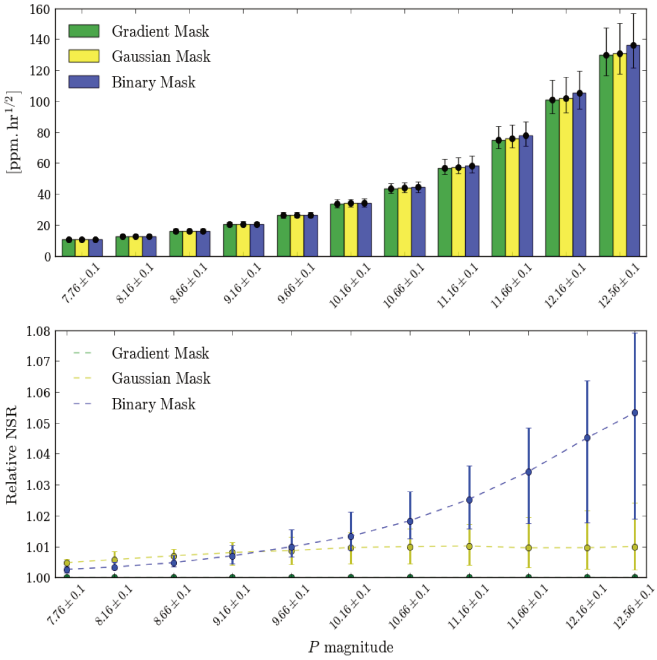


Fig. 15. *Top*: median values (black dots) of $\mathrm{NSR}_{1\,\mathrm{h}}^{\mathrm{jitter}}$ ($N_\mathrm{T} = 24$) as a function of target P magnitude and the applied mask model. *Bottom*: relative $\mathrm{NSR}_{1\,\mathrm{h}}^{\mathrm{jitter}}$, where the unit stands for the best NSR. In both plots, interval bars represent dispersions at 90% confidence level.

contaminant SPR (SPR_k) was computed for all ~3.25 million stars located within a 10 pixel radius from those targets. Both plots show that the binary mask collects significantly less contaminant flux overall, and more particularly when the contaminant sources are located at more than 2 pixels distant from the targets. To give a rough idea, for about 80% of the contaminant sources SPR_k is at least three times greater for the weighted masks. This result however is not surprising because gradient and Gaussian masks are typically larger to best fit the shape of the PSF. This is the reason why they typically give lower NSR, as shown in the previous section.

5.3. Detectability of planet transits

With both NSR and SPR determined, we are now capable of estimating the number $N_\mathrm{TCE}^\mathrm{good}$ of target stars with sufficiently low NSR permitting the detection of eventual planets orbiting them. Tables 9 and 10 show the values for $N_\mathrm{TCE}^\mathrm{good}$ for the case of an Earth-like planet orbiting a Sun-like star, respectively, for the scenarios of $\mathrm{SPR}_\mathrm{tot}$ as given by Fig. 17 (no background correction) and $\mathrm{SPR}_\mathrm{tot} = 0$ (perfect background correction).

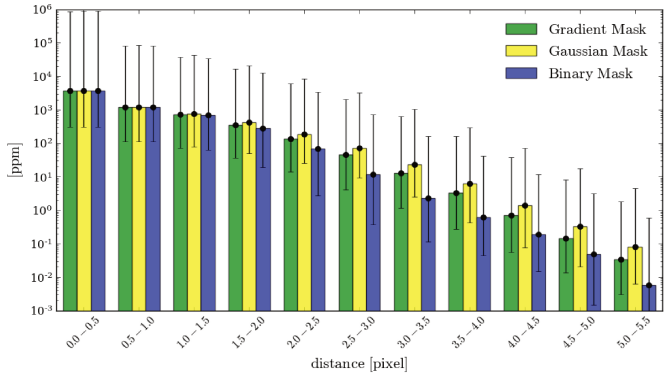


Fig. 16. Median values (black dots) of SPR_k (Eq. (16)) as a function of the distance in pixels between the contaminants sources and their respective targets, and the applied mask model. Interval bars represent dispersions at 90% confidence level.

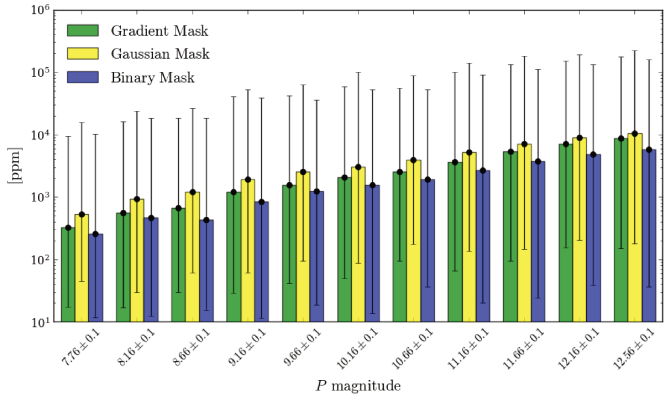


Fig. 17. Median values (black dots) of SPR_{tot} (Eq. (17)) as a function of target P magnitude and the applied mask model. Interval bars represent dispersions at 90% confidence level.

The results show that the advantage of weighted masks regarding NSR performance, which is up to ~7.5% better with respect to the binary mask for the faintest and most numerous targets (see Fig. 15), does not translate into a proportionally better sensitivity in detecting true planet transits. Indeed, the mask with lowest NSR, called the gradient mask, provides only ~0.8% more chance of detecting Earth-like planets orbiting Sun-like stars at 1au. The difference between Gaussian and binary masks is even smaller, i.e. ~0.4%. All three masks are equally capable of detecting Jupiter-like planets, no matter the number of telescopes observing the host star. To understand this, we need to compare Figs. 15 and 18. Taking the case of detecting Earth-like planets at about 1 au from Sun-like stars, our analyses show that the limiting magnitude[17] for aperture photometry is of the order of $P \sim 11.7$ ($V \sim 12$ at 6000 K) at 7.1σ, $n_{tr} = 3$ and $N_T = 24$. Therefore, for most of the magnitude range ($11 \lesssim P \leq 12.66$) where the binary mask present the most degraded NSR performance with respect to the weighted masks, the latter do not provide any advantage in detecting such planets after all. Thus the small differences in N_{TCE}^{good} between binary and weighted masks, for the considered scenario, are consistent.

Correcting for the background results in an almost negligible impact (≲0.6% increase) in the overall sensitivity of the apertures in detecting true planet transits. Also, it has no significant impact in the comparative basis analysis between the different aperture models. We stress however that inefficient background correction may significantly limit the accuracy with which planet transit depths can be determined.

Hence, from a planet transit finding perspective, designating an optimal solution for extracting photometry from the P5 stellar sample now becomes substantially less obvious. To this extent, looking at how each aperture performs in terms of false planet transit rejection may give us a hint about which is effectively the most appropriate choice.

[17] We note that this threshold is likely to be diminished by the presence of stellar activity in the noise (see Gilliland et al. 2011).

Table 9. Number N_{TCE}^{good} of target stars for which $\eta \geq \eta_{min}$, as a function of the number N_T of telescopes observing them and the applied aperture model.

N_T	Gradient mask	Gaussian mask	Binary mask
24	19 063 (38.1%)	18 674 (37.3%)	18 201 (36.4%)
18	15 105 (30.2%)	14 753 (29.5%)	14 469 (28.9%)
12	10 629 (21.3%)	10 368 (20.7%)	10 202 (20.4%)
6	5528 (11.1%)	5395 (10.8%)	5357 (10.7%)
weighted	10 067 (20.1%)	9833 (19.7%)	9667 (19.3%)

Notes. We present above the case of an Earth-like planet with $\delta_p = 84$ ppm, $t_d = 13$ h, $n_{tr} = 3$ and SPR_{tot} given by the simulated values presented in Fig. 17 (i.e. assuming no background correction). The values in this table were determined from our dataset of 50 000 target stars. The weighted values correspond to the effective N_{TCE}^{good}, obtained by assuming uniform star distribution and a fractional field of view as given in Table 1.

Table 10. Same as Table 9, but for $SPR_{tot} = 0$ (i.e. assuming a perfect background correction).

N_T	Gradient mask	Gaussian mask	Binary mask
24	19 608 (39.2%)	19 319 (38.6%)	18 637 (37.3%)
18	15 510 (31.0%)	15 264 (30.5%)	14 806 (29.6%)
12	10 909 (21.8%)	10 701 (21.4%)	10 441 (20.9%)
6	5625 (11.2%)	5527 (11.1%)	5456 (10.9%)
weighted	10 318 (20.6%)	10 141 (20.3%)	9884 (19.8%)

Notes. A scatter plot of η, as a function of target P magnitude, is illustrated in Fig. 18 for $N_T = 24$.

5.4. Sensitivity to background false transits

We now compare the parameters SPR_k^{raw} (Eq. (26)) and SPR_k^{app} (Eq. (31)) with SPR_k^{crit} (Eq. (28)), to determine the number N_{TCE}^{bad} of contaminant stars with sufficiently high average flux to generate false positives. Two scenarios are considered herein: N_{TCE}^{bad} representing the number of contaminant sources for which $SPR_k^{raw} \geq SPR_k^{crit}$, which supposes no background correction in the photometry; and N_{TCE}^{bad} representing the number of contaminant sources for which $SPR_k^{app} \geq SPR_k^{crit}$, which supposes a perfect background correction in the photometry. In both cases, we define SPR_k^{crit} with $\delta_{back,k} = 8.5\%$ (~0.1 mag), $t_d = 4$ h, $\eta = 7.1$, and $n_{tr} = 3$. The chosen value for $\delta_{back,k}$ corresponds to the median depth of the sources in the *Kepler* Eclipsing Binary Catalogue (Third Revision)[18], considering both primary (pdepth) and secondary (sdepth) depths together. The

[18] http://keplerebs.villanova.edu/

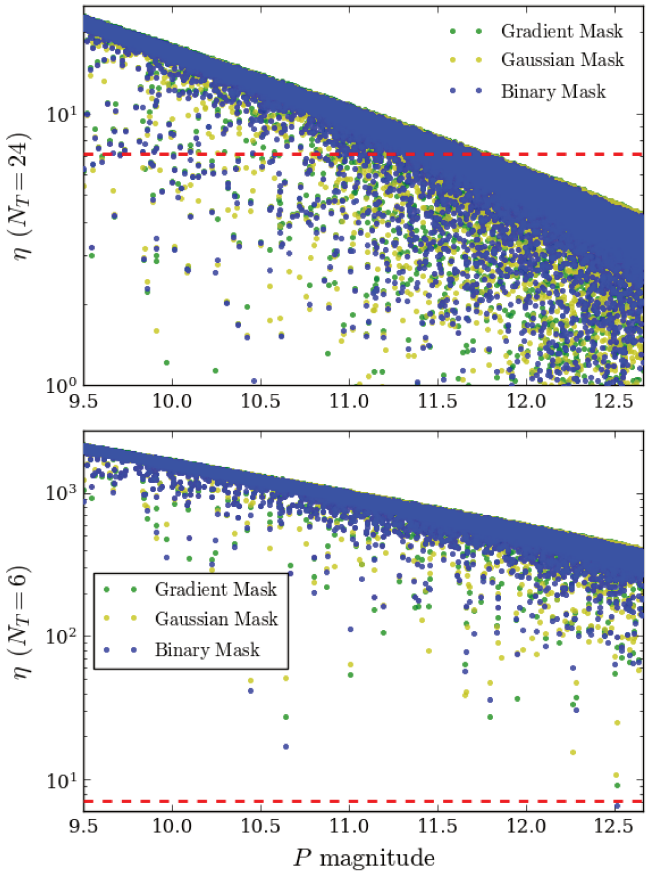


Fig. 18. Scatter plot of the statistical significance η (Eq. (23)) computed for 50 000 target stars, as a function of their respective P magnitude and the applied aperture model. The red dashed line represents the threshold $\eta_{\rm min}$ = 7.1. Values of $N_{\rm TCE}^{\rm good}$ are provided in Tables 9 and 10. *Top*: statistics for an Earth-like planet with $\delta_{\rm p}$ = 84 ppm, $t_{\rm d}$ = 13 h, $n_{\rm tr}$ = 3, SPR$_{\rm tot}$ = 0, and $N_{\rm T}$ = 24. *Bottom*: statistics for an Jupiter-like planet with $\delta_{\rm p}$ = 0.1, $t_{\rm d}$ = 29.6 h, $n_{\rm tr}$ = 3, SPR$_{\rm tot}$ = 0, and $N_{\rm T}$ = 6.

chosen value for $t_{\rm d}$ corresponds to the median transit duration of the offset false positive sources listed in the Certified False Positive Table at NASA Exoplanet Archive[19]. The transit duration values themselves were retrieved form the Threshold Crossing Events Table, by crossmatching the ID columns (KepID) from both tables.

Looking at the obtained results for $N_{\rm TCE}^{\rm bad}$, which are presented in Tables 11 and 12, the important thing to notice at first glance is the fact that all tested aperture models have, fortunately, an intrinsically very low (less than 5%) overall sensitivity to detect mimicked planet transits caused by background eclipsing objects. In other words, these models are all insensitive to most of the potential false planet transits that may be produced by the contaminant sources in regions IV and VIII of Fig. 19. This is surely mostly because of the high enclosure energy of PLATO PSFs, but the optimization scheme applied to each aperture model, privileging low NSR, is also key in this context. Nevertheless, the results also clearly show that compared to the binary mask employing weighted masks substantially increases the predicted occurrence of events mimicking planet transits. The Gaussian mask is expected to deliver up to ~40% higher $N_{\rm TCE}^{\rm bad}$ than the binary mask, which is notably a huge discrepancy. The differences between gradient and binary masks are smaller, but still very significant: $N_{\rm TCE}^{\rm bad}$ is up to ~20% higher for the gradient mask. Either correcting for the background or not, these differences rest roughly the same, so background correction has no significant impact in the comparative basis analysis between the different aperture models. In absolute terms though, the results indicate that fully removing the background leads to an overall increase of more than 10% in $N_{\rm TCE}^{\rm bad}$, which is consistent with the analysis presented in Sect. 4.5.

Overall, the obtained results for $N_{\rm TCE}^{\rm bad}$, in comparison to those of $N_{\rm TCE}^{\rm good}$ presented in the previous section, makes the scenario of choosing weighted masks become highly unfavoured even though that kind of mask provides better overall performance in terms of NSR. Still, it would be legitimate to ask whether the obtained values for $N_{\rm TCE}^{\rm bad}$ are indeed significant in an absolute sense, since they represent less than 5% of our full set of contaminant stars composed of ~3.25 million sources. Properly answering this question requires carefully modelling the parameters $\delta_{\rm back,k}$ and $t_{\rm d}$ for the PLATO target fields, which is though beyond the scope of this paper. However, it is possible to obtain a rough idea of the occurrence of EBs ($N_{\rm beb}$) that could potentially result from the weighted values shown in Tables 11 and 12. First, we need to consider that these values refer to about 20% of the minimum number of expected targets for the P5 sample. Second, we may assume that the frequency of EBs ($F_{\rm eb}$) for the PLATO mission might be of the order of 1%[20]. Accordingly, the expected occurrence of EBs at 7.1σ, for the P5 sample could be approximately estimated with $N_{\rm beb} \sim 5 \times N_{\rm TCE}^{\rm bad} \times 1\%$. From the weighted values presented in Tables 11 and 12, that gives $1600 \lesssim N_{\rm beb} \lesssim 2500$ (all three tested aperture models comprised). This allows us to conclude that $N_{\rm TCE}^{\rm bad}$ is thus not negligible. Moreover, considering that the total number of targets in the P5 sample is comparable to the total number of observed targets by the *Kepler* mission, we verified that our approximative estimate on the expected $N_{\rm beb}$, for the P5 sample, is very consistent to the statistics of background false positives of the *Kepler* mission. Indeed, the Certified False Positive Table on the NASA exoplanet archive gives at the present date 1287 offset false positives out of 9564 *Kepler* objects of interest. Such a consistency attests that our study is satisfactorily realistic. We stress however that accurate false positive estimates for the P5 sample cannot be provided by our study alone, in particular because it needs to be consolidated with PLATO's science exoplanet pipeline.

As a complement to the results presented in this section, Fig. 20 shows, for each aperture model, a two-dimensional histogram containing the distribution of contaminant stars having SPR$_{\rm k}^{\rm app} \geq$ SPR$_{\rm k}^{\rm crit}$, as a function of the differential P magnitude and the Euclidean distance between these sources and the corresponding targets. The parameters used to calculate SPR$_{\rm k}^{\rm crit}$ were $\delta_{\rm back,k}$ = 0.8 mag, $t_{\rm d}$ = 4 h, $N_{\rm T}$ = 24, η = 7.1, and $n_{\rm tr}$ = 3. This plot is of particular interest since it illustrates that the contaminant stars having sufficiently high average flux to produce background false positives are typically less than ~10 mag brighter and located at less than ~4 pixels away from the targets. Consequently, from the point of view of the distances, we verified that our approach of considering contaminant sources located at up to 10 pixels distant from the targets was largely enough for the purposes of this work. From the point of view of the differential

[19] https://exoplanetarchive.ipac.caltech.edu

[20] Fressin et al. (2013) give $F_{\rm eb}$ = 0.79% for the *Kepler* mission. They defined it as being the fraction of EBs found by *Kepler*, including detached, semi-detached, and unclassified systems, divided by the number of *Kepler* targets.

Table 11. Number $N_{\rm TCE}^{\rm bad}$ of contaminant stars for which $\rm SPR_k^{raw} \geq SPR_k^{crit}$, which supposes photometry with no background correction, as a function of the number $N_{\rm T}$ of telescopes observing the host star and the aperture model.

$N_{\rm T}$	Binary mask (I + II) = (V + VI)	Gradient mask (I + III)	Gaussian mask (V + VII)
24	40 135	48 005	55 520
18	36 835	43 690	50 785
12	32 830	38 485	44 565
6	26 545	31 050	35 995
weighted	31 591	37 178 (+17.7%)	43 073 (+36.3%)

Notes. The presented values were determined from our dataset of ~3.25 million contaminant stars. The $\rm SPR_k^{crit}$ was computed with $\delta_{\rm back,k} \sim 0.1$ mag, $t_{\rm d} = 4$ h, $\eta = 7.1$, and $n_{\rm tr} = 3$. The roman numerals correspond to the areas indicated in Fig. 19. The percentiles indicate the amount of deviation of the values from weighted masks with respect to those from binary mask. The weighted values in the lower row correspond to the effective $N_{\rm TCE}^{\rm bad}$, obtained by assuming uniform star distribution and a fractional field of view as given in Table 1.

Table 12. Same as Table 11, but now representing the contaminant stars for which $\rm SPR_k^{app} \geq SPR_k^{crit}$, which supposes photometry with perfect background correction.

$N_{\rm T}$	Binary mask (I + II) = (V + VI)	Gradient mask (I + III)	Gaussian mask (V + VII)
24	45 180	54 720	63 555
18	41 575	50 185	58 540
12	37 055	44 380	51 885
6	30 185	35 820	41 570
weighted	35 731	42 774 (+19.7%)	49 814 (+39.4%)

magnitude, three important aspects need to be considered when interpreting the results.

First, we note that stars in our input catalogue are limited in magnitude to $P \sim 21.1$. This means that for the faintest (and most numerous) P5 targets, for which P magnitude is as high as 12.66, the maximum differential magnitude from their contaminants is therefore as small as $21.1 - 12.66 = 8.44$ mag, i.e. smaller than the limit of ~10 mag suggested by the histograms. In contrast, P5 has targets as bright as 7.66 mag, so that the differential magnitude may be as high as $21.1 - 7.66 = 13.44$ mag. Hence, well above that limit.

Second, we notice in Fig. 20 some supposedly missing stars at distances near zero, in particular at differential magnitudes above 5 mag. We understand such an anomaly to be related to what we have already pointed out in Sect. 3.3 concerning bad estimates of the fluxes of stars fainter than $G \sim 17$ in the DR2 catalogue. This issue is reported in Evans et al. (2018) and assumed to be caused by factors such as poor background estimation, observation taken in the proximity of bright sources, binarity, and crowding. In these conditions, the capability to isolate stars is therefore compromised. Taking into account that the most problematic cases were removed from the DR2 release according to the authors, the lack of stars in the above mentioned areas of Fig. 20 is justified. Yet scenarios of differential magnitude higher than 10 mag, at the same time that $\rm SPR_k^{app} \geq SPR_k^{crit}$, should mostly occur at distances shorter than ~0.5 pixel, where the occurrence of contaminant stars is substantially smaller than that at longer distances (see Fig. 10).

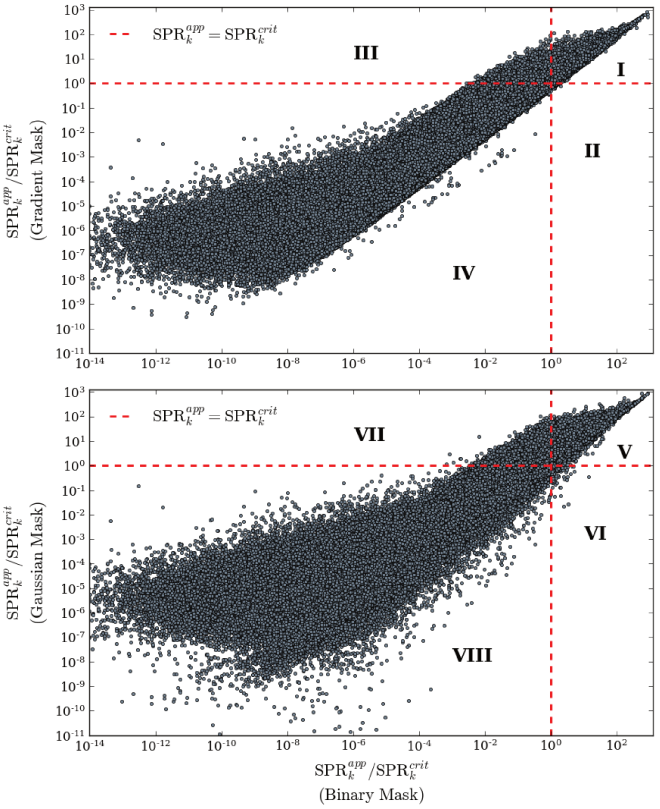


Fig. 19. Scatter plot of $\rm SPR_k^{app}$ normalized by $\rm SPR_k^{crit}$, computed for ~3.25 million contaminant stars. This illustration represents the particular case where $\rm SPR_k^{crit}$ is computed with $\delta_{\rm back,k} \sim 0.1$ mag; $N_{\rm T} = 12$; $n_{\rm tr} = 3$; $t_{\rm d} = 4$ h. Values of $N_{\rm TCE}^{\rm bad}$ are provided in Tables 11 and 12. *Top*: comparison between the values given by the gradient mask (vertical axis) and by the binary mask (horizontal axis). Region I: both masks exceed $\rm SPR_k^{crit}$. Region II: only the binary mask exceeds $\rm SPR_k^{crit}$. Region III: only the gradient mask exceeds $\rm SPR_k^{crit}$. Region IV: no mask exceeds $\rm SPR_k^{crit}$. *Bottom*: comparison between the values given by the Gaussian mask (vertical axis) and the binary mask (horizontal axis). Regions V to VIII are analogous to I, II, III, and IV, respectively.

Third, the parameters used to build the histograms of Fig. 20 correspond in practice to a near worst case scenario in terms of the expected occurrences of false transits caused by background eclipsing objects. Indeed, it considers photometry perfectly corrected for the background; contaminants stars being observed by 24 cameras (maximum sensitivity to transit signatures); and contaminant stars generating background transit depths of 0.8 mag, which is significantly high. This means that the maximum differential magnitude is typically much smaller than 10 mag.

Taking into account all the above considerations, we conclude that Fig. 20 gives a sufficiently realistic and unbiased representation of distances and differential magnitudes of contaminant stars that are likely to cause background false planet transits, regardless of the limitation in maximum magnitude of our input catalogue. Furthermore, we note that the missing fraction (~0.01%) of PSF energy in the images of Fig. 4 entails no significant impact in our analysis. This small fractional energy may be non-negligible uniquely in cases in which the differential magnitude between target and contaminant stars is $\lesssim -4$ mag. These are however extremely rare scenarios in our input stellar field, and thus statistically insignificant to our analysis. Indeed, less than 0.5% of the contaminant sources in Fig. 20 have differential magnitude smaller than −2.6 mag.

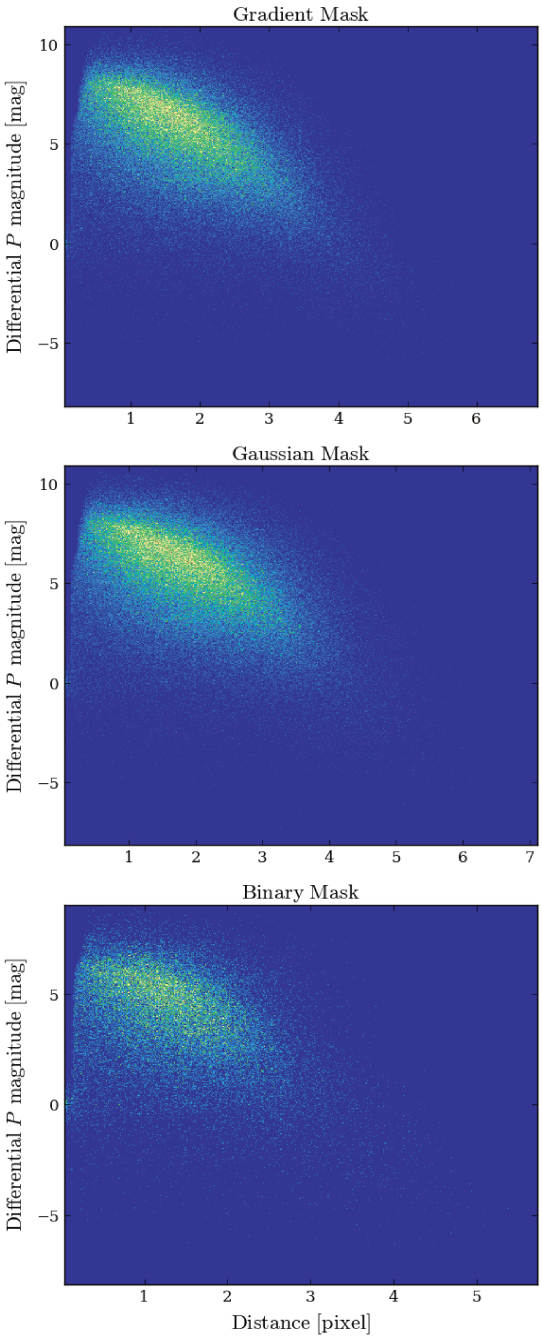


Fig. 20. Two-dimensional histograms of the distribution of contaminant stars with $SPR_k^{app} \geq SPR_k^{crit}$, for gradient (*top*), Gaussian (*centre*), and binary (*bottom*) masks. The vertical axis indicates the differential P magnitude between the contaminants and their respective targets, whereas the horizontal axis indicates the corresponding Euclidean distances. The parameters used to calculate SPR_k^{crit} are $\delta_{back,k} = 0.8$ mag, $N_T = 24$, $n_{tr} = 3$, and $t_d = 4$ h.

Ultimately, we extract the unique set of contaminant stars from all three histograms presented in Fig. 20, and use it to build a histogram of the fractional distribution of contaminant stars having $SPR_k^{app} \geq SPR_k^{crit}$ (i.e. the fractional distribution of N_{TCE}^{bad}) as a function of Galactic latitude. The resulting plot is shown in Fig. 21. It suggests that the occurrences of false positives caused by background eclipsing stars might increase exponentially towards the Galactic plane, which is consistent with the distributions of offset transit signals presented in Bryson et al. (2013). We note that since the distribution of stars within our IF privileges certain latitudes, because of its its circular shape, we avoid propagating such bias to the data of Fig. 21 by considering contaminant sources within a sufficiently narrow Galactic longitude range $l_{LoS} \pm 1.5$ [deg] (see Table 4).

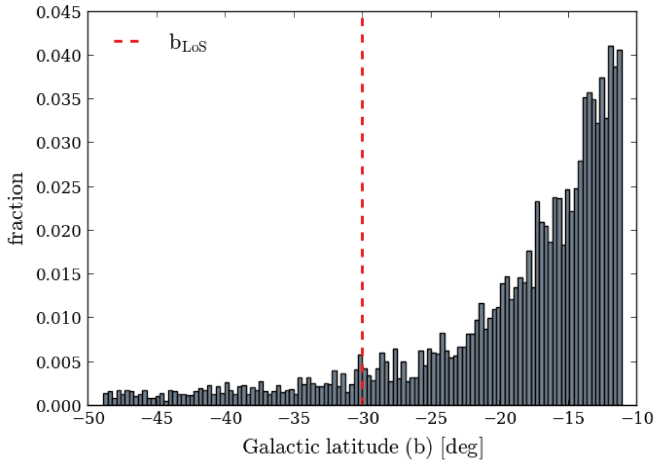


Fig. 21. Fractional distribution of N_{TCE}^{bad}, as a function of Galactic latitude (all three mask models comprised). This histogram (cf. Bryson et al. 2013) was built with contaminant sources that have Galactic longitude within the range $l_{LoS} \pm 1.5$ [deg]. The red vertical line indicates the Galactic latitude b_{LoS} of IF_{LoS} (see IF coordinates in Table 4).

5.5. Implementation constraints

5.5.1. Updating the masks on board

As explained earlier in Sect. 2.2, the pixels of PLATO detectors are relatively broad compared to the size of the PSFs. During observations, this causes aperture photometry to be sensitive to the long-term star position drift occurring on the focal plane. For PLATO, this effect is expected to be caused notably by the orbital differential velocity aberration[21] and the thermo-elastic distortion from the optical bench, and might be as large as 1.3 pixel over three months. Consequently, mask-target assignments performed during each calibration phase become, soon or later, no longer optimal, since the flux distributions of the targets significantly change as these move across the pixels. Therefore, the NSR of the resulting light curves substantially increases.

To compensate for this effect, the proposed solution consists in tracking the targets by updating the placement of their apertures on board, as explained in Samadi et al. (2019). This will involve both ground and flight segments of the mission, as the apertures will first be computed on the ground and then transmitted to the spacecraft. Both the criteria and timescale on which such actions will be performed are yet to be defined.

5.5.2. Uploading the masks on board

The performance results presented in this paper were obtained by assigning apertures for each target individually, following the computation schemes presented in Sect. 4.6. Taking the case of the gradient mask, which provides the lowest values of NSR, having such performance on board requires a unique mask shape per target to be uploaded to the flight software. This demands, in turn, prohibitive telemetry and time resources. In addition,

[21] http://www.stsci.edu/hst/fgs/documents/datahandbook

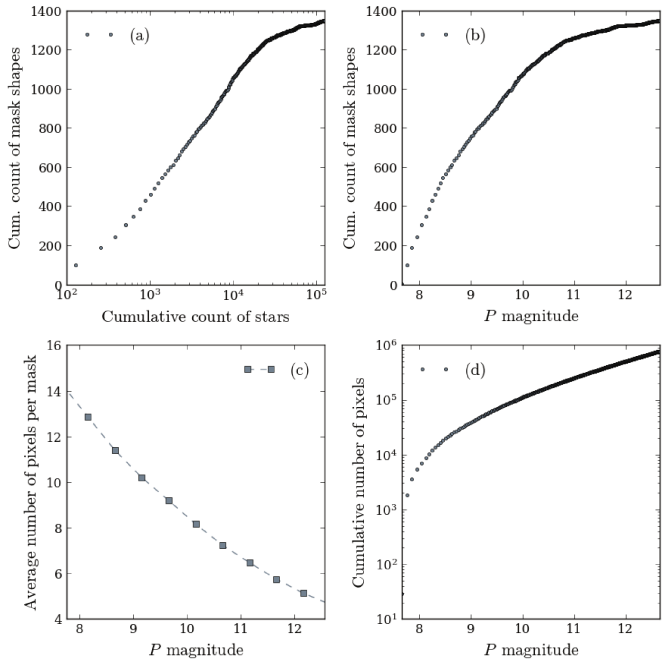


Fig. 22. Statistics on the morphology of binary masks. The above results are based on all ∼127 thousand target stars within IF (see Sect. 3.3). (*a*) Cumulative count of unique binary mask shapes as a function of the cumulative count of target stars. (*b*) Cumulative count of unique binary mask shapes as a function of target P magnitude. (*c*) Average number of pixels composing the binary masks as a function of target P magnitude. (*d*) Cumulative number of pixels composing the binary masks as a function of target P magnitude.

as explained in Sect. 5.5.1, the masks will have to be regularly updated in flight to compensate for long-term star position drift, thus making the employment of gradient masks unfeasible.

For the Gaussian mask the outlook is not much more favourable, as this solution would require a massive set of widths (practically one per target) to guarantee the NSR performance results presented in Fig. 15. Otherwise, we could in principle take advantage of the fact that the Gaussian mask has an analytical form – with small number of parameters – to apply simplification schemes to avoid the need for having one particular mask per target. For instance, a feasible approach would consist of employing polynomial surfaces or fixed widths to cover the multiple combination scenarios in terms of magnitude and intra-pixel location of the targets. Nevertheless, that would inevitably reduce the overall NSR performance, which is the major benefit of using weighted masks.

The binary mask, in turn, provides a virtually unbeatable capacity for compressing combinations of mask shapes without loss of performance. We can visualize this by looking at the data concerning the morphology of binary masks provided in Fig. 22a. These data give the accumulated unique combinations of binary mask shapes computed from the set of binary masks used to extract photometry from all ∼127 thousand target stars in our adopted IF. We verify that the unique set saturates to about only 1350 mask shapes, thereby giving a compression factor of almost 99%. This represents another significant advantage of employing the binary mask, since no weighted mask is actually capable of providing such compression capabilities while keeping the original performance of the full set of masks unchanged.

We note that the statistics on the number of binary mask shapes and pixels, presented in Fig. 22, are valid for non-saturated stars. It implies that only one mask is attributed to each target, and each mask is limited in size by the (6 × 6) shape of an imagette. This is a fundamental assumption for the study presented herein. In the current instrument design, PLATO detectors are expected to exhibit saturation at pixels observing stars brighter than $P \sim 8.16 \pm 0.5$ (i.e. $V \sim 8.5 \pm 0.5$ at 6000 K) after a 21 s exposure (normal cameras). The exact saturation limit depends on the location of the star in the CCD and where its barycentre falls within a pixel. The brightest stars in our study are thus at the very lower bound of this broad saturation threshold. In the context of the PLATO science pipeline, the photometry of saturated stars will be extracted exclusively from the ground on the basis of extended imagettes, that is, nominal imagettes appropriately extended such as to capture the charges spilt along the CCD columns from the saturation.

6. Conclusions and discussions

Light curves will be produced in flight for potentially more than 250 000 PLATO targets (the P5 stellar sample) by employing aperture photometry. To maximize the scientific exploitability of the resulting data, an appropriate aperture model needs to be determined. Aiming to fulfil this objective, we presented in this paper a detailed photometric performance analysis based on three different strategies: a weighted aperture providing global minimum NSR (gradient mask) obtained through a novel direct calculation method; a weighted Gaussian aperture giving sub-optimal NSR; and a narrower binary aperture to reduce the impact of contamination. Each aperture model was applied to 50 000 synthetic imagettes containing a realistic stellar distribution with more than three million sources, extracted from the *Gaia* DR2 catalogue. The stellar population was obtained from one of the expected long-pointing fields for the mission, namely the southern PLATO field.

For a more appropriate estimate of stellar fluxes reaching the instrument's detectors, we established a synthetic PLATO P photometric passband derived from the spectral response of the instrument and calibrated in the VEGAMAG system. This allows us to avoid the inconvenience of having colour dependency when estimating stellar fluxes from V magnitudes. The photometric relationships $V - P$ and $G - P$ are included. In addition, we used a zodiacal light semi-analytical model from the literature to derive an expression for estimating the intensities of scattered background light entering the PLATO cameras.

To determine the optimal aperture model for extracting photometry from the P5 targets, we adopted an innovative criterion that is based on two science metrics: a simulated number of target stars for which a planet orbiting it would be detected, denoted as $N_{\rm TCE}^{\rm good}$ (to be maximized); and a simulated number of contaminant stars that are sufficiently bright to generate background false positives when eclipsed, denoted as $N_{\rm TCE}^{\rm bad}$ (to be minimized). Both metrics depend on NSR, SPR, and simulated frequency of TCEs at 7.1σ; they allow a direct evaluation of the scientific performance of apertures in detecting true and false planet transit signatures. The *Kepler* and TESS missions adopt, analogous to our stellar pollution (SPR), the crowding metric r (Batalha et al. 2010) and the dilution parameter D (Sullivan et al. 2015), respectively, to quantitatively distinguish photometric fluxes originating from targets and other sources. However, these are instrumental level parameters and are not taken into account for choosing their apertures.

From our results we conclude that, compared to the binary mask, weighted masks (gradient and Gaussian) best fit the instrumental PSF at pixel resolution, thus providing lower NSR in

general, but their larger wings inevitably encompass more fractional flux from contaminant stars. From a science perspective, all three mask models present comparable overall efficiency in detecting legitimate planet transits, but the binary mask is substantially (up to ∼30%) less likely to produce background false positives with respect to the weighted masks. These results led us to select the binary mask as the optimal solution for extracting photometry in flight from P5 targets, since this provides the best compromise between maximizing $N_{\rm TCE}^{\rm good}$ and minimizing $N_{\rm TCE}^{\rm bad}$. Besides, this mask model offers a significant implementation advantage, since it requires a relatively small number of unique mask shapes to extract photometry from a large set of stars; we found that about 1350 unique binary masks are sufficient to extract optimal photometry from ∼127k targets.

Our approach currently represents a consistent contribution to the science of exoplanet searches. It confirms that the ordinary concept adopted in the literature for finding apertures, which typically relies on noise minimization for maximum transit detection without directly taking into account the impact from false positives, is not necessarily the best strategy. This statement was initially raised as a hypothesis earlier in this paper, and our results confirm that it holds for the PLATO P5 sample. Indeed, the conventional approach would suggest the use of weighted masks instead of the binary mask.

Beyond the P5 sample, the weighted masks may be exploited as additional photometry extraction methods for the targets whose light curves will be produced from the ground from imagettes. Compared to more complex methods based on PSF fitting photometry (e.g. Libralato et al. 2015; Nardiello et al. 2016), our gradient and Gaussian masks are much simpler and faster to calculate. They might be suitable for not too crowded fields or in situations in which the existence of contaminants may not be too critical (e.g. for asteroseismology targets). We note however that these masks adapt their size to the presence of contaminant stars. This is possible since our expression for the NSR (Eq. (11)) takes into account the fluxes coming from contaminant companions, so whenever their signals are sufficiently strong compared to those of the targets the masks are reduced in width to keep NSR as low as possible. Moreover, our weighted masks can be implemented with ease in both *Kepler* and TESS data processing pipelines, so their usage is not limited to PLATO targets. We expect that the ensemble of results and discussions derived from this work might be particularly useful during the next steps of the preparation phases of the PLATO mission, in particular for the definition of algorithms in the exoplanet validation pipeline, for the construction of the PIC, and later on for the selection of targets.

Finally, despite the relevant contributions of the present study towards minimizing the frequency of background false positives in the P5 sample, a particular concern might still arise with regard the potential difficulties in properly identifying, based on the light curves alone, the false positives from the P5 detections. We highlight however that for an observation scenario covering two long pointing fields the P5 photometry includes, in addition to light curves, a dedicated data share comprising more than 9000 imagettes – with 25 seconds cadence – and COBs for 5% of the targets (see ESA 2017). Allocating these resources to the P5 targets is expected to be flexible enough so that they can be employed following the principle of an alert mode, for example whenever transit signals are detected in the light curves available on the ground. Therefore, the P5 sample will be composed of a photometry extraction method (binary masks) that is intrinsically insensitive to detect most of the potential background false positives, plus a non-negligible number of imagettes and COBs that can be strategically allocated to targets of interest. Overall, that should be enough to identify properly a substantial fraction of the TCEs, which will be dominated by short period transits, in the P5 sample. Aside from that, the PLATO data processing team is currently studying the feasibility and effectiveness of applying imagette-independent methods for identifying background false positives from the P5 detections.

Acknowledgements. This work has benefited from financial support by the French National Centre for Space Studies (CNES), the Paris Observatory – PSL and the Brazilian National Council for Scientific and Technological Development (CNPq). This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This research was achieved using the POLLUX database (http://pollux.graal.univ-montp2.fr) operated at LUPM (Université Montpellier – CNRS, France with the support of the PNPS and INSU. This research uses solar spectrum data from the National Renewable Energy Laboratory, operated for the U.S. Department of Energy (DOE) by the Alliance for Sustainable Energy, LLC ("Alliance"). This research made use of Astropy (http://www.astropy.org), a community-developed core Python package for Astronomy (Astropy Collaboration 2013); Scipy (http://www.scipy.org/), a Python-based ecosystem of open-source software for mathematics, science, and engineering; and SunPy (http://www.sunpy.org), an open-source and free community-developed solar data analysis package written in Python (SunPy Community et al. 2015). The authors express their deep gratitude to the anonymous referee for her/his constructive remarks and suggestion; give special thanks of gratitude to Valerio Nascimbeni for his courtesy in providing the illustration of Fig. 7; and gratefully acknowledge the helpful suggestions by David Brown and Benoît Mosser and the language editing by Amy Mednick.

References

Aigrain, S., Hodgkin, S. T., Irwin, M. J., Lewis, J. R., & Roberts, S. J. 2015, MNRAS, 447, 2880
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411
Batalha, N. M., Borucki, W. J., Koch, D. G., et al. 2010, ApJ, 713, L109
Bessell, M. S. 1990, PASP, 102, 1181
Bohlin, R. C. 2007, in The Future of Photometric, Spectrophotometric and Polarimetric Standardization, ed. C. Sterken, 364, 315
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977
Bryson, S. T., Jenkins, J. M., Gilliland, R. L., et al. 2013, PASP, 125, 889
Bryson, S. T., Tenenbaum, P., Jenkins, J. M., et al. 2010, ApJ, 713, L97
Busso, G., Cacciari, C., Carrasco, J. M., et al. 2018, Gaia DR2 documentation Chapter 5: Photometry, Tech. Rep.
Casagrande, L., & VandenBerg, D. A. 2018, MNRAS, 479, L102
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]
Drummond, R., Lapeyrere, V., Auvergne, M., et al. 2008, A&A, 487, 1209
ESA. 2017, PLATO - Definition Study Report (RED BOOK). ESA-SCI(2017) 1, Tech. rep., European Space Agency
Evans, D. W., Riello, M., Angeli, F. D., et al. 2018, A&A, 616, A4
Fialho, F. D. O., Lapeyrere, V., Auvergne, M., et al. 2007, PASP, 119, 337
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, ApJS, 197, 6
Goupil, M. 2017, EPJ Web conf., 160, 01003
Guyonnet, A., Astier, P., Antilogus, P., Regnault, N., & Doherty, P. 2015, A&A, 575, A41
Handberg, R., & Lund, M. N. 2016, A&A, 597, A36
Howell, S. B. 1989, PASP, 101, 616
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2008, ApJ, 873, 111
Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010a, in Software and Cyberinfrastructure for Astronomy, eds. N. M. Radziwill, & A. Bridger (SPIE)
Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010b, ApJ, 713, L87
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Leinert, C., Bowyer, S., Haikala, L. K., et al. 1998, A&AS, 127, 1

Libralato, M., Bedin, L. R., Nardiello, D., & Piotto, G. 2015, MNRAS, 456, 1137
Llebaria, A., & Guterman, P. 2006, in The CoRoT Mission Pre-Launch Status – Stellar Seismology and Planet Finding, eds. M. Fridlund, A. Baglin, J. Lochard, & L. Conroy, ESA Spec. Pub., 1306, 293
Lund, M. N., Handberg, R., Davies, G. R., Chaplin, W. J., & Jones, C. D. 2015, ApJ, 806, 30
Lund, M. N., Handberg, R., Kjeldsen, H., Chaplin, W. J., & Christensen-Dalsgaard, J. 2017, EPJ Web of Conf., 160, 01005
Massey, R., Schrabback, T., Cordes, O., et al. 2014, MNRAS, 439, 887
Nardiello, D., Libralato, M., Bedin, L. R., et al. 2016, MNRAS, 463, 1831
Naylor, T. 1998, MNRAS, 296, 339
Palacios, A., Gebran, M., Josselin, E., et al. 2010, A&A, 516, A13
Pickles, A. J. 1998, PASP, 110, 863
Rauer, H., Catala, C., Aerts, C., et al. 2014, Exp. Astron., 38, 249
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, J. Astron. Telescopes Instrum. Syst., 1, 014003
Samadi, R., Deru, A., Reese, D., et al. 2019, A&A, 624, A117
Short, A., Crowley, C., de Bruijne, J. H. J., & Prod'homme, T. 2013, MNRAS, 430, 3078
Smith, J. C., Morris, R. L., Jenkins, J. M., et al. 2016, PASP, 128, 124501
Stetson, P. B. 1987, PASP, 99, 191
Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, ApJ, 809, 77
SunPy Community, T., Mumford, S. J., Christe, S., et al. 2015, Comput. Sci. Discovery, 8, 014009
Twicken, J. D., Clarke, B. D., Bryson, S. T., et al. 2010, in Software and Cyberinfrastructure for Astronomy, eds. N. M. Radziwill, & A. Bridger (SPIE)
Twicken, J. D., Catanzarite, J. H., Clarke, B. D., et al. 2018, PASP, 130, 064502
Weiler, M. 2018, A&A, 617, A138
Widenhorn, R. 2010, Opt. Eng., 49, 044401
Wolf, C., Onken, C. A., Luvaul, L. C., et al. 2018, PASA, 35

A&A 624, A117 (2019)
https://doi.org/10.1051/0004-6361/201834822

Astronomy & Astrophysics

The PLATO Solar-like Light-curve Simulator

A tool to generate realistic stellar light-curves with instrumental effects representative of the PLATO mission

R. Samadi[1], A. Deru[1], D. Reese[1], V. Marchiori[1,2], E. Grolleau[1], J. J. Green[1], M. Pertenais[3], Y. Lebreton[1,4], S. Deheuvels[5], B. Mosser[1], K. Belkacem[1], A. Börner[3], and A. M. S. Smith[6]

[1] LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Univ. Paris Diderot, Sorbonne Paris Cité, 5 Place Jules Janssen, 92195 Meudon, France
e-mail: reza.samadi@obspm.fr
[2] Escola Politécnica – Departamento de Engenharia de Telecomunicações e Controle, Universidade de São Paulo, Av. Prof. Luciano Gualberto, 05508-010 São Paulo, Brazil
[3] Institute of Optical Sensors Systems, German Aerospace Center (DLR), Rutherfordstrasse 2, 12489 Berlin, Germany
[4] Univ. Rennes, CNRS, IPR (Institut de Physique de Rennes) – UMR 6251, 35000 Rennes, France
[5] IRAP, Université de Toulouse, CNRS, CNES, UPS, Toulouse, France
[6] Institute of Planetary Research, German Aerospace Center, Rutherfordstrasse 2, 12489 Berlin, Germany

Received 11 December 2018 / Accepted 25 February 2019

ABSTRACT

Context. ESA's PLATO space mission, to be launched by the end of 2026, aims to detect and characterise Earth-like planets in their habitable zone using asteroseismology and the analysis of the transit events. The preparation of science objectives will require the implementation of hare-and-hound exercises relying on the massive generation of representative simulated light-curves.
Aims. We developed a light-curve simulator named the PLATO Solar-like Light-curve Simulator (PSLS) in order to generate light-curves representative of typical PLATO targets, that is showing simultaneously solar-like oscillations, stellar granulation, and magnetic activity. At the same time, PSLS also aims at mimicking in a realistic way the random noise and the systematic errors representative of the PLATO multi-telescope concept.
Methods. To quantify the instrumental systematic errors, we performed a series of simulations at pixel level that include various relevant sources of perturbations expected for PLATO. From the simulated pixels, we extract the photometry as planned on-board and also simulate the quasi-regular updates of the aperture masks during the observations. The simulated light-curves are then corrected for instrumental effects using the instrument point spread functions reconstructed on the basis of a microscanning technique that will be operated during the in-flight calibration phases of the mission. These corrected and simulated light-curves are then fitted by a parametric model, which we incorporated in PSLS. Simulation of the oscillations and granulation signals rely on current state-of-the-art stellar seismology.
Results. We show that the instrumental systematic errors dominate the signal only at frequencies below $\sim 20\,\mu$Hz. The systematic errors level is found to mainly depend on stellar magnitude and on the detector charge transfer inefficiency. To illustrate how realistic our simulator is, we compared its predictions with observations made by *Kepler* on three typical targets and found a good qualitative agreement with the observations.
Conclusions. PSLS reproduces the main properties of expected PLATO light-curves. Its speed of execution and its inclusion of relevant stellar signals as well as sources of noises representative of the PLATO cameras make it an indispensable tool for the scientific preparation of the PLATO mission.

Key words. asteroseismology – stars: oscillations – techniques: image processing – techniques: photometric – methods: numerical

1. Introduction

ESA's PLATO[1] space mission is expected to be launched by the end of 2026 with the goal of detecting and characterising Earth-like planets in the habitable zone of dwarf and sub-giant stars of spectral types F to K (Rauer et al. 2014). The age and mass of planet-hosting stars will be determined by applying stellar seismic techniques to their solar-like oscillations (see e.g. Gizon et al. 2013; Van Eylen et al. 2014, 2018; Huber et al. 2019). The determination of these stellar parameters is a complex procedure since it relies on both the precise seismic analysis of the individual mode frequencies and the use of sophisticated stellar modelling techniques (see e.g. Lebreton et al. 2014a,b). To develop and test such complex procedures, realistic simulated light-curves are needed. These simulated light-curves are, for instance, typically used to conduct hare-and-hounds exercises[2] involving various teams in charge of the seismic analysis and stellar modelling (see e.g. Reese et al. 2016, and references therein). They are also used to conduct massive Monte Carlo simulations that enable one to assess the performances of seismic analysis pipelines (e.g. de Assis Peralta et al. 2018, and

[1] https://platomission.com/

[2] Hare-and-hounds exercises typically involve several teams: one team produces a set of artificial observations while the other teams try to infer the physical model/properties behind these observations.

reference therein). The simulated light-curves must be sufficiently realistic to accurately account for the properties of the modes but also for the other sources of stellar noise such as the granulation noise and the instrumental random noise that – to a large extent – limit the precision of the age and mass determination. Similar hare-and-hounds exercises are also planned to be carried out to test the efficiency of planet detection and the accuracy of the derived transit parameters. Since planetary transits are expected to last several hours, their analysis is quite sensitive to the noises occurring at low frequencies (typically below a few ten of μHz). Finally, simulated light-curves are also used to prepare the analysis of the PLATO light-curves for a variety of other scientific objectives that are also relevant at low frequencies. We can, for instance, mention the characterisation of stellar granulation, the detection and characterisation of rotational modulations, among others. Accordingly, it is necessary to simulate in a realistic way the different sources of noise that dominate the signal at low frequencies. Among them, we have predominantly the stellar activity signal, but systematic instrumental errors may also intervene.

The CoRoT (Baglin et al. 2006b,a) and *Kepler* (Borucki et al. 2010) space missions, allowed us to carry out seismic studies of several thousands of pulsating red-giant stars (De Ridder et al. 2009; Kallinger et al. 2010; Stello et al. 2013) thus enabling important progress in our understanding of stellar interiors (see e.g. the reviews by Mosser & Miglio 2016; Hekker & Christensen-Dalsgaard 2017). These observations opened up the path to what we now call ensemble asteroseismology (see e.g. Huber et al. 2011; Belkacem et al. 2013; Miglio et al. 2015) with various applications in the field of Galactic archaeology (Miglio et al. 2017). PLATO can potentially observe a large number of faint red giants. The number of targets that can be observed in addition to the targets of the core program is nevertheless limited to about 40 000 per pointing. An optimal choice of those targets can rely on the seismic performance tool of Mosser et al. (2019). On the other hand, the design and the development of seismic analysis pipelines that are able to process in an automatic way a large number of red giants require the generation of simulated light-curves representative of such stars.

To our knowledge the light-curve simulator developed by De Ridder et al. (2006) in the framework of the Eddington space project is the first code made available to the community that simulates solar-like oscillations together with the stellar granulation noise and the instrumental sources of noise. This simulator relies on a description of the modes and stellar granulation noise that predates CoRoT and *Kepler* space missions. However, our knowledge of solar-like oscillations and stellar granulation has greatly improved since that time. Very recently, Ball et al. (2018) proposed a light-curve simulator dedicated to the TESS mission and that includes an up-to-date description of solar-like oscillators and the granulation background. However, in this simulator, white noise is the only non-stellar source of noise; this means that systematic errors are not included. However, the latter, which are very specific to a given instrument and its space environment, are in general frequency dependent and can only be realistically quantified with simulations made at detector pixel level. Furthermore, the level of the white noise (random noise) also strongly depends on the implemented photometry method and the performance of the instrument. Finally, these simulators do not include planetary transits and are not suited for red giant stars. Indeed, red giants show the presence of numerous mixed-modes, and calculating mixed-mode frequencies with pulsation codes requires a very high number of mesh points in the stellar models thus making the massive generation of corresponding simulated light-curves numerically challenging.

The PLATO mission has some characteristics that make it very different from other space-based mission based on high-precision photometry such as CoRoT, *Kepler* or TESS. Indeed, one of the main specificities of the mission is that it relies on a multi-telescope concept. Among the 26 cameras that compose the instrument, two of them are named "fast" cameras and work at a 2.5 s cadence while the remaining 24 are named "normal" cameras and work at a 25 s cadence. The normal cameras are divided into four groups of six cameras, with large fields of view (~1100 square degrees) that partially overlap. Each camera is composed of four Charge Couple Devices (CCD hereafter) which are read out at the cadence of 25 s with a time-shift of 6.25 s between each of them. Accordingly, the observations made for a given target by various groups of camera will be time-shifted thereby allowing us to perform super-Nyquist seismic analysis (Chaplin et al. 2014). Because of the large field of view and the long-term change of the pointing direction of each individual camera, star positions will slowly drift on the camera focal plane by up to 1.3 pixels during the 3-month uninterrupted observation sequences. As a consequence, stars will slowly leave the aperture photometry (i.e. masks), leading obviously to a long-term decrease of their measured intensities. Furthermore, during the life of the mission, the instrument will be continuously exposed to radiation (mostly proton impacts). This will generate more and more traps in the CCD thus increasing the Charge Transfer Inefficiency (CTI hereafter, see e.g. Massey et al. 2014, and references therein) over time. Coupled with the long-term drift of the stellar positions, the CTI will induce an additional long-term variability of the photometric measurements.

To mitigate the flux variations induced by the instrument and the observational conditions, the aperture masks used on-board will be updated on a quasi-regular basis. This will nevertheless leave residual flux variations of about several % over three months, which remain high w.r.t. the science requirements. The residual flux variations will fortunately be corrected a posteriori on-ground on the basis of the knowledge of the instrumental point spread function (PSF). Nevertheless, such a correction will leave systematic errors in the power spectrum that will rapidly increase with decreasing frequency. All of these instrumental systematic errors together with the stellar activity noise component can in principle impact the detection and characterisation of the planetary transits, limit the seismic analysis of very evolved red giant stars, and affect any science analysis of the signal at rather low frequencies.

The Plato Stellar Light-curve Simulator[3] (PSLS) aims at simulating stochastically-excited oscillations together with planetary transits, stellar signal (granulation, activity) and instrumental sources of noise that are representative of the PLATO cameras. The simulator allows us to simulate two different types of oscillation spectra: (i) oscillation spectra computed on the basis of the so-called Universal Pattern by Mosser et al. (2011) optionally including mixed-modes following the asymptotic gravity mode spacing (Mosser et al. 2012b) and (ii) oscillation spectra computed using a given set of theoretical

[3] The PSLS source code is available for download from the PSLS website (`http://psls.lesia.obspm.fr`) as well as from Zenodo.org (`http://doi.org/10.5281/zenodo.2581107`). The source code is free: you can redistribute it and/or modify it under the terms of the GNU General Public License (for more details see `http://www.gnu.org/licenses`). The present paper describes the version 0.8.

frequencies pre-computed with the ADIPLS pulsation code (Christensen-Dalsgaard 2008).

The instrumental noise level is quantified by carrying out realistic simulations of the instrument at CCD pixel level using the Plato Image Simulator (PIS) for three-month observation sequences. These simulations are performed for different stellar magnitudes, and for both the beginning of life (BOL[4]) and end of life (EOL[5]) observation conditions. The photometry is extracted from these simulated images in the same manner as planned on-board, that is using binary masks that minimise the noise-to-signal ratio (NSR) of each target. The corresponding simulated light-curves are then corrected using PSFs reconstructed on the basis of a microscanning technique, which will be operated in-flight before each three-month observation sequence and which we also simulate in the present work. This set of simulated light-curves, corrected for the instrumental errors, then enables us to quantify the expected level of residual systematic errors. These simulations are then used to derive – as a function of the stellar magnitude – a parametric model of the residual errors in the time domain. This model is in turn implemented into PSLS.

Finally, the other components of the stellar signal (granulation signal, and planetary transits) are included in PSLS following prescriptions found in the literature.

2. General principle

The stochastic nature of the different phenomena (i.e. white noise, stellar granulation and stochastically-excited oscillations) are simulated following Anderson et al. (1990, see also Baudin et al. 2007). As detailed below, the properties of the simulated stellar signal are first modelled in the Fourier domain, we next add a random noise to simulate the stochastic nature of the signal, and finally we perform an inverse Fourier transform to come back into the time domain and derive the corresponding time-series (i.e. light-curve). We note that other authors (e.g. Chaplin et al. 1997; De Ridder et al. 2006) have proposed instead to work directly in the time domain. Although, rigorously equivalent, it is more convenient to describe the stellar signal in the Fourier domain since this is the common way signals (such as pulsation, granulation, and activity) are analysed in solar-like pulsators.

Let $\mathcal{F}(\nu)$ be the Fourier Transform (FT hereafter) of the simulated light-curve $\mathcal{S}(t)$, and $\overline{\mathcal{P}}(\nu)$ the expectation of the Power Spectral Density (PSD) associated with the stellar signal (i.e. the PSD one would have after averaging over an infinite number of realisations). If the frequency bins of the PSD are uncorrelated, we can then show that

$$\mathcal{F}(\nu) = \sqrt{\overline{\mathcal{P}}}\,(u + i\,v)\,, \tag{1}$$

where u, and v are two uncorrelated Normal distributions of zero mean and unit variance, and i is the imaginary unit ($i^2 = -1$). We finally compute the inverse Fourier Transform of $\hat{F}(\nu)$ to derive the simulated light-curve $\mathcal{S}(t)$ for a given realisation. We note that the PSD $\mathcal{P}(\nu)$ associated with a given realisation verifies

$$\mathcal{P}(\nu) = |\mathcal{F}(\nu)|^2 = \overline{\mathcal{P}}\left(u^2 + v^2\right). \tag{2}$$

Our PSD is "single-sided", which means that the integral of the PSD from $\nu = 0$ (excluded) to the Nyquist frequency is equal to the variance of the time-series.

[4] I.e. in the absence of CTI.

[5] I.e. with the level of CTI expected at the end of the mission, that is 6 years after launch by definition.

Here, the expectation $\overline{\mathcal{P}}(\nu)$ is the sum of an activity component $\mathcal{A}(\nu)$, the granulation background $G(\nu)$, and the oscillation spectrum $O(\nu)$, that is

$$\overline{\mathcal{P}}(\nu) = \mathcal{A}(\nu) + G(\nu) + O(\nu). \tag{3}$$

In accordance with our initial hypothesis, all these components are uncorrelated. However, some interferences can in principle exist between the various stellar signal components, such as the activity, the granulation and the oscillations. For instance there are some observational evidences about correlations between granulation (i.e. convection) and modes. Indeed, solar mode profiles slightly depart from symmetric Lorentzian profiles (Duvall et al. 1993). Likewise, pieces of evidence for similar asymmetries were recently found in stars observed by *Kepler* (Benomar et al. 2018). Helioseicmic data clearly show that this asymmetry is reversed between velocity and intensity measurements (e.g. Duvall et al. 1993; Nigam et al. 1998; Barban et al. 2004). This reversal is believed to be the signature of a correlation between convection and oscillations (Roxburgh & Vorontsov 1997; Nigam et al. 1998). However, the departures from symmetric Lorentzian profiles are small w.r.t. the mode linewidths. Hence, we consider this as an indication of a small level of correlation between convection (i.e. granulation) and oscillations. Finally, concerning possible interferences between activity and convection, to our knowledge there are no pieces of evidence. For these reasons, in this work, we decided to neglect the correlations between the stellar signal components.

Once the FT associated with the stellar signal is simulated on the basis of Eq. (1), we perform an inverse Fourier transform to come back into the time domain. This then provides the stellar signal as a function of time. However, in order to take into account the fact that each group of cameras are time-shifted by $\Delta t = 6.25$ s, we multiply Eq. (1) by the phase term $e^{i2\pi\Delta t}$ prior to calculating its inverse Fourier Transform.

The instrumental signal component (i.e. the systematic errors plus the instrumental random sources of noise) is simulated in the time domain as explained in Sect. 4. Finally, once the instrumental signal is simulated, it is multiplied by the stellar signal and the planetary transit (which as the instrumental component is simulated in the time domain) to get finally the simulated light-curve averaged over a given number of cameras. We describe in the following sections the way each simulated component is modelled.

3. Solar-like oscillations

In this section, we describe the modelling of the oscillation spectrum $O(\nu)$. It is the sum over the different normal modes

$$O(\nu) = \sum_i \mathcal{L}_i(\nu), \tag{4}$$

where each individual resolved mode of frequency ν_i is described by a Lorentzian profile

$$\mathcal{L}_i(\nu) = \frac{H_i}{1 + \left(2\,(\nu - \nu_i)\,/\Gamma_i\right)^2}, \tag{5}$$

where H_i is the mode height, and Γ_i its linewidth. A mode is considered to be resolved when $\Gamma_i > 2\delta f$ where δf is the frequency resolution (or equivalently the inverse of the observation duration). In contrast, for an unresolved mode the profile is given by (see, e.g. Berthomieu et al. 2001),

$$\mathcal{L}_i(\nu) = \frac{\pi\,\Gamma_i\,H_i}{2\delta\nu}\,\mathrm{sinc}^2\left[\pi\,(\nu - \nu_i)\right], \tag{6}$$

where $\delta\nu$ is the resolution of the spectrum.

To go further, one needs to determine the mode frequencies, heights, and line-widths. To do so, we consider two different methods for the frequencies. For main-sequence and sub-giant stars, the method consists in computing a set of theoretical mode frequencies using the ADIPLS adiabatic pulsation code while for red giant stars we consider the method developed by Mosser et al. (2011), which relies on what is commonly known as the Universal Pattern. This distinction is motivated by the difficulty to compute red giant frequencies. Indeed, for evolved stars, a proper modelling of the normal frequencies requires an important number of grid points in the innermost layers. While still feasible, this makes the computation more demanding. We therefore adopt a more flexible and affordable method based on asymptotic considerations to ensure the possibility of using the simulator on a massive scale.

3.1. Main-sequence and sub-giant stars

The oscillation spectrum is constructed using a set of theoretical eigenfrequencies computed using the ADIPLS code (Christensen-Dalsgaard 2008). The program allows one to include uniform rotational splittings as specified by an input surface rotation period $T_{\rm rot} = 2\pi/\Omega_{\rm surf}$ where $\Omega_{\rm surf}$ is the surface rotation rate. The set of frequencies included in the model are

$$\nu^{(0)}_{n,\ell,m} = \nu_{n,\ell} + \frac{m}{T_{\rm rot}}\,(1 - c_{n,\ell}), \tag{7}$$

where n is the radial order, ℓ the angular (or harmonic) degree, m the azimuthal order, and $c_{n,\ell}$ the Ledoux constant (see, e.g. Unno et al. 1989) provided by ADIPLS. We consider all the modes from $n = 1$ up to the cut-off frequency, with angular degrees ranging from $\ell = 0$ to 3 inclusive. Near-surface effects are eventually added using the empirical correction proposed by Sonoi et al. (2015):

$$\nu_{n,\ell,m} = \nu^{(0)}_{n,\ell,m} + a\,\nu_{\rm max}\left(1 - \frac{1}{1 + \left(\nu^{(0)}_{n,\ell,m}/\nu_{\rm max}\right)^b}\right), \tag{8}$$

where a and b are two parameters, which are expressed in terms of $T_{\rm eff}$ and $\log\, g$ thanks to the scaling laws provided in Eqs. (10) and (11) of Sonoi et al. (2015), respectively.

The mode height of each given mode is computed according to

$$H_{n,\ell,m} = G(\nu_{n,\ell,m}; \delta\nu_{\rm env})\; V_\ell^2\; r^2_{n,\ell,m}(i)\, H_{\rm max}, \tag{9}$$

where V_ℓ is the mode visibility ($V_0 = 1$, $V_1 = 1.5$, $V_2 = 0.5$, $V_3 = 0.05$), $H_{\rm max}$ the mode height at the peak frequency, and $r_{n,\ell,m}$ the (relative) visibility of a mode of azimuthal order m within a multiplet for a given inclination angle i. The ratio $r_{n,\ell,m}$ is computed according to Dziembowski (1971, see also Gizon & Solanki 2003) and represents – at fixed values of n and ℓ – the ratio of the mode height for a given inclination angle i to the mode height at $i = 0°$. Finally, G is the Gaussian envelope defined as

$$G(\nu_{n,\ell,m}; \delta\nu_{\rm env}) = \exp\left[\frac{-(\nu_{n,\ell,m} - \nu_{\rm max})^2}{\delta\nu^2_{\rm env}/4\ln 2}\right], \tag{10}$$

where $\delta\nu_{\rm env}$ is the full width at half maximum, which is supposed to scale as (Mosser et al. 2012a):

$$\delta\nu_{\rm env} = 0.66\,\nu^{0.88}_{\rm max}. \tag{11}$$

This scaling relation was established for red giants. The applications presented in Sect. 6 show that it provides rather good results for less evolved stars.

To compute Eq. (9), we now need to specify $H_{\rm max}$. For a single-side PSD, the mode height is related to the mode linewidth as (see, e.g. Baudin et al. 2005)[6]

$$H_{\rm max} = \frac{2\,A^2_{\rm max}}{\pi\,\Gamma_{\rm max}}, \tag{12}$$

where $A_{\rm max}$ is the rms of the mode amplitude at the peak frequency. The latter is related to the bolometric amplitude $A_{\rm max,bol}$ using the correction proposed for *Kepler*'s spectral band by Ballot et al. (2011)

$$A_{\rm max} = A_{\rm max,bol}\left(\frac{T_{\rm eff}}{5934\,{\rm K}}\right)^{-0.8}. \tag{13}$$

We note that the CoRoT spectral band results in very similar corrections (see Michel et al. 2009). Finally, $A_{\rm max,bol}$ is derived from the scaling relations derived by Corsaro et al. (2013) and defined as

$$\begin{aligned}\ln(A_{\rm max,bol}) = \ln(A_{\rm max,bol,\odot}) &+ (2s - 3t)\ln(\nu_{\rm max}/\nu_{\rm max,\odot})\\ &+ (4t - 4s)\ln(\Delta\nu/\Delta\nu_\odot)\\ &+ (5s - 1.5t - r + 0.2)\ln(T_{\rm eff}/T_{\rm eff,\odot}) + \ln(\beta),\end{aligned} \tag{14}$$

where $A_{\rm max,bol,\odot} = 2.53$ ppm (rms) is the maximum of the bolometric solar mode amplitude (Michel et al. 2009), and s, t, r and β are coefficients that depend on the star's evolutionary status (see Tables 3 and 4 in Corsaro et al. 2013).

Finally, one needs to specify the mode line-widths. To this end, we note that the product of the mode line-width and the mode inertia has a parabolic shape (Belkacem et al. 2011, see Fig. 2). Therefore,

$$\Gamma_{n,\ell,m} = \Gamma_{\rm max}\left(\frac{I_{\rm max}}{I_{n,\ell}}\right)\gamma(\nu_{n,\ell,m}), \tag{15}$$

where $I_{n,\ell}$ is the mode inertia, $I_{\rm max}$ is the mode inertia of the radial modes interpolated at $\nu = \nu_{\rm max}$, $\Gamma_{\rm max}$ is the mode linewidth at $\nu = \nu_{\rm max}$ derived from two different scaling relations (see below), and the function $\gamma(\nu)$ models the frequency dependence of the product $\Gamma_{n,\ell,m}I_{n,\ell}$ around $\nu_{\rm max}$. The latter is modelled empirically as follows

$$\gamma(\nu) = 1 + A\,\left(1 - G(\nu_{n,\ell,m}; 2\delta\nu_{\rm env})\right), \tag{16}$$

where G is the Gaussian function defined by Eq. (10), A is a constant, and $\delta\nu_{\rm env}$ is given by the scaling relation of Eq. (11). With $A = 2$ for $\nu \geq \nu_{\rm max}$ and $A = 6$ for $\nu < \nu_{\rm max}$, Eq. (16) reproduces rather well the variation with frequency of the solar mode linewidths. Given the objectives targeted by the simulator, we assume that this empirical description is sufficiently representative for other stars. An alternative approach would have been to use the relation describing the frequency dependence derived from *Kepler* observations by Appourchaux et al. (2014, see its corrigendum in Appourchaux et al. (2016)). However, this relation was established for a limited number of targets and hence in limited ranges in effective temperatures, surface gravities and surface metal abundances. Therefore, to avoid extrapolations we prefer to adopt Eq. (15). In addition, the relation inferred by

[6] The additional factor of two comes from the fact we assume here a single-sided PSD while Baudin et al. (2005) assumed a double-sided one.

Appourchaux et al. (2014) was established on limited frequency intervals. Since the mode line-widths scale as the inverse of the mode inertia (Eq. (15)), this scaling relation allows us instead to derive the frequency dependence of $\Gamma_{n,\ell,m}$ for the whole acoustic spectrum of a given star.

Finally, the mode line-width at the peak frequency, $\Gamma_{\rm max}$, is determined on the basis of the scaling relation derived by Appourchaux et al. (2012) from main-sequence *Kepler* targets, that is

$$\Gamma_{\rm max} = \Gamma_{\rm max,0} + \beta \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}} \right)^s , \tag{17}$$

where $\Gamma_{\rm max,0} = 0.20\,\mu$Hz, $\beta = 0.97$, and $s = 13.0$.

3.2. Red-giant stars

Each mode frequency $\nu_{n,\ell}$ is computed according to the Universal Pattern proposed by Mosser et al. (2011)

$$\nu_{n,\ell,m} = n + \frac{\ell}{2} + \varepsilon(\Delta\nu) - d_{0\ell}(\Delta\nu) + \frac{\alpha_\ell}{2}\left(n - \frac{\nu_{\rm max}}{\Delta\nu}\right)^2 \Delta\nu + m\,\delta\nu_{\rm rot} + \delta_{n,\ell}, \tag{18}$$

where ε is an offset, $d_{0\ell}$, the small separation, α_ℓ the curvature, $\Delta\nu$ the large separation, $\delta\nu_{\rm rot}$ the rotational mode splitting (included only for dipolar modes, as will be explained later on), and finally $\delta_{n,\ell}$ a term that accounts for a possible coupling with the gravity modes, which results in the deviation of the mode frequency from its uncoupled solution ("pure" acoustic mode) and gives the mode its mixed-mode nature. For a dipole mode, $\delta_{n,\ell}$ is computed according to the asymptotic gravity-mode spacing (Mosser et al. 2012b)

$$\delta_{n,\ell} = \frac{\Delta\nu}{\pi} \arctan\left[q \tan\pi\left(\frac{1}{\Delta\Pi_1 \nu_{n,\ell}} - \epsilon_g \right) \right], \tag{19}$$

where q is the coupling coefficient, $\Delta\Pi_1$ the asymptotic period spacing of the (pure) dipole g modes, and ϵ_g an offset fixed to the value 0.25, which is representative for most red giants (Mosser et al. 2017). For radial modes, one obviously has $\delta_{n,0} = 0$, while for all modes with angular degree $\ell \geq 2$ we neglect the deviation and assume $\delta_{n,\ell} = 0$.

The mode height of each given mode (n, ℓ, m) is given by

$$H_{n,\ell} = G(\nu_{n,\ell})\, V_\ell^2\, H_{\rm max}, \tag{20}$$

where $G(\nu_{n,\ell})$ is given by Eq. (15), V_ℓ, is the mode visibility determined from Mosser et al. (2012a) and $H_{\rm max}$ is the maximum of the mode heights derived from the scaling relation established by de Assis Peralta et al. (2018), that is

$$H_{\rm max} = 2.01 \times 10^7\, \nu_{\rm max}^{-1.9}. \tag{21}$$

Concerning the mode linewidths $\Gamma_{n,\ell}$, they are assumed to be constant with frequency. This assumption is motivated by the fact that modes are observed in a relatively small frequency range compared to main-sequence and sub-giant stars. This constant value is determined from the theoretical scaling relation of Vrard et al. (2018), which depends on the effective temperature, $T_{\rm eff}$, and stellar mass as follows

$$\Gamma_{\rm max} = \Gamma_{\rm max,0} \left(\frac{T_{\rm eff}}{4800\,{\rm K}} \right)^{\alpha_T}, \tag{22}$$

where $\Gamma_{\rm max,0} = 0.1\,\mu$Hz and α_T is a coefficient which depends on the stellar mass range (see Vrard et al. 2018). The dipolar mixed modes have, however, much smaller line-widths than their associated "pure" acoustic modes. This is mainly because their inertia is much larger as a consequence of the fact they behave as gravity modes in the inner layers. Indeed, the mode line-width scales as the inverse of the mode inertia (see, e.g., Belkacem & Samadi 2013). Let $I^m_{n,\ell}$ (resp. $\Gamma^{(m)}_{n,\ell}$) be the mode inertia (resp. mode line-width) of a dipolar mixed-mode and $I^0_{n,\ell}$ (resp. $\Gamma^{(0)}_{n,\ell}$) that of a "pure" acoustic mode of the same radial order. We then have

$$\Gamma^{(m)}_{n,\ell} = \Gamma^{(0)}_{n,\ell} \left(\frac{I^0_{n,\ell}}{I^m_{n,\ell}} \right), \tag{23}$$

where according to our previous assumption $\Gamma^{(0)}_{n,\ell} = \Gamma_{\rm max}$ for any couple (n, ℓ). In Eq. (23), it is assumed that radiative damping in the radiative interior of red giants is negligible. The validity of this assumption has been thoroughly investigated by Grosjean et al. (2014).

To go further, we use the following relation from Goupil et al. (2013):

$$\frac{I^0_{n,\ell}}{I^m_{n,\ell}} \simeq 1 - \frac{I_{\rm core}}{I} = 1 - \zeta, \tag{24}$$

where $I_{\rm core}$ is the contribution of the core to the mode inertia, and ζ is calculated according to Eq. (4) in Gehan et al. (2018). Finally, the rotational splitting for dipolar modes (the term $\delta\nu_{\rm rot}$ in Eq. (18)) is computed on the basis of Eq. (22) in Goupil et al. (2013) by neglecting the surface rotation (see e.g. Mosser et al. 2015; Gehan et al. 2018). Accordingly, we have

$$\delta\nu_{\rm rot} = \frac{\zeta}{2} \left(\frac{\Omega_{\rm core}}{2\pi} \right), \tag{25}$$

where $\Omega_{\rm core}$ is the core rotation rate (in rad/s).

The oscillation spectrum is then constructed by summing a Lorentzian profile for each mode. We include modes with radial orders ranging from $n = 1$ up to $n =$ integer $(\nu_{\rm c}/\Delta\nu)$, where $\nu_{\rm c}$ is the cutoff-frequency (see Eq. (28)), and with angular degrees from $\ell = 0$ to $\ell = 3$.

The simulator requires three main input parameters, $\nu_{\rm max}$, $T_{\rm eff}$ and $\Delta\nu$, from which all the other parameters are established using scaling relations, except $\Delta\Pi_1$ and q which can be provided as optional inputs (otherwise no mixed modes are included). In case $\Delta\nu$ is not provided, it is computed according to the scaling relation (Mosser et al. 2013)

$$\Delta\nu = 0.274\, \nu_{\rm max}^{0.757}. \tag{26}$$

The stellar mass used for the granulation scaling relations is determined by combining the scaling relation for $\nu_{\rm max}$ and $\Delta\nu$ (see Belkacem 2012; Mosser et al. 2010, and references therein):

$$m = M_\odot \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}} \right)^3 \left(\frac{\Delta\nu}{\Delta\nu_\odot} \right)^{-4} \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}} \right)^{3/2}. \tag{27}$$

Finally, the cutoff frequency $\nu_{\rm c}$ is derived from the following scaling relation:

$$\nu_{\rm c} = \nu_{\rm c,\odot}\, \frac{g}{g_\odot} \sqrt{\frac{T_{\rm eff,\odot}}{T_{\rm eff}}}, \tag{28}$$

where $\nu_{\rm c,\odot} = 5300\,\mu$Hz.

4. Instrumental errors

Our objective here is to quantify the instrumental sources of error, namely the systematic error and the random noise, and to implement them into PSLS. For the former, a set of simulations at CCD pixel level is carried out while for the random noise we rely on the work made by Marchiori et al. (2019) as explained in Sect. 4.4.

4.1. The Plato instrument

PLATO is composed of 24 cameras (named normal cameras) working at a cadence of 25 s and two cameras (named fast cameras) working at a cadence of 2.5 s. Each group of cameras is composed of six normal cameras that see half of the full field of view (2200 square degrees). The fast cameras point towards the centre of the field of view, and provide the platform with pointing errors for the Attitude Control System. Four Charge Coupled Devices (CCDs) are mounted on the focal plane of each camera. The pixels have a size of 18 μm and their projected size in the sky represents approximately 15 arcsec.

Every three months, the platform is rotated by 90° in order to maintain the solar panel in the direction of the Sun. Due to the thermal distortion of the platform, changes in the pointing direction of each individual camera are expected during the uninterrupted three-month observation sequences. These variations will lead to long-term star drifts on the focal plane of up to 0.8 pixels in three months. Furthermore, because of the large field of view, the kinematic aberration of light will induce drifts of the stellar positions of up to 0.5 pixels in three months at the edge of the field of view. Both effects add together and result in drifts of up to 1.3 pixels in three months (in the worst case, at the edge of field of view).

4.2. The Plato Image Simulator

To quantify the instrumental systematic errors, we generate time-series of small imagettes with the Plato Image Simulator (PIS). This simulator, developed at the LESIA-Observatoire de Paris since the early phases of the PLATO project, has very similar capabilities as the PLATOSim code (Marcos-Arenal et al. 2014). PIS can simulate imagettes representative of PLATO CCDs. It includes various sources of perturbations, such as shot-noise (photon noise), readout noise, background signal, satellite jitter, long-term drift, smearing, digital saturation, pixel response non-uniformity (PRNU), intra pixel response non-uniformity (IPRNU), charge diffusion, and charge transfer inefficiency (CTI). Since our goal is to quantify systematic errors, we turned off all random sources of noise in our instrumental simulations, except in the calculation of the NSR, see Sect. 4.4; these are the shot-noise, the readout-noise, and the satellite jitter. CTI is simulated following Short et al. (2013) and activated for end-of-life (EOL) simulations only. Charge diffusion within the CCD pixels is not activated because we still lack a reliable estimate of its amplitude (see the discussion in Sect. 7).

To take into account the impact of long-term drifts of the stellar positions, simulations are generated over 90 days and include a linear drift of 1.3 pixels in three months. To be more realistic, the instrumental point spread functions (PSF) used during these simulations include optical manufacturing errors and integration and alignment tolerances to the nominal design for the nominal focus position. These input PSFs do not include effects due to the detector or the spacecraft (such as the satellite jitter). However, most of them (like PRNU, IPRNU, CTI, and satellite jitter) are in any case included in PIS.

Table 1. Simulation parameters used with the PIS code.

Parameters		Value
Reference flux at $V = 11$	BOL	2.17×10^5 e-/exp.
	EOL	2.13×10^5 e-/exp.
Sky background		120 e-/s/pixels
PRNU		1.00%
IPRNU		0.50%
Integration time		21s
Readout time		4s
Gain		25 e-/ADU
Electronic offset		1000 ADU
Photon noise		Disabled
Readout noise		Disabled
e-Satellite jitter		Disabled

4.3. Simulation parameters and data sets

The flux of each simulated star behaves differently according to their magnitude, position over the CCD, and even position within a pixel (hereafter named intra-pixel position). In order to cover the largest combination of these factors, we use PIS to run 630 artificial star simulations using a combination of:

- 9 stellar magnitudes (from $V = 9$ to $V = 13$ with a step of 0.5),
- 14 focal plane positions over the focal plan (from 1.41° to 18.08° from the optical centre),
- 5 intra-pixel positions for each of the 14 focal plane positions.

The simulations are carried out using the parameters relevant for BOL and EOL conditions. Thus, regarding EOL simulations, the CTI is enabled and the mean optical transmission is assumed to be lower than the BOL one. The CTI model used by PIS requires specifying the number of trap species and their characteristics in terms of density, release time, and cross sections. To this end, Prod'homme et al. (2016) have studied CTI on a representative PLATO CCD that has been irradiated on purpose. This study allowed the authors to identify four trap species and to calibrate their corresponding parameters. We used the parameters derived by Prod'homme et al. (2016). However, the trap densities are re-scaled so that the level of CTI reaches the mission specifications at the EOL. The adopted values of the simulation parameters are reported in Table 1.

4.4. Photometry extraction

Of the ~120 000 targets observed by each camera during a given pointing, about 14 000 of them will have their 6×6 imagettes downloaded on-ground at a cadence of 25 s. For these targets, the photometry will be extracted on-ground on the basis of more sophisticated methods, which are not yet fully established. The photometry of the remaining targets will necessarily have to be performed on-board.

Before computing the photometry, we start with a basic pre-processing of the imagettes aiming to subtract the electronic offset and the background, convert ADU to electrons using the gain, and finally subtract the smearing for each column of the imagette.

Photometry extraction is performed on-board by integrating the stellar flux over a collection of pixels called the aperture or the mask. Different strategies for determining the most adequate aperture shape have been the subject of a detailed study

(Marchiori et al. 2019), leading to the adoption of binary masks as the best compromise between NSR and stellar contamination ratio. For a given target, its associated binary mask is defined as the subset of the imagette pixels giving the minimum noise-to-signal ratio. It is computed through the following scheme

1. Arrange all pixels n from the target imagette in increasing order of noise-to-signal ratio NSR_n

$$\mathrm{NSR}_n = \frac{\sqrt{\sigma^2_{F_{T_n}} + \sum_{k=1}^{N_C} \sigma^2_{F_{C_{n,k}}} + \sigma^2_{B_n} + \sigma^2_{D_n} + \sigma^2_{Q_n}}}{F_{T_n}}. \quad (29)$$

2. Compute the aggregate noise-to-signal $\mathrm{NSR}_{\mathrm{agg}}(m)$, as a function of the increasing number of pixels $m = \{1, 2, 3, \ldots, 36\}$, stacking them conforming to the arrangement in the previous step and starting with the pixel owning the smallest NSR_n

$$\mathrm{NSR}_{\mathrm{agg}}(m) = \frac{\sqrt{\sum_{n=1}^{m} \left(\sigma^2_{F_{T_n}} + \sum_{k=1}^{N_C} \sigma^2_{F_{C_{n,k}}} + \sigma^2_{B_n} + \sigma^2_{D_n} + \sigma^2_{Q_n} \right)}}{\sum_{n=1}^{m} F_{T_n}}. \quad (30)$$

3. Define as the aperture the collection of pixels m providing minimum $\mathrm{NSR}_{\mathrm{agg}}(m)$.

In Eqs. (29) and (30), F_{T} is the target star's mean flux, $\sigma_{F_{\mathrm{T}}}$ the target star's photon noise, F_{C} the contaminant star's mean flux, $\sigma_{F_{\mathrm{C}}}$ the contaminant star's photon noise, σ_{B} the background noise from the zodiacal light, σ_{D} the overall detector noise (including readout, smearing and dark current noises) and σ_{Q} the quantization noise. Figure 1 illustrates how the NSR typically evolves as the binary mask gets larger following the above scheme. We note that the noise due to satellite jitter is not included in the definition of the mask (Eq. (30)). Including the contribution of the jitter noise in the definition of the mask is not trivial because its contribution depends on the final shape of the mask (see e.g. Fialho & Auvergne 2006). Nevertheless, it turns out that for PLATO, the jitter noise is small enough that its does not play a role in the mask shape, the dominant sources of noise being the photon noise for brighter stars and the background and readout noise for fainter stars. Accordingly, once the mask is defined, we include the jitter a posteriori in the estimation of the NSR.

The NSR was estimated for a large sample of targets (so far about 50 000) with magnitudes ranging from 9 to 13. The targets and their associated contaminant stars were extracted from the *Gaia* DR2 catalogue (Gaia Collaboration 2018). In total about 3.5 million contaminant stars with magnitudes up to $G = 21$ were included in the calculation.

The calculation of the NSR takes into account the various sources of noise described above and also the fact that the shape of the PSF varies across the field of view. The latest version of the instrument parameters were also considered (details will be given in Marchiori et al. 2019). Typical values of the NSR are given in Table 2 for a single camera and 24 cameras as a function of the PLATO magnitude, P, which is defined in Marchiori et al. (2019) and is by definition directly connected to the flux collected by a PLATO camera. For comparison with the mission specifications (Rauer et al. 2014), we also provide the V magnitude, which is defined here as the flux collected in the Johnson V filter for a reference PLATO target of $T_{\mathrm{eff}} = 6000$ K.

For comparison, we also reported the values of the NSR in the photon noise limit (i.e. when there is only the photon noise due to the target). The relative contribution of random noises that add (quadratically) to the target photon noise increase with increasing stellar magnitude from 20% at magnitude $V = 8.5$ up to 65% at $V = 13$.

Unless the NSR value is imposed by the user, the latter is obtained by interpolating the values given in Table 2 for a given V magnitude.

Table 2. NSR as a function of target V and P magnitudes.

V	P	NSR 1 camera (ppm h$^{1/2}$)	NSR 24 cameras (ppm h$^{1/2}$)	Photon noise limit 24 cameras (ppm h$^{1/2}$)
8.1	7.76	51.9	**10.6**	10.5
8.5	8.16	63.2	**12.9**	12.7
9.0	8.66	80.3	**16.4**	16.0
9.5	9.16	101.9	**20.8**	20.1
10.0	9.66	130.8	**26.7**	25.4
10.5	10.16	169.0	**34.5**	32.2
11.0	10.66	219.5	**44.8**	40.8
11.5	11.16	290.0	**59.2**	52.0
12.0	11.66	387.5	**79.1**	66.1
12.5	12.16	523.2	**106.8**	84.3
12.9	12.56	678.5	**138.5**	102.6

Notes. The values are given for a single camera and for 24 cameras, and were extracted from Marchiori et al. (2019). The rightmost column gives the photon noise limit, that is the NSR one would have if we were limited only by the photon noise of the target. Values in boldface represent the performance in terms of NSR to be archived with stars observed by all the 24 cameras. Values given for a single camera should not be used as a reference for the mission performance.

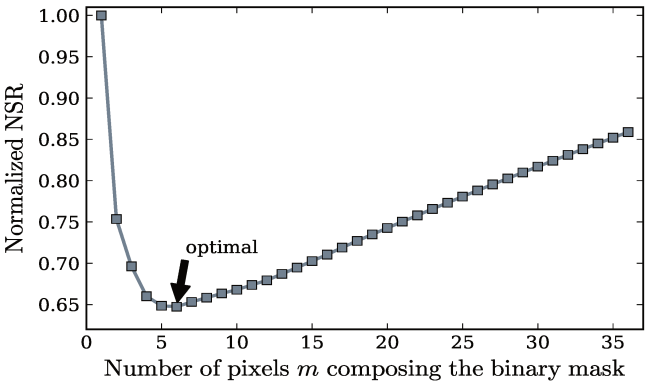


Fig. 1. Typical NSR evolution curve as a function of increasing aperture size. Pixels with the smallest noise-to-signal ratio are added-up successively. The collection of pixels giving the lowest aggregate NSR defines the binary mask for a given target.

4.5. Mask update

An example of a light-curve obtained with a fixed optimal binary mask is shown in Fig. 2 (top) for a target of magnitude V=11. Because of the long-term drift of the star and the fact that the mask is maintained at the same position during the three-month observation sequence, we observe a significant long-term decrease of the stellar flux. In this worst case scenario (a displacement of 1.3 pixels in three months), the flux decreases by about 15%, which subsequently results in an increase of the NSR by about 8% (see Fig. 2 – bottom). The NSR increase obviously has an impact on the science objectives of the mission,

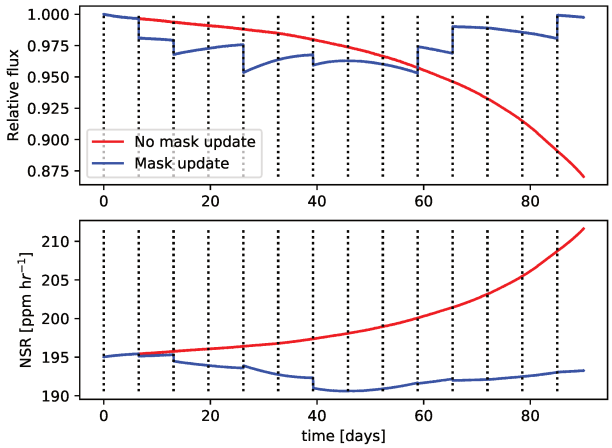


Fig. 2. *Top*: examples of light-curves generated from a simulated time-series of CCD imagettes. The red curve corresponds to the light-curve generated with a fixed binary mask while the blue one includes a series of mask updates. The dotted vertical lines identify the times at which the masks were updated. *Bottom*: corresponding time variations of the NSR.

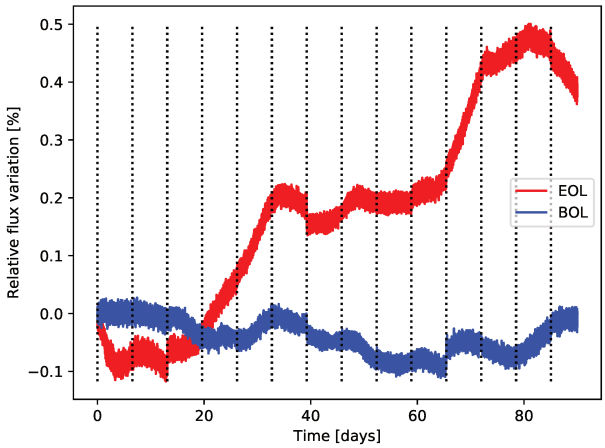


Fig. 3. Examples of single-camera light-curves generated from a simulated time-series of CCD imagettes and after correcting it for the long-term drift of the stellar position as explained in Sect. 4.6. We represent the relative flux variation in %. The red (reps. blue) curve corresponds to EOL (resp. BOL) observation conditions. The dotted vertical lines identify the times at which the masks were updated.

in particular on the planet detection rate. Indeed, a higher NSR at a given magnitude reduces the number of targets for which the NSR is lower than a given threshold, which subsequently lowers the number of detected planets.

To mitigate the impact of the long-term drift and to maintain the NSR as low as possible, the proposed solution is to update the mask when required during the three-month observation sequences. An example is shown in Fig. 2 (top). We see in this example that the peak-to-peak variation of the flux is maintained within 4% (top panel) while the variation of the NSR is maintained within less than 2% (lower panel). Therefore, the mask updates always guarantee that one reaches the best possible NSR. Furthermore, it is also found that the mask updates partially mitigate the impact of CTI.

It is interesting to note that some mask updates simultaneously reduce the flux and the NSR. This is because the NSR does not scale linearly with the flux. Indeed, due to the presence of the readout noise and the complex shape of the PSF, two masks collecting the same amount of flux can have a different number of pixels and hence different contributions of the readout noise.

Obviously, each mask update will introduce a discontinuity at a well known instant, which for a given target will be different from one camera to another. For instance in the example shown in Fig. 2, the binary mask has been updated seven times. It is however possible to reduce the number of updates by increasing the threshold above which a variation of the NSR since the last update must trigger a mask update. Furthermore, as explained below, the discontinuities induced by each update as well as the long-term flux variations induced by the long-term drifts can be efficiently corrected a posteriori on-ground.

4.6. Light-curve correction

Knowing the PSF at any position across the field of view and the time displacements of a given target, it is possible – given its associated aperture mask – to reconstruct a synthetic light-curve, which exactly mimics the time variation of the star flux induced only by the long-term drift over the CCD plane as well as the discontinuity induced by each mask update. The light-curve correction then consists in building such a synthetic light-curve assuming that the star has an unit intensity, then dividing the real stellar light-curve by the synthetic one. The quality of this correction intrinsically depends on our ability to construct the PSF and to derive the stellar displacement in time. As explained in detail in Sect. 4.7, the stellar PSFs will be reconstructed during the in-flight calibration phases on the basis of a microscanning sequence coupled with a dedicated inversion method.

Concerning the stellar displacements, the fast camera will provide information about the short-term variations of the satellite attitude (i.e. satellite jitter) with a high cadence (2.5 s) and a sufficient accuracy. This information will then be directly translated in terms of the short-term variations of the attitude of each given normal camera. Concerning the long-term displacements, centroids of a larger set of targets will be measured using the imagettes registered on-board at a cadence of 25 s. These centroids will be used to derive the attitude of each camera at any instant. The combination of the two sets of information will finally provide both the short-term (i.e. jitter) and long-term time-displacements of any target.

Finally, it has been established that the PRNU is a limiting factor for this correction. However, prior to the launch, the PRNU will be measured with an accuracy better than 0.1% (rms), which is sufficient to leave a negligible level of residual error in the corrected light-curves (Samadi 2015).

Two examples of corrected (individual) light-curves are displayed in Fig. 3 for a $V = 11$ PLATO target. The upper light-curve corresponds to BOL observation conditions while the lower one to the EOL. It is clearly seen that the residual flux variations are larger at the EOL: for that particular target the peak-to-peak flux variations is as high as about 1% at the EOL while it remains within about 0.2% at the BOL. This is explained by the combined effect of the CTI and the star drift. Indeed, as the star moves, the energy distribution in the different pixels vary with time and so does the CTI. This effect is named differential CTI.

The small discontinuities seen in the light-curve occur each times the mask has been updated. These discontinuities are of the order 500 ppm. Hence, they remain small compared to the photon noise, which is about 2000 ppm for this target and for a single camera. It is also worth noting that for a given target observed with several cameras the instants at which the mask updates occur are different between the different cameras. Accordingly,

the systematic errors induced by these updates are uncorrelated between the cameras and accordingly, their impacts on the final light-curves (obtained after averaging several individual light-curves) will be significantly reduced (see Sect. 4.10).

4.7. Point spread function reconstruction

One of the challenges with the PLATO mission is the relatively large size of the camera pixels (approximately 15 arcsec as projected on the field-of-view) compared to the typical size of the point spread functions (PSFs)[7]. Accordingly, raw camera pictures do not provide a sufficient resolution of the PSFs, thus requiring the use of a specific strategy in order to obtain the PSFs with a sub-pixel resolution. In the PLATO mission, the adopted strategy is similar to the one applied to *Kepler* observations (Bryson et al. 2010): a microscanning sessions in which a series of imagettes with sub-pixel displacements are obtained (Green 2011). High resolution PSFs will then be reconstructed by inverting the imagettes along with a precise knowledge of the displacements. Such PSFs will be obtained for a number of reference stars across the field-of-view. The PSF at any position will then be obtained via interpolation using the reference PSFs. The resultant PSFs will subsequently be used in correcting the light-curves sent down by PLATO as explained in Sect. 4.6.

4.7.1. Microscanning sessions

The microscanning sessions will typically last for 3 h and lead to a series of 430 imagettes composed of 6×6 pixels encompassing the target stars. The telescope will be pointing in a slightly different direction for each imagette resulting in small sub-pixel displacements of the target stars. A continuous microscanning strategy has been opted for, that is the position will be changing continuously throughout the manoeuvre rather than stopping for each imagette and then starting again (Ouazzani et al. 2015). The displacements do not need to fulfil stringent criteria in order to be suitable for the inversions, but only to form a path which roughly covers a pixel uniformly (for more details see Reese 2018a). Accordingly, this path has primarily been determined based on technological constraints. However, a precise knowledge of the displacements is essential for carrying out successful inversions. Various tests have shown that the fast cameras are able to obtain this information from the centre-of-brightness of reference stars.

The displacements will form an Archimedean spiral such that the distance, D, between consecutive images is approximately constant, and the distance between consecutive spiral arms is $D\sqrt{3}/2$, thus leading to the formation of near-equilateral triangles depending on the relative positions of imagettes on consecutive arms. Furthermore, the spiral needs to approximately cover 1 pixel. The combination of these constraints leads to a spiral like the one illustrated in Fig. 4.

4.7.2. PSF inversions

In order to carry out the inversions, it is necessary to discretise the PSF by expressing it as a sum of basis functions:

$$f(x, y) = \sum_i a_i \phi_i(x, y), \tag{31}$$

[7] We note that the point spread function changes significantly across the relatively large field-of-view (about 20° in radius).

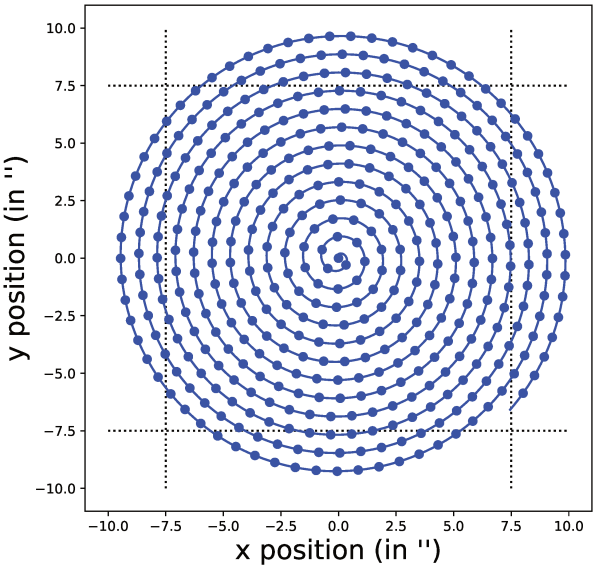


Fig. 4. Archimedean spiral used for the microscanning strategy. The dotted lines correspond to the CCD pixel borders.

where the a_i are unknown coefficients which will be determined via the inversion, and the ϕ_i basis functions. Typical choices of basis functions include sub-pixel indicator functions, or Cartesian products of cubic B-splines. The typical resolution used for these basis functions is 1/20th of a pixel (along both the x and y directions), given the number of imagettes from the microscanning session. This high resolution representation of the PSF then needs to be integrated over the pixels of the imagettes. Equating the resultant integrals with the observed intensities in these pixels leads to the following equation

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{32}$$

where $\boldsymbol{b}$ is a vector composed of the observed intensities from the imagettes, $\boldsymbol{x}$ a vector composed of the coefficients a_i, and A the discretisation matrix. The inverse problem is then to extract the high resolution PSF, x, knowing A and b. Given that the number of unknowns does not necessarily equal the number of observables, this problem needs to be inverted in a least-squares sense. Furthermore, some form of regularisation needs to be included in order to obtain well-behaved solutions. Finally, the resultant high-resolution PSF needs to remain positive. A sufficient (though not necessary) condition for this is to impose that the coefficients a_i are positive, provided the basis functions ϕ_i remain positive.

Two inversion techniques have been used for solving Eq. (32). The first is an iterative approach called the Multiplicative Algebraic Reconstruction Technique (MART; Censor 1981; Green & Wyatt 2006), which starts from a positive smooth solution and iteratively corrects it using one constraint at a time. Given that the corrections are applied in a multiplicative manner, the solution remains positive. The number of iterations is then used to control the degree of regularisation. The second approach is a regularised least-squares approach with a positivity constraint on the coefficients. The regularisation term consists of a 2D Laplacian multiplied by a weight function which leads to a higher amount of regularisation in the wings of the PSF. Accordingly, cubic B-splines are used with this approach given that these are continuously twice-differentiable. This term is then multiplied by a tunable regularisation parameter. As shown in Reese (2018b), this second approach leads to better results in

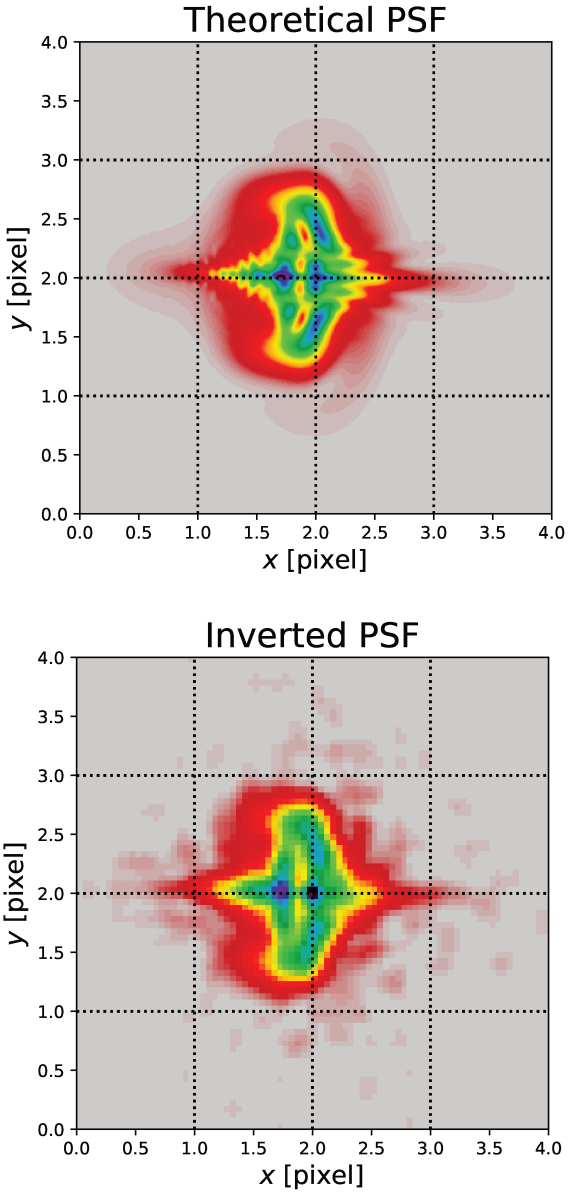


Fig. 5. *Top*: original PSF. *Bottom*: PSF obtained by inversion on the basis of the microscanning technique (see Sect. 4.7). The dotted lines correspond to the CCD pixel borders.

most cases (see an example in Fig. 5 bottom panel) and is accordingly the preferred approach.

4.8. Analysis of systematic errors in terms of PSD

We compute the PSD associated with each corrected light-curve, both for the BOL and EOL data sets. Two examples are shown in Fig. 6 at the BOL and the EOL for a star located at a given position in the field of view and at the sub-pixel position P_0 (pixel corner). These PSDs are compared with the PLATO requirements in terms of allowed systematic errors at $V = 11$. At the EOL, the requirements are marginally exceeded in the frequency range [10 μHz–100 μHz]. This is a consequence of the presence of CTI. We stress that we expect to be able to correct for the CTI. However, this correction is not yet fully modelled and hence cannot yet be reliably quantified. Accordingly, the predictions for the EOL have to be considered conservative at this moment.

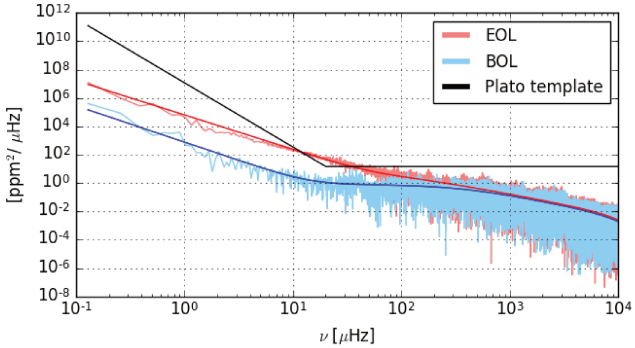


Fig. 6. PSD of the residual light-curve obtained with two three-month PIS simulations, one representative of the BOL (cyan) and the second of the EOL (pink). The results shown here correspond to a star of magnitude 11. The solid coloured line represents the fitted model defined in Eq. (33). The solid black line represents the PLATO requirements in terms of systematic errors translated for a single camera by assuming that they are uncorrelated between the different camera (see text).

We find that the PSD of the residual light-curve can satisfactory be fitted with a function of the form:

$$\mathcal{I}(\nu) = H_1 \left(\frac{\nu_1}{\nu}\right)^{\alpha_1} + \frac{H_2}{1 + (2\pi\tau_2\nu)^{\alpha_2}}, \tag{33}$$

where H_1, α_1, H_2, τ_2 and α_2 are the fitted parameters, and $\nu_1 = 1/T_1$ with $T_1 = 90$ days (three months). For convenience, we further define σ^2 as the variance of the residual light-curve, which is also the integral $\mathcal{I}(\nu)$. The quantity σ corresponds to the amplitude of the systematic errors and is related to the other parameters according to the relation

$$\sigma^2 = \frac{H_1\nu_1}{\alpha_1 - 1} + \frac{H_2}{2\,\tau_2\,\sin(\pi/\alpha_2)}. \tag{34}$$

We fit each residual light-curve with the function given by Eq. (33). Depending on the parameters, we find that the parameter values significantly vary with the sub-pixel positions. This is not surprising since about 90% of the star's intensity is concentrated in a square of 2.2×2.2 pixels (an example of such a PSF is displayed in Fig. 5). As a result, a change of the sub-pixel position of the star's centroid induces important changes in the charge distribution. In contrast, at fixed sub-pixel positions, changes of the parameter values with the stellar field of view are in general weaker. Finally, the parameters controlling the amplitudes of the systematic errors (namely H_1 and H_2) are found to strongly vary with the stellar magnitude.

Figure 7 highlights the impact of the star magnitude. Indeed, the 90th percentile of the quantity σ [in ppm] is displayed as a function of the star magnitude, both for BOL and EOL conditions. In general, the residual systematic errors increase with increasing star magnitude. As expected, the systematic errors at EOL are systematically much higher than at BOL by about a factor ten at magnitude 8.5 and down to about a factor four at magnitude 13. However, they hardly exceed 0.5% (rms) in the magnitude range considered here.

4.9. Modelling the systematic errors in the time domain

The systematic errors were analysed in the previous section in terms of PSD because this is the most convenient way to compare directly with the mission requirements in terms of allowed systematic errors. Indeed, the latter are specified in terms of the

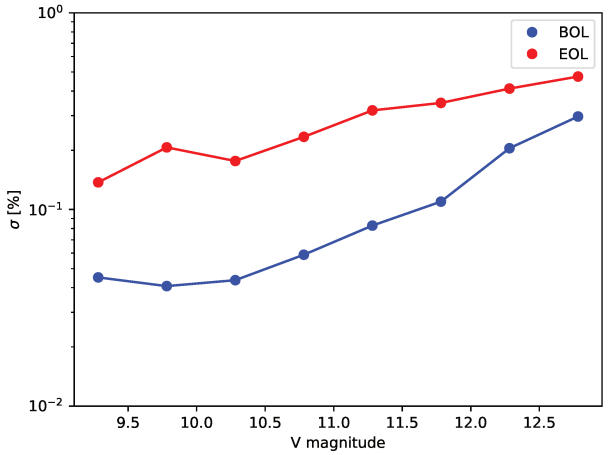


Fig. 7. Amplitude of the systematic errors (σ) as a function of the stellar magnitude. This quantity is computed according to Eq. (34). At each stellar magnitude, only the 90th percentile is displayed.

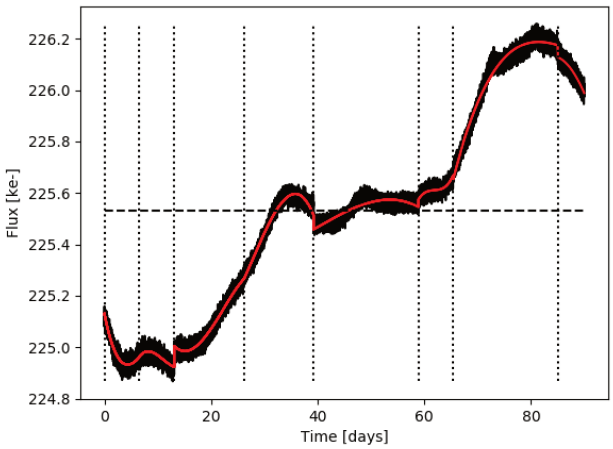


Fig. 8. Example of a generated instrumental light-curve (single camera) fitted with the piecewise polynomial decomposition of Eq. (35). The black line represents the generated light-curve and the red one the result of the fit.

PSD. However, modelling the systematic errors this way has the obvious consequence of destroying the phase of the instrumental or operational perturbations (e.g. the discontinuities induced by the mask updates or other effects). While this is in principle not a problem when the stellar signal is analysed in terms of the PSD (e.g. as this is typically done for the granulation or the solar-like oscillations), this can be misleading for the analysis taking place in the time domain, as for instance the detection and the characterization of planetary transits. To overcome this, we decide instead to model the systematic errors in the time domain.

Due to the quasi-regular mask updates, the residual light-curve is piecewise continuous (each piece corresponding to an interval of time where the aperture mask is unchanged). We find that each piece can be well reproduced by a third order polynomial. Accordingly, we decompose each generated instrument light-curve as follows

$$s(t) = \overline{s} \sum_{i=1}^{N} \Pi\left(\frac{t-t_i}{d_i}\right)\left(1 + p_{3,i} + p_{2,i}x + p_{1,i}x^2 + p_{0,i}x^3\right), \quad (35)$$

where $\overline{s}$ is the light-curve time-average, N the number of masks used for a given imagette time-series, i the mask index, t_i the time the mask is first applied or updated, d_i the time during which it is maintained, $x \equiv (t - t_i)/\tau_0$, τ_0 a time constant (set arbitrarily to 90 days), $p_{j,i}$ the polynomial coefficients associated with the mask i, and finally $\Pi(x)$ a function defined as

$$\Pi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1 \\ 0 & \text{if } x < 0 \text{ or } x \ge 1. \end{cases}$$

While the coefficient $p_{3,i}$ informs us about the amplitude of the discontinuity induced by a given mask i, the three other coefficients ($p_{0,i}$, $p_{1,i}$ and $p_{2,i}$) inform us about the long-term variations of the instrument residuals obtained with that mask.

Each of the generated instrument light-curve is fitted by the model given by Eq. (35). An example of such a fit is given in Fig. 8. In most cases, this polynomial model reproduces very well the main characteristics of the systematic errors, in particular the jumps induced by the mask updates as well as the long-term variations induced by the long-term star drifts.

4.10. Implementation into PSLS

The model for the systematic errors presented in the previous section is implemented into PSLS as follows: we have at our disposal a set of p coefficients for each stellar magnitude, focal plane position (i.e. PSF), and sub-pixel position. We first identify the positions (focal plane and sub-pixel) corresponding to the magnitude the closest to that of the star we want to simulate. For each position, the number of p coefficients depends on the number of masks used at that position. Then, each individual light-curve simulated by PSLS is divided into quarters. For each quarter, we randomly select the set of coefficients p among the ensemble previously selected. We proceed in the same way for each quarter and for each individual camera. By proceeding this way, we simulate the fact that each star will have different PSFs and sub-pixel positions in the different cameras and that these PSFs and sub-pixel positions will change after the rotations of the spacecraft by 90° every 3 months. An example of such simulated light-curves is shown in Fig. 9.

The above example also illustrates the benefit of averaging the light-curves over several cameras. Indeed as shown in Fig. 9 averaging over, for example 24 cameras, substantially reduces the residual errors because the systematic errors are not always in phase and the mask updates do not always occur at exactly the same times. However, the figure also highlights some degree of correlation between the individual light-curves. Indeed, it can clearly be seen that some light-curves are close to being in phase. These correlations are expected as explained and discussed in Sect. 7.2.

5. Other signal components

5.1. Stellar granulation

The granulation background is simulated by assuming two pseudo-Lorentzian functions

$$G(\nu) = \sum_{i=1,2} \frac{h_i}{1 + (2\pi\tau_i\nu)^{\beta_i}}, \quad (36)$$

where h_i is the height, τ_i the characteristic time-scale, and β_i the slope of the Lorentzian function. The values of h_i and τ_i are determined from the scaling relations established by

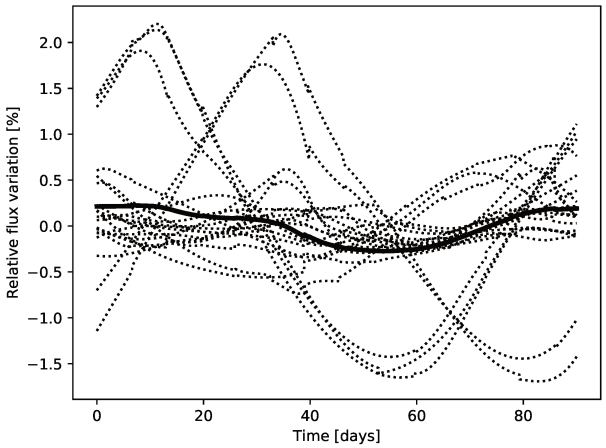


Fig. 9. Simulated instrument residual light-curves (systematic errors) over 90 days for a star of mangitude $V = 11$ and for the EOL conditions. The light-curves are plotted in terms of relative variations and were generated using Eq. (35) and as explained in Sect. (4.10). Each dotted line corresponds to an individual light-curve (here 24 in total) while the thick solid line corresponds to the light-curve obtained by averaging the 24 simulated light-curves.

Kallinger et al. (2014) with *Kepler* observations of red giants, sub-giants and main-sequence stars. These scaling relations are a function of peak frequency $\nu_{\max}$ of the oscillations and the stellar mass M. Following Kallinger et al. (2014), the values of the two slopes (β_1 and β_2) are both fixed to four.

5.2. Stellar activity

The stellar activity signal is simulated assuming a Lorentzian function

$$\mathcal{A}(\nu) = \frac{2\,\sigma_{\rm A}^2\,\tau_{\rm A}}{1 + (2\pi\tau_{\rm A}\nu)^2}, \tag{37}$$

where $\sigma_{\rm A}$ is the amplitude and $\tau_{\rm A}$ is the characteristic time-scale of the activity component. Both parameters have to be specified by the user (but see the discussion in Sect. 7).

5.3. Planetary transit

Planetary transit light-curve are simulated on the basis of Mandel & Agol (2002)'s formulation and using the Python implementation by Ian Crossfield at UCLA[8]. This model allows us to specify several parameters controlling the characteristics of the transit light-curve. Among them, PSLS allows us to specify the planet radius, the orbital period, the semi-major axis and finally the orbital angle. We have adopted a quadratic limb-darkening law (cf. Sect. 4 of Mandel & Agol 2002) and assumed default values for the corresponding two coefficients (namely $\gamma_1 = 0.25$ and $\gamma_2 = 0.75$). However, these coefficients can be set by the user.

6. Simulated stellar oscillation spectra

As a preliminary remark, we stress that the goal of the simulator is not to provide state-of-the-art modelling of a given target but rather to be able to mimic the main characteristics of objects of interest. Accordingly, we did not carry out a quantitative and extensive comparison between outputs of our simulator and light-curves (or equivalently PSD) obtained from high-precision photometric observations, from space missions such as CoRoT and *Kepler*. However, to illustrate the quality and the relevance of the simulated light-curves, we performed a qualitative comparison with *Kepler* observations. Three *Kepler* targets were selected according to the quality of the data and their evolutionary status: a main sequence star (16 Cyg B–KIC 12069449), a sub giant (KIC 12508433) and a giant on the Red Giant Branch (KIC 009882316). For each of them, a simulation was generated with stellar parameters and models as close as possible to the corresponding target.

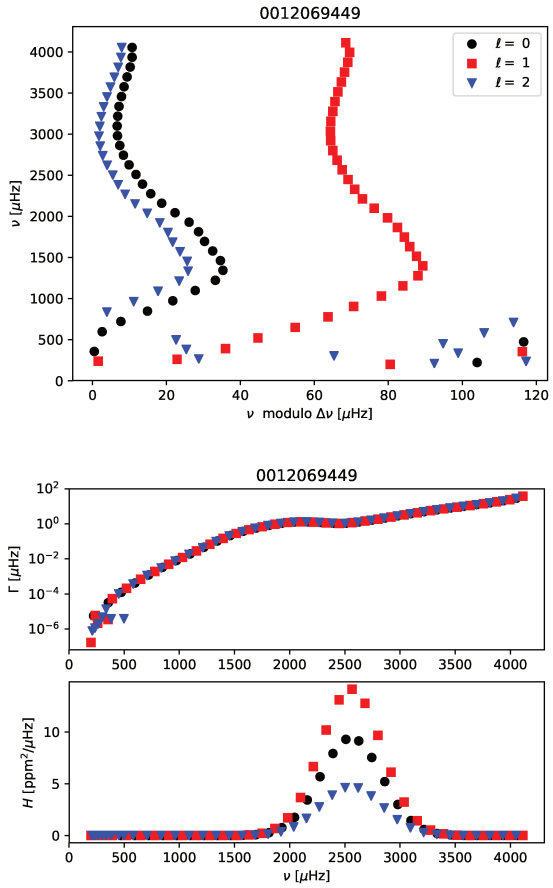


Fig. 10. *Top panel*: Échelle diagram corresponding to the frequencies used as input for the simulations made for 16 Cyg B (KIC 12069449). A mean large separation of $\Delta\nu = 118.9\,\mu$Hz was assumed when plotting the échelle diagram. *Bottom panel*: corresponding mode linewidths (*top*) and mode heights (*bottom*).

6.1. Main sequence star

16 Cyg B (KIC 12069449) belongs to the *Kepler* legacy sample (Lund et al. 2017a,b). As input for PSLS, we considered a set of theoretical adiabatic mode frequencies computed with ADIPLS, using one of the stellar models considered in Silva Aguirre et al. (2017). The effective temperature and surface gravity were adjusted in accordance with the 1D stellar model while the seismic indices $\nu_{\max}$ and $\Delta\nu$ were taken from Lund et al. (2017a).

We generated an initial light-curve assuming a $V = 10.0$ PLATO target observed with 24 cameras in EOL conditions and

[8] http://www.astro.ucla.edu/~ianc/

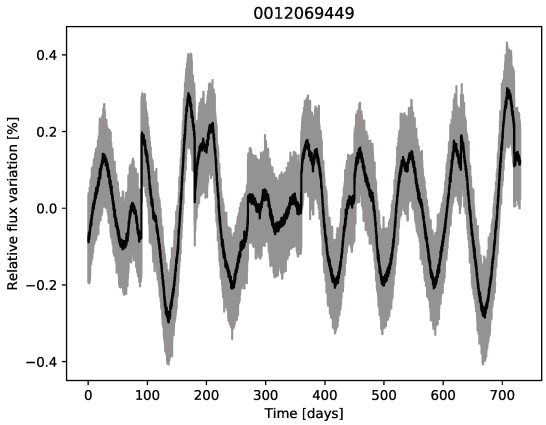


Fig. 11. Simulated light-curve for the main sequence star 16 Cyg B (KIC 12069449) with simulation parameters representative of a $V = 10.0$ PLATO target observed in EOL conditions with 24 cameras (see the details given in Sect. 6.1). The grey curve corresponds to the raw light-curve while the black ones is the light-curve averaged over one hour.

for a duration of 2 years. The choice of magnitude is motivated by the fact that we expect to derive stellar ages with the required accuracy (10%) with 24 cameras up to the magnitude $V = 10.0$ (Goupil 2017)[9]. The two parameters controlling the activity component have been adjusted so that it matches that of the activity component seen in the 16 Cyg B Kepler light-curve. The corresponding PSLS configuration file is given in Appendix A. The mode frequencies, line-widths, and heights used as input for the simulation are displayed in Fig. 10. The modes for which the frequencies significantly depart from the general trends are mixed modes. However, they have such low amplitudes that in practice they are not at all detectable.

The corresponding simulated light-curve is displayed in Fig. 11 while the corresponding PSD is plotted in Fig. 12, where we have depicted the various contributions to the PSD. As can been seen, the systematic errors start to dominate over the stellar signal only below ~20 μHz. On the other hand, they remain negligible in the frequency domain where the solar-like oscillations and stellar granulation signal lie. As expected at that magnitude, the random noise (white noise) dominates the signal in this domain. Nevertheless, the presence of the oscillations in the PSD is clearly discerned when zooming and smoothing the PSD in this frequency domain (see bottom panel of Fig. 12).

The simulated PSD cannot be directly compared with *Kepler* observations for that star because PLATO and *Kepler* have different characteristics and furthermore 16 Cyg B is so bright that its image on the CCD is saturated. Therefore, to perform a comparison we adjusted the white noise level (equivalently the NSR value) so that it matches the level of the white noise seen at high frequency in the *Kepler* light-curve. We compare in Fig. 13 the simulated PSD with the *Kepler* observations. Qualitatively, we note a fair agreement between the simulation and the observations. The figure however highlights some differences, in particular in terms of mode heights and the width of the oscillations envelope. As the characteristics of the oscillations are obtained through scaling relations, we do not except the match to be perfect, and in any case this is not the ultimate goal of the simulator.

[9] This threshold is obviously lower for stars observed with less than 24 cameras.

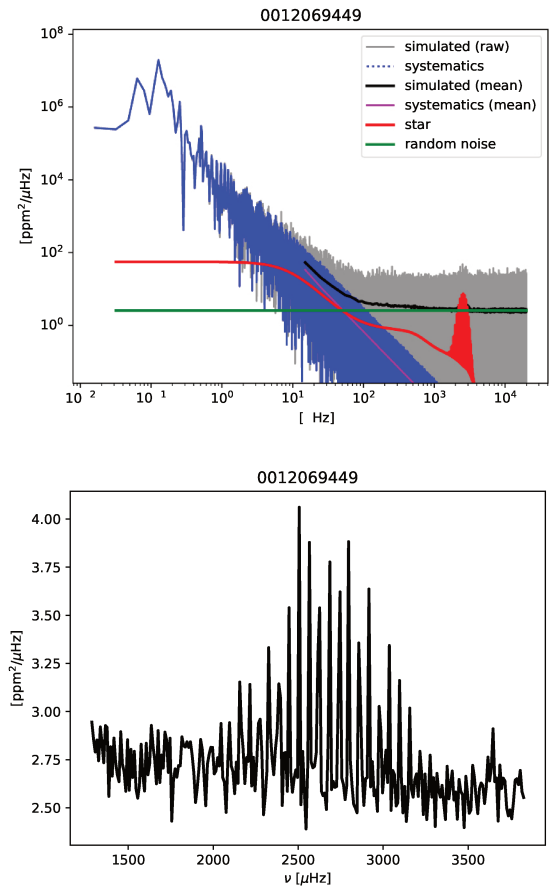


Fig. 12. PSD of the simulated light-curve of 16 Cyg B (KIC 12069449) shown in Fig. 11. *Top*: full PSD. The grey curve represents the raw PSD (i.e. un-smoothed PSD) while the black line corresponds to the PSD obtained after applying a running average over a width of 10 μHz. The coloured lines represent the various contributions to the signal (see the associated legend). *Bottom*: zoom in the frequency domain where solar-like oscillations are detected. Only the smoothed PSD is shown.

6.2. Sub-giant star

The sub-giant star KIC 12508433 observed by *Kepler* is among the sub-giant stars studied in detail by Deheuvels et al. (2014). As an input for PSLS, we use the same stellar parameters as in this study as well as the set of theoretical mode frequencies that best fits the seismic constraints. As for 16 Cyg B (KIC 12069449), we adjusted the white noise level so that it matches the level of the white noise seen at high frequency in the corresponding *Kepler* light-curve. The comparison between the simulated PSD and the one computed from the *Kepler* light-curve is shown in Fig. 14. Here also, we have a good qualitative agreement between both PSDs. Nonetheless, some mismatch is visible by eye, especially concerning the mode heights and the width of the oscillation envelope.

6.3. Red-giant star

The red giant KIC 9882316 has been studied extensively since the *Kepler* era. Precise measurements of its seismic indexes ($\Delta\nu$, $\nu_{\max}$, $\Delta\Pi$ and q) have been published for example in Mosser et al. (2015). We generated for this red giant a simulated light-curve on the basis of the UP method (see Sect. 3.2). The latter requires specifying the seismic indexes as well as the effective

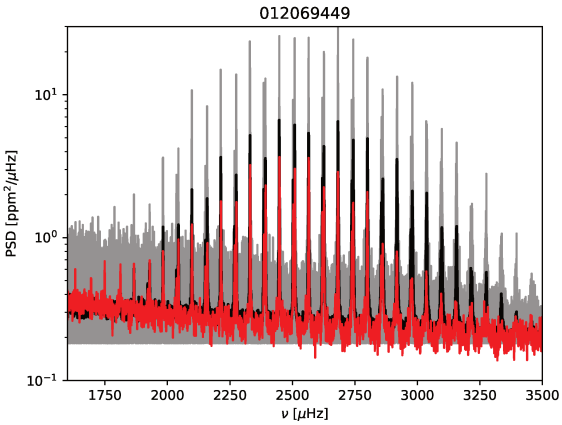


Fig. 13. Comparison of a simulated PSD and the one obtained from *Kepler* observations of the main sequence star 16 Cyg B (KIC 12069449). The grey and black lines have the same meaning as in Fig. 12. The red curve corresponds to the smoothed PSD obtained from the observations.

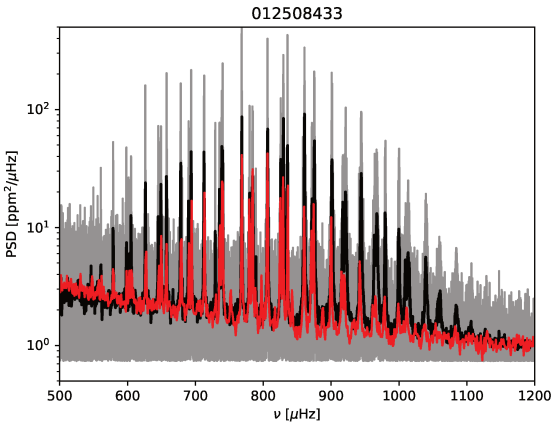


Fig. 14. Same as Fig. 13 for the sub-giant star KIC 12508433 observed with *Kepler*.

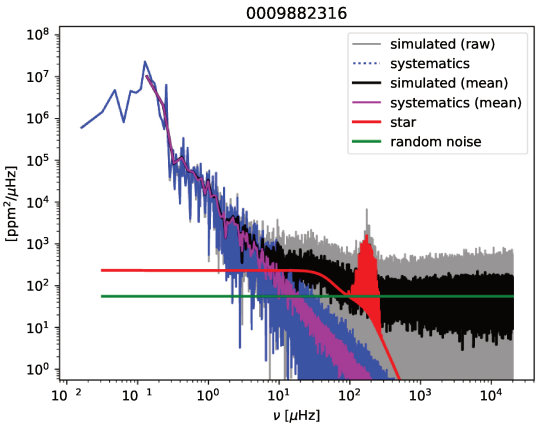


Fig. 15. PSD of the simulated light-curve for the *Kepler* red giant KIC 9882316 seen as a PLATO target of $V = 12.5$ in EOL conditions with 24 cameras. The curves have the same meaning as the top panel of Fig. 12.

temperature ; all these parameters are taken from Mosser et al. (2015). To illustrate the quality of the light-curve expected for red giants with PLATO, we first perform a simulation for a $V = 12.5$ PLATO target observed with 24 cameras in EOL conditions for a duration of 2 years. The corresponding PSLS configuration file is given in Appendix A. We note that solar-like oscillations are also expected to be detectable in fainter red giants, but we limit ourselves to this magnitude because the systematic errors were not quantified for fainter stars. The PSD of the simulated light-curve is displayed in Fig. 15, where we have also plotted the different contributions to the signal. As can been seen, the systematic errors remain negligible compared to the solar-like oscillations and stellar granulation. On the other hand, they dominate below $\nu \sim 20\,\mu$Hz.

Finally, we compare the predictions made by PSLS with *Kepler* observations. We again adjust the white noise level to match the *Kepler* observations for that target and considered a simulation duration of 4 years. The comparison is shown in Fig. 16. The agreement between the simulation and the *Kepler* observations is rather good. In particular, we see that the mixed-mode frequencies and heights are quite well reproduced.

7. Discussion

We discuss in this section the limitations of the current approach and possible future improvements.

7.1. Instrument model

As far as the modelling of the instrument is concerned, there is still an important effect missing in the image simulator (PIS), which is the Brighter Fatter Effect (BFE hereafter). Indeed, there are several pieces of evidence showing that spot images using CCDs do not exactly scale with the spot intensity: bright spots tend to be broader than faint ones, using the same illumination pattern (see Guyonnet et al. 2015, and references therein). The BFE is fundamentally due to the self-electrostatic interaction between charges in different pixels. This broadening, which mainly affects bright targets, would not be a problem as long as these interactions are stable in time. However this cannot be the case since the long-term drift of the stellar position changes the charge distribution in the different pixels. Analytical models (e.g.. Guyonnet et al. 2015) can be easily implemented into PIS, and that can subsequently be used to correct both the CCD imagettes and the light-curves generated on-board. Such analytical models involve several free coefficients that can be calibrated on-ground with the test bench dedicated to the calibration of the flight PLATO CCDs. The BFE is not expected to evolve with time so that the parameters obtained with the on-ground calibration can be used throughout the mission. The calibration procedure for the PLATO CCD is not yet established but can in principle follow the one proposed in Guyonnet et al. (2015). As soon as we have at our disposal calibrated values of the BFE coefficients, it will be possible to update our simulations and derive new prescriptions to account for this additional source of systematic errors.

Charge diffusion within the CCD was neglected in this work since we still lack reliable estimates of its amplitude in the case of the PLATO CCD. However, charge diffusion is expected to a have non-negligible impact on the performance since it enlarges somewhat the width of the PSF and leads to the suppression of the small-scale structures of the optical PSF. It has been shown for example by Lauer (1999) that in standard rear-illuminated CCDs this phenomena can be modelled by performing the convolution of the optical PSF with a 2D Gaussian kernel with a given width, which strongly depends on the wavelength and type of CCD device. To what extent this model is applicable to the

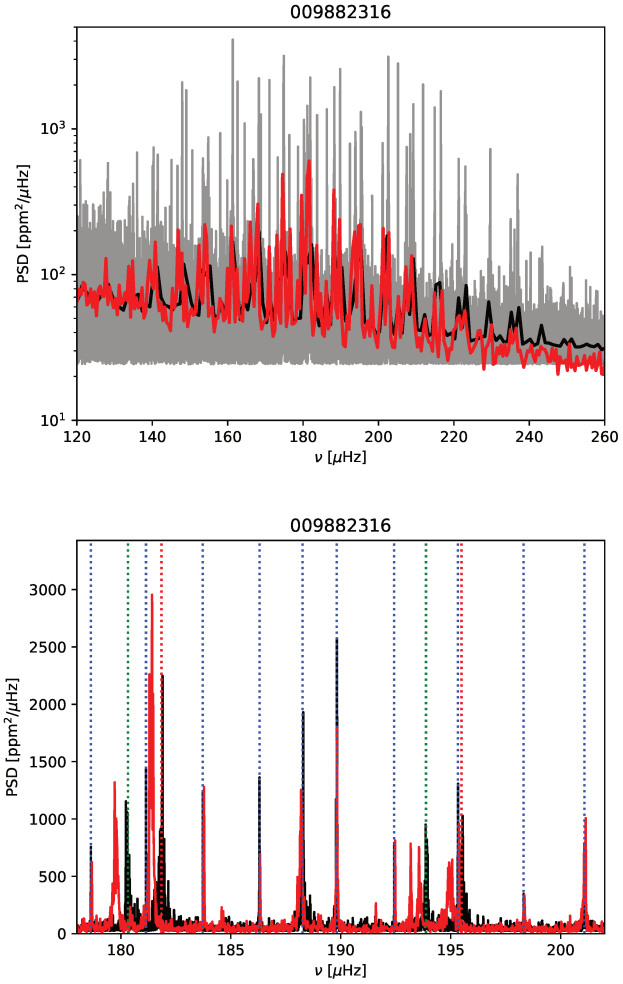


Fig. 16. Same as Fig. 13 for the red giant star KIC 9882316 observed with *Kepler*. *Top*: entire oscillation spectrum. *Bottom*: zoom around the maximum peak frequency. Here we did not apply a running average over the PSD. The dotted red/blue/green vertical lines represent the frequency locations of the radial/dipole/quadrupole modes respectively.

PLATO CCD and what typical width to use for representative PLATO targets are still open questions. Fortunately, tests on representative PLATO CCD are currently taking place at the ESTEC and will provide feedback on this issue that should allow us to improve the present performance assessment.

Besides the above mentioned effects taking place at the detector level, PLATO will be subject to many others perturbations that are not yet taken into account in the present simulator. Among them, we can in particular mention thermal trends after the rotations of the spacecraft by 90° every 3 months, the regular thermal perturbations induced by the daily downlinks, the momentum wheels de-saturations, and finally the residual outliers that would not have been detected by the outlier detection algorithms. All these perturbations are not yet well characterized but will be better known in the future, in particular with the deeper involvement of the prime contractor of the platform in the project and progress in the definition of the data processing pipeline.

7.2. Single to multiple instrument simulations

Strong correlations between the light-curves coming from different cameras are expected. For instance, stellar drifts along the focal plane are expected to be strongly correlated between the cameras. Although for each given target, the associated PSF and aperture mask can differ between cameras, variations in the stellar flux induced by stellar drifts will present some degree of correlation, which only pixel-level simulations made for several cameras can quantify.

PSLS generates the instrument systematic errors of each individual light-curve individually. However, it uses the model parameters derived from pixel-level simulations made for a single camera only (see Sect. 4.9). Each observed target will have different PSFs and sub-pixel positions in the different cameras with which it is observed. Accordingly, to simulate this diversity, PSLS randomly selects the model parameters derived at various positions (for a given star magnitude).

However, this approach is to some extent conservative. Indeed we use the systematic errors evaluated for the same camera. Each individual light-curve will be corrected a posteriori on-ground on the basis of auxiliary data (such as the PSF) obtained by calibrating independently each individual camera. There are good reasons to believe that the systematic errors will be different from one camera to another. Indeed, the systematic errors made on each individual calibration are expected to be different because the cameras are not exactly identical. Indeed, the cameras do not have the exact same alignments of the CCD over the focal plane, same focal plane flatness, same PRNU, same optical manufacturing and alignment errors, etc. However, to confirm this it is required to simulate a statistically sufficient number of cameras with slightly different setups (results from a limited attempt can be found in Deru et al. 2017).

7.3. Stellar contamination

While the presence of contaminant stars was taken into account in our calculation of the NSR as a function of the stellar magnitude (see Sect. 4.4), this is not the case for the systematic errors. It is, in principle, possible to take into account the contaminant stars in the reconstruction of the stellar PSF and the generation of the three-month imagette time-series (Reese & Marchiori 2018). However, this is numerically challenging since due to the high diversity in terms of configuration, statistically reliable quantification of the impact of stellar contamination would require a much larger sample of simulations. Accordingly, we plan to perform, in the near future, simulations on the basis of a sufficiently large stellar field extracted from the *Gaia* DR2 catalogue.

7.4. Stellar activity and rotational modulations

Although the presence of stellar activity has little impact on solar-like oscillations, its presence is critical for the detection of planetary transits. At the present time, the parameters of the activity component still need to be specified by the user. Hence, our objective is to implement into the simulator some empirical descriptions of the magnetic activity sufficiently realistic to be representative of solar-like pulsators in the context of PLATO. To this end, we plan to analyse a large set of *Kepler* targets and derive from their spectra, in a similar way as for example in de Assis Peralta et al. (2018), two main characteristic parameters of the activity, namely the characteristic time-scale and the amplitude associated with the activity component. Once these parameters are derived for a large sample of stars, we believe it will be possible to derive some relations between these parameters and some stellar parameters, such as the surface rotation period and the Rossby number which is the ratio of the rotation period and convective turnover time. Indeed, the differential

rotation existing at the interface between the convective envelope and the internal radiative zone is believed to be at the origin of the stellar dynamo while convection is believed to be responsible for the diffusion of the magnetic field in the convective zone (see e.g. Montesinos et al. 2001, and references therein).

Finally, one other missing activity-related signal is the rotational modulation due to the presence of rather large spots on the stellar surface. It is hence planned to implement in the near future some of the existing spot models (for a review on this problem see Lanza 2016). However, one difficulty is to have at our disposal representative prescriptions for the model parameters, for instance typically the number of spots, their sizes and their lifetimes. To our knowledge, such prescriptions do not exist yet. Therefore, as a starting point we plan to let the user chose these parameters.

8. Conclusion

We have presented here a light-curve simulator, named the PLATO Solar-like Light-curve Simulator (PSLS), that aims at simulating, as realistically as possible, solar-like oscillations together with other stellar signals (granulation, activity, planetary transits) representative of stars showing such pulsations. One of the specificities of this tool is its ability to account for instrumental and observational sources of errors that are representative of ESA's PLATO mission. The latter were modelled on the basis of the Plato Image Simulator (PIS), which simulates the signal at the CCD pixel level. At the Beginning Of Life, we show that the systematic errors are always compliant with the specifications, whereas at the End Of Life they marginally exceed the specifications between 10 μHz and 100 μHz approximately (see Fig. 6) as a result of Charge Transfer Inefficiency (CTI). However, some mitigation options for the CTI are currently under study (e.g. charge injection, increasing the camera shielding). Although the procedure is not yet fully established, existing correction algorithms can be implemented in the context of PLATO (e.g. Short et al. 2013; Massey et al. 2014).

The PIS code is however not adapted to generating in a massive way simulated long-duration light-curves (e.g. up to two years in the case of PLATO). This is why a parametric description of the systematic errors expected in the time domain has been derived from the PIS simulations. This model reproduces both the residual long-term flux variations due to the instrument as well as the jumps induced by the mask-updates for those of the targets (the large majority of the targets of sample P5) for which photometry is extracted on-board. Implemented into PSLS, this parametric model enables us to mimic in a realistic and efficient way the instrument systematic errors representative of the PLATO multi-telescope concept. Hence, with the inclusion of stellar signal components that are the most representative for the PLATO targets together with a realistic description of the instrument response function, this light-curve simulator becomes an indispensable tool for the preparation of the mission. Its adaptation to other future space missions is in principle possible, provided that some analytical prescriptions for the instrumental and environmental sources of errors representative of the mission are available.

Light-curves simulated with PSLS allow us to conclude that the systematic errors remain negligible above about 100 μHz and only start to dominate over the stellar signal below ~20 μHz. Accordingly, they should not impact the core science objectives of PLATO. One the other hand, they can potentially impact the analysis of the signal below $\nu \sim 20\,\mu$Hz. In both cases, however, firm conclusions deserve dedicated studies, which are beyond the scope of the present work. It must further be made clear that the level of systematic errors predicted by the present modelling is, strictly speaking, only representative for those targets for which the photometry is extracted on-board (i.e. the large majority of the sample P5). For all the other samples, in particular the main sample (P1), the photometry will be extracted on-ground and thus will not suffer from the quasi-regular mask updates. Therefore, a lower level of systematic errors are expected for these samples. Accordingly, the use of PSLS must be considered as a conservative approach for these samples.

This simulator is based on our current knowledge of the instrument and of the current development of the correction pipeline. Although already well advanced, this knowledge will improve in the near future as soon as a first flight model of the camera will be available and fully characterized (around the beginning of 2021). At that time, it will be relatively easy to update our pixel-level simulations and subsequently the parameters used by the model for the systematic errors as well as the Noise-to-Signal Ratio (NSR) table.

Acknowledgements. This work has benefited from financial support by CNES in the framework of its contribution to the PLATO mission. It has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. We acknowledge T. Prod'homme and D. Tiphéne for the useful discussions we had about various CCD effects. We thank the anonymous referee for his/her remarks which enabled us to improve significantly the simulator and the present article.

References

Anderson, E. R., Duvall, T. L., & Jefferies, S. M. 1990, ApJ, 364, 699
Appourchaux, T., Benomar, O., Gruberbauer, M., et al. 2012, A&A, 537, A134
Appourchaux, T., Antia, H. M., Benomar, O., et al. 2014, A&A, 566, A20
Appourchaux, T., Antia, H. M., Benomar, O., et al. 2016, A&A, 595, C2
Baglin, A., Auvergne, M., Barge, P., et al. 2006a, in The CoRoT Mission Pre-Launch Status - Stellar Seismology and Planet Finding, eds. M. Fridlund, A. Baglin, J. Lochard, & L. Conroy, ESA SP, 1306, 33
Baglin, A., Auvergne, M., Boisnard, L., et al. 2006b, in 36th COSPAR Scientific Assembly, COSPAR Meet., 36
Ball, W. H., Chaplin, W. J., Schofield, M., et al. 2018, ApJS, 239, 34
Ballot, J., Gizon, L., Samadi, R., et al. 2011, A&A, 530, A97
Barban, C., Hill, F., & Kras, S. 2004, ApJ, 602, 516
Baudin, F., Samadi, R., Goupil, M.-J., et al. 2005, A&A, 433, 349
Baudin, F., Samadi, R., Appourchaux, T., & Michel, E. 2007, ArXiv e-prints [arXiv:0710.3378]
Belkacem, K. 2012, in SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, eds. S. Boissier, P. de Laverny, N. Nardetto, et al., 173
Belkacem, K., & Samadi, R. 2013, in Lect. Notes Phys., eds. M. Goupil, K. Belkacem, C. Neiner, F. Lignières, & J. J. Green (Berlin: Springer Verlag), 865, 179
Belkacem, K., Goupil, M. J., Dupret, M. A., et al. 2011, A&A, 530, A142
Belkacem, K., Samadi, R., Mosser, B., Goupil, M. J., & Ludwig, H. G. 2013, in Progress in Physics of the Sun and Stars: A New Era in Helio- and Asteroseismology, eds. H. Shibahashi, & A. E. Lynas-Gray, ASP Conf. Ser., 479, 61
Benomar, O., Goupil, M., Belkacem, K., et al. 2018, ApJ, 857, 119
Berthomieu, G., Toutain, T., Gonczi, G., et al. 2001, ESA SP-464: SOHO 10/GONG 2000 Workshop: Helio- and Asteroseismology at the Dawn of the Millennium, 411
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977
Bryson, S. T., Tenenbaum, P., Jenkins, J. M., et al. 2010, ApJ, 713, L97
Censor, Y. 1981, SIAM Rev., 23, 444
Chaplin, W. J., Elsworth, Y., Howe, R., et al. 1997, MNRAS, 287, 51
Chaplin, W. J., Elsworth, Y., Davies, G. R., et al. 2014, MNRAS, 445, 946
Christensen-Dalsgaard, J. 2008, Ap&SS, 316, 113
Corsaro, E., Fröhlich, H.-E., Bonanno, A., et al. 2013, MNRAS, 430, 2313

de Assis Peralta, R., Samadi, R., & Michel, E. 2018, Astron. Nachr., 339, 134
De Ridder, J., Arentoft, T., & Kjeldsen, H. 2006, MNRAS, 365, 595
De Ridder, J., Barban, C., Baudin, F., et al. 2009, Nature, 459, 398
Deheuvels, S., Doğan, G., Goupil, M. J., et al. 2014, A&A, 564, A27
Deru, A., Samadi, R., & Grolleau, E. 2017, Impacts of the Pointing Knowledge on the Long-term Drift Correction, Tech. Rep. PLATO-LESIA-PDC-TN-0029_i1.0, LESIA, Observatoire de Paris
Duvall, Jr., T. L., Jefferies, S. M., Harvey, J. W., Osaki, Y., & Pomerantz, M. A. 1993, ApJ, 410, 829
Dziembowski, W. A. 1971, Acta Astron., 21, 289
Fialho, F., & Auvergne, M. 2006, in The CoRoT Mission Pre-Launch Status - Stellar Seismology and Planet Finding, eds. M. Fridlund, A. Baglin, & J. Lochard, ESA SP, 1306, 289
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gehan, C., Mosser, B., Michel, E., Samadi, R., & Kallinger, T. 2018, A&A, 616, A24
Gizon, L., Ballot, J., Michel, E., et al. 2013, Proc. Natl Acad. Sci., 110, 13267
Gizon, L., & Solanki, S. K. 2003, ApJ, 589, 1009
Goupil, M. 2017, Eur. Phys. J. Web Conf., 160, 01003
Goupil, M. J., Mosser, B., Marques, J. P., et al. 2013, A&A, 549, A75
Green, J. 2011, Microscanning, Tech. Rep. PLATO-GS-TN-231-LESIA_i1.0, LESIA, Observatoire de Paris
Green, J. J., & Wyatt, L. R. 2006, J. Atm. Ocean. Technol., 23, 501
Grosjean, M., Dupret, M.-A., Belkacem, K., et al. 2014, A&A, 572, A11
Guyonnet, A., Astier, P., Antilogus, P., Regnault, N., & Doherty, P. 2015, A&A, 575, A41
Hekker, S., & Christensen-Dalsgaard, J. 2017, A&ARv, 25, 1
Huber, D., Bedding, T. R., Stello, D., et al. 2011, ApJ, 743, 143
Huber, D., Chaplin, W. J., Chontos, A., et al. 2019, AJ, accepted [arXiv:1901.01643]
Kallinger, T., Weiss, W. W., Barban, C., et al. 2010, A&A, 509, A77
Kallinger, T., De Ridder, J., Hekker, S., et al. 2014, A&A, 570, A41
Lanza, A. F. 2016, in Lect. Notes Phys., eds. J. P. Rozelot, & C. Neiner (Berlin: Springer Verlag), 914, 43
Lauer, T. R. 1999, PASP, 111, 1434
Lebreton, Y., Goupil, M. J., & Montalbán, J. 2014a, EAS Pub. Ser., 65, 99
Lebreton, Y., Goupil, M. J., & Montalbán, J. 2014b, EAS Pub. Ser., 65, 177
Lund, M. N., Silva Aguirre, V., Davies, G. R., et al. 2017a, ApJ, 835, 172
Lund, M. N., Silva Aguirre, V., Davies, G. R., et al. 2017b, ApJ, 850, 110
Mandel, K., & Agol, E. 2002, ApJ, 580, L171
Marchiori, V., Samadi, R., Fialho, F., Paproth, C., & Santerne, A. 2019, A&A, submitted
Marcos-Arenal, P., Zima, W., De Ridder, J., et al. 2014, A&A, 566, A92
Massey, R., Schrabback, T., Cordes, O., et al. 2014, MNRAS, 439, 887
Michel, E., Samadi, R., Baudin, F., et al. 2009, A&A, 495, 979
Miglio, A., Chiappini, C., Mosser, B., et al. 2017, Astron. Nachr., 338, 644
Miglio, A., Girardi, L., Rodrigues, T. S., Stello, D., & Chaplin, W. J. 2015, in Asteroseismology of Stellar Populations in the Milky Way, eds. A. Miglio, P. Eggenberger, & L. Girardi , Astrophys. Space Sci. Proc., 39, 11
Montesinos, B., Thomas, J. H., Ventura, P., & Mazzitelli, I. 2001, MNRAS, 326, 877
Mosser, B., & Miglio, A. 2016, IV.2 Pulsating Red Giant Stars, ed. CoRot Team, 197
Mosser, B., Belkacem, K., Goupil, M.-J., et al. 2010, A&A, 517, A22
Mosser, B., Belkacem, K., Goupil, M. J., et al. 2011, A&A, 525, L9
Mosser, B., Elsworth, Y., Hekker, S., et al. 2012a, A&A, 537, A30
Mosser, B., Goupil, M. J., Belkacem, K., et al. 2012b, A&A, 540, A143
Mosser, B., Samadi, R., & Belkacem, K. 2013, in SF2A-2013: Proceedings ofthe Annual meeting of the French Society of Astronomy and Astrophysics, eds. L. Cambresy, F. Martins, E. Nuss, & A. Palacios, 25
Mosser, B., Vrard, M., Belkacem, K., Deheuvels, S., & Goupil, M. J. 2015, A&A, 584, A50
Mosser, B., Pinçon, C., Belkacem, K., Takata, M., & Vrard, M. 2017, A&A, 600, A1
Mosser, B., Michel, E., Samadi, R., et al. 2019, A&A, 622, A76
Nigam, R., Kosovichev, A. G., Scherrer, P. H., & Schou, J. 1998, ApJ, 495, L115
Ouazzani, R.-M., Green, J. J., & Samadi, R. 2015, Eur. Phys. J. Web Conf., 101, 06050
Prod'homme, T., Verhoeve, P., Beaufort, T., et al. 2016, in High Energy, Optical, and Infrared Detectors for Astronomy VII, Proc. SPIE, 9915, 99150U
Rauer, H., Catala, C., Aerts, C., et al. 2014, Exp. Astron., 38, 249
Reese, D. R. 2018a, Testing and Choosing the Microscanning Path, Tech. Rep. PLATO-LESIA-PDC-TN-0045_i1.0, LESIA, Observatoire de Paris
Reese, D. R. 2018b, The Weighted Positive RLS Inversion Method, Tech. Rep. PLATO-LESIA-PDC-TN-0037_i1.1, LESIA, Observatoire de Paris
Reese, D. R., & Marchiori, V. A. 2018, PSF Inversions in the Presence of Contaminating Stars, Tech. Rep. PLATO-LESIA-PDC-TN-0039_i1.0, LESIA, Observatoire de Paris
Reese, D. R., Chaplin, W. J., Davies, G. R., et al. 2016, A&A, 592, A14
Roxburgh, I. W., & Vorontsov, S. V. 1997, MNRAS, 292, L33
Samadi, R. 2015, Photometric Impact of the Long-term Drift of the Star Position and Efficiency of its Correction, Tech. Rep. PLATO-LESIA-PDC-RP-0020_i1.0, LESIA, Observatoire de Paris
Short, A., Crowley, C., de Bruijne, J. H. J., & Prod'homme, T. 2013, MNRAS, 430, 3078
Silva Aguirre, V., Lund, M. N., Antia, H. M., et al. 2017, ApJ, 835, 173
Sonoi, T., Samadi, R., Belkacem, K., et al. 2015, A&A, 583, A112
Stello, D., Huber, D., Bedding, T. R., et al. 2013, ApJ, 765, L41
Unno, W., Osaki, Y., Ando, H., Saio, H., & Shibahashi, H. 1989, Nonradial Oscillations of Stars, 2nd edn. (Tokyo: University of Tokyo Press)
Van Eylen, V., Lund, M. N., Silva Aguirre, V., et al. 2014, ApJ, 782, 14
Van Eylen, V., Dai, F., Mathur, S., et al. 2018, MNRAS, 478, 4866
Vrard, M., Kallinger, T., Mosser, B., et al. 2018, A&A, 616, A94

Appendix A: Configuration file for the main-sequence star KIC 12069449 (16 Cyg B)

```
# PLATO Solar-like Light-curve Simulator (PSLS) configuration file (V 0.8)

# Observations conditions
Observation:
  Duration : 730. #  [days]
  MasterSeed : 1704040900 # Master seed of the pseudo-random number generator

# Instrument parameters
Instrument:
  Sampling : 25. # Sampling cadence of each camera [s]
  IntegrationTime : 21. # Integration time [s]
  NGroup : 4 # Number of camera groups (1 -> 4)
  NCamera : 6 # Number of camera per group (1->6)
  TimeShift : 6.25 # Time shift between camera groups [s]
  RandomNoise:
    Enable: 1
    Type: PLATO #  either 'User' or 'PLATO'. In the first case the NSR value is specified by the
                user (see below) while in the second case the NSR value is obtained by interpolating
                for the give magnitude NSR values expected for PLATO
    NSR :  73. # Noise to signal ratio [ppm/hr], for one single camera, this value takes into account
                all random noises but does not include systematic errors
  Systematics:
    Enable : 1
    Table : PLATO_systematics_EOL.npy # The binary file containing  the systematics parameters

# Stellar parameters
Star:
  Mag : 10. #  Magnitude
  ID : 12069449  #  star ID
  ModelDir : # Directory containing the single models or the grid (parameters and associated
              theoretical frequencies)
  ModelType: single  # Type of model: 'grid' or 'single' or 'UP'
  ModelName: 0012069449  # Name of the input model, to be specified when ModelType = 'single', ,
                         the program will then load the corresonding .gsm file generated with ADIPLS
  ES : ms  # Evolutionary status: 'ms' for the main-sequence phase, 'sg' for the sub-giant phase, 'rg'
            for redgiants (Red Giant Branch or clump stars)
  Teff : 5750.  # Effective temperature [K]
  Logg : 4.353 # Surface gravity, ignored for the UP
  numax : 179.3 # frequency of the maximum power [muHZ], used only with the UP
  delta_nu : 13.68 # Mean large separation [muHz], used only for UP, -1 if you want this parameter to
                     be derived from scaling relation
  DPI : 80.58 # Asymptotic values of the gravity mode period spacing [s], used only with the UP, -1
               if you don't want mixed modes to be included
  q : 0.15 # Mixed modes coupling factor, used only with the UP
  SurfaceEffects: 1 # Include mode near-surface effects, not implemented for the UP
  SurfaceRotationPeriod :  0. # Surface rotation period [days], not used with the UP
  CoreRotationFreq :  0. # Core rotation frequency [muHz], this is by definition Omega/2pi*1e6 where
                         Omega is the angular rate [rad/s], parameter used for UP only

  Inclination : 0. # Inclination angle [deg.]

Activity :
  Enable: 1
  Sigma : 40. # Amplitude of the activity component [ppm]
  Tau : 0.2 # Time-scale of the  activity component [days]

Granulation :
  Enable: 1

# Transit parameters
Transit :
  Enable: 0
  PlanetRadius : 0.5   # in jupiter radius
  OrbitalPeriod : 10.  # in days
  PlanetSemiMajorAxis : 1.  # in U.A.
  OrbitalAngle : 0. # in deg
  LimbDarkeningCoefficients: [0.25,0.75]
```

Appendix B: Configuration file for the red giant star KIC 9882316

```
# PLATO Solar-like Light-curve Simulator (PSLS) configuration file (V 0.8)

# Observation conditions
Observation:
  Duration : 730. #  [days]
  MasterSeed : 1704040900 # Master seed of the pseudo-random number generator

# Instrument parameters
Instrument:
  Sampling : 25. # Sampling period of each camera [s]
  IntegrationTime : 22. # Integration time [s]
  NGroup : 4 # Number of camera groups (1 -> 4)
  NCamera : 6 # Number of cameras per group (1 -> 6)
  TimeShift : 6.25 # Time shift between camera groups [s]
  RandomNoise:
    Enable: 1
    Type: PLATO #  either 'User' or 'PLATO'. In the first case the NSR value is specified by the
                user (see below) while in the second case the NSR value is obtained by interpolating,
                at the given magnitude, the NSR values expected for PLATO.
    NSR :  1. # Noise to signal ratio [ppm/hr], for a single camera. This value takes into account
              all random noise sources but does not include systematic errors.

  Systematics:
    Enable : 1
    Table: PLATO_systematics_EOL.npy # The binary file containing  the systematics parameters

# Stellar parameters
Star:
  Mag : 12.5 #  Magnitude
  ID : 9882316 #  star ID (integer)
  ModelDir : # Directory containing the single models or the grid (parameters and associated
             theoretical frequencies)

  ModelType: UP # Type of model: 'grid' or 'single' or 'UP'
  ModelName: KIC9882316  # Name of the input model, to be specified when ModelType = 'single'
  ES : rg  # Evolutionary status: 'ms' for the main-sequence phase, 'sg' for the sub-giant phase,
           'rg' for redgiants (Red Giant Branch or clump stars)
  Teff : 5400.  # Effective temperature [K]
  Logg : 3.934 # Surface gravity, ignored for the UP
  numax :  181.77120898 # frequency at maximum power [muHZ], used only with the UP
  delta_nu : 13.70885191 # Mean large separation [muHz], used only with the UP, -1 if you want this
                         parameter to be derived from a scaling relation
  DPI : 80.30739925 # Asymptotic values of the gravity mode period spacing [s], used only with the
                    UP, -1 if you don't want mixed modes to be included
  q :  0.15611347  # Mixed mode coupling factor, used only with the UP
  SurfaceEffects: 0 # Include near-surface effects in mode frequencies, not implemented for the UP
  SurfaceRotationPeriod :  0. # Surface rotation period [days], not used with the UP
  CoreRotationFreq :  0. # Core rotation frequency [muHz], this is by definition Omega/2pi*1e6 where
                         Omega is the angular rate [rad/s], used only with the UP
  Inclination : 0. # Inclination angle [deg.]

Activity :
  Enable: 0
  Sigma : 1000. # Amplitude of the activity component [ppm]
  Tau : 30. # Time-scale of the activity component [days]

Granulation :
  Enable: 1

# Transit parameters
Transit :
  Enable: 0
  PlanetRadius : 0.5   # in jupiter radii
  OrbitalPeriod : 10.  # in days
  PlanetSemiMajorAxis : 1.  # in A.U.
  OrbitalAngle : 0. # in deg
  LimbDarkeningCoefficients: [0.25,0.75]
```

RÉSUMÉ

PLAnetary Transits and Oscillations of stars (PLATO) est une mission scientifique spatiale européenne dédiée à la sismologie stellaire et à la recherche d'exoplanètes, et dont le développement est coordonné par l'Agence Spatiale Européenne. Avec un intérêt majeur sur des planètes du type terrestre située dans la zone habitable des étoiles du type solaire de la séquence principale, cette mission repose sur de la photométrie de très haute précision et exige une très grande stabilité des mesures. Elle s'appuie sur des techniques bien éprouvées : la méthode de transits pour la détection des planètes, suivie de mesures de vitesses radiales réalisées au sol, et l'analyse des oscillations stellaires pour la caractérisation des leurs étoiles hôtes. Grace à son très large champ de vue couvrant plus de deux mille dégrées carrés du ciel, l'instrument PLATO sera capable d'observer plusieurs centaines de milliers d'étoiles de magnitude apparente inférieure à treize dans le visible, et de détecter des milliers de systèmes planétaires. Par ailleurs, en raison des contraintes télémétriques du satellite, l'extraction photométrique de la majorité des étoiles devra être effectuée à bord en s'appuyant sur des méthodes optimisées. Pour autant, la méthode d'extraction photométriques par masques a été adoptée pour sa faisabilité compte-tenu des contraintes à bord. Dans ce contexte, la problématique de développement des masques photométriques optimaux représente le cœur du travail de recherche présenté dans cette thèse. Dans les missions précédentes de la même catégorie de PLATO, à savoir CoRoT, *Kepler* et TESS, des masques photométriques ont été conçus selon une approche reposant uniquement sur la minimisation du rapport bruit sur signal, car la sensibilité à laquelle un transit planétaire peut être identifié, dans une courbe de lumière, est fortement corrélée à son niveau de bruit. En revanche, plus il est facile d'identifier une planète, en raison d'un niveau de bruit suffisamment faible, plus élevée est la probabilité que des objets en arrière-plan entrant dans la scène (par exemple des systèmes binaires reproduisant des vrais transits planétaires) soient détectés. Étant donné que la plupart des étoiles-cible de PLATO n'auront pas des images au sol pour l'identification des faux positifs, une conception de masques photométriques reposant uniquement sur la sensibilité de détection des vrais transits planétaires, sans faire suffisamment attention aux potentiels faux positifs, n'est donc pas forcément la meilleure stratégie. Pour vérifier cette hypothèse, deux métriques scientifiques ont été introduites nous permettant de quantifier directement la sensibilité d'un masque à la détection des vrais, d'une part, et de faux transits planétaires, d'autre part. Ainsi, le masque optimal a été défini comme étant celui qui donne le meilleur compromis entre ces deux métriques. Cette approche, originale à cette thèse, s'est avérée décisive pour la détermination d'un modèle de masque statistiquement capable de détecter des planètes de façon quasi-optimale, tout en étant substantiellement moins sensible aux faux positifs. Globalement, ces travaux constituent une étape importante dans la conception des chaînes de traitement des données bord et sol de la mission PLATO, ainsi que pour ses algorithmes de détection et de caractérisation des transits planétaires.

MOTS CLÉS

PLATO – instrumentation – photométrie – réduction des données bord – masques photométriques – détection des planètes

ABSTRACT

PLAnetary Transits and Oscillations of stars (PLATO) is a European spatial scientific mission dedicated to asteroseismology and searching for exoplanets, and whose development is being carried out by the European Space Agency. With focus on Earth-like planets orbiting the habitable zone of main-sequence Sun-like stars, the mission relies on very high precision photometry and requires great stability of measurements. The mission is founded upon well-proven techniques: the transit method for detecting exoplanets, along with radial velocity follow-up from the ground, and the analysis of stellar oscillations for characterizing their host stars. Thanks to its very large field of view encompassing more than two thousand square degrees of the sky, the PLATO instrument will be able to observe several hundreds of thousands of stars with apparent magnitude lower than thirteen in the visible band, and thousands of planetary systems. In contrast, because of satellite telemetry constraints, photometry will have to be extracted in flight for most of the PLATO targets. For that, mask-based (aperture) photometry was adopted because of its sufficiently high performance and relatively low complexity for implementing on board. In this context, the development of optimal photometric apertures represents the core of the research work presented in this thesis. In the previous missions of the same category of PLATO (i.e. CoRoT, *Kepler* and TESS), photometric apertures were designed following an approach based uniquely on the minimization of the nose-to-signal ratio, because the sensitivity at which a planet transit can be found in a light curve is strongly correlated to its noise level. On the other hand, the higher the ease in identifying a transit-like signal because of a sufficiently low noise level, the higher the probability that background objects in the scene (e.g. binary systems reproducing legitimate planet transits) are detected. Since most of the PLATO targets will not have images available on ground for the identification of false positives, conceiving photometric masks based solely on how well a transit-like signal can be detected, paying no attention to potential false positives may not be the best strategy. To verify the consistence of this hypothesis, two science metrics were introduced allowing one to directly quantify the sensitivity of an aperture in detecting true and false planet transits. Then, the optimal aperture was defined as that which gives the best compromise between these two metrics. Such an approach, novel to this thesis, has been proven to be decisive for the determination of a mask model capable to provide near maximum planet yield and substantially reduced occurrence of false positives. Overall, this work constitutes an important step in the design of both on-board and on-ground science data processing pipelines of the PLATO mission.

KEYWORDS

PLATO – instrumentation – photometry – on-board data reduction – photometric masks – planet detection