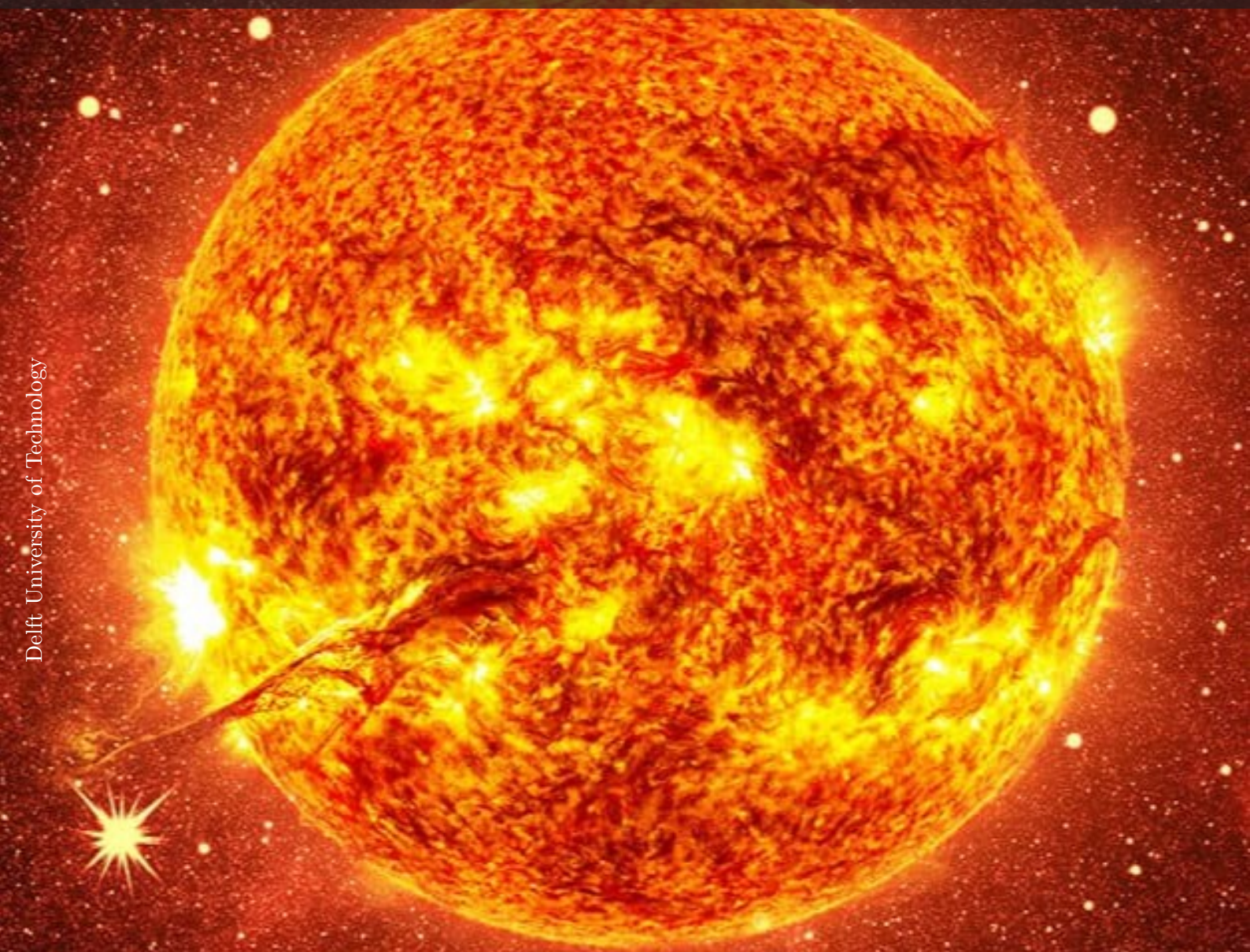
Building the stellar calibrator catalogue for the Ariel mission

Elena Tonucci



Delft University of Technology

TU Delft

Building the stellar calibrator catalogue for the Ariel mission

by

Elena Tonucci

Student number:	5593409
Academic supervisor:	Dr. M. A. Kenworthy
External supervisor:	Dr. T. A. van Kempen
University:	Delft University of Technology
Collaborative institution:	Netherlands Institute for Space Research (SRON)

Cover credit: iStock / Getty Images Plus

TU Delft

SRON Netherlands Institute for Space Research

Preface

To everyone who walked by my side,
the knots in the great scheme of things:
I'm just a small part of the plan,
but you're the real deal, I believe.

Look for paths, reasons, and signs,
the world wouldn't spin otherwise
but again as the fire ignites
the rain will come back to harmonise.

What is left is the will to understand
and to take a high leap of faith
so trivial, yet so hard to commence
when we think that we can control fate.

So, ultimately, this is actually for
all who left, ignored, or denied,
and to those who felt like fools too:
Now watch and rejoice (or surrender)
as the universe unfolds.

Elena Tonucci
Delft, September 2023

Table of contents

List of Figures

List of Tables

List of Symbols and Abbreviations

2MASS - Two Micron All Sky Survey
AIRS - Ariel InfraRed Spectrometer
Ariel - Atmospheric Remote-sensing Infrared Exoplanet Large-survey
BTJD - TESS Barycentric Julian Day
CBV - Cotrending Basis Vector
CCD - Charge-Coupled Device
CME - Coronal mass ejection
CoRoT - Convection, Rotation and planetary Transits
DR2 or DR3 - Data Release 2 or 3
EChO - Exoplanet Characterisation Observatory
ESA - European Space Agency
FFI - Full Frame Image
FGS - Fine Guidance System
FOV - Field of view
HR - Hertzsprung–Russell
HST - Hubble Space Telescope
IAP - Institute d'Astrophysique de Paris
ICRS - International Celestial Reference System
ID - Identifier
IR - Infrared
IUE - International Ultraviolet Explorer
JWST - James Webb Space Telescope
LC - Light curve
LSP - Lomb-Scargle periodogram
MAST - Barbara A. Mikulski Archive for Space Telescope
NASA - National Aeronautics and Space Administration
NIR - Near infrared
NIRSpec - Near-IR Spectrometer
PDC - Pre-search Data Conditioned
PLATO - PLAnetary Transits and Oscillations of stars
PSD - Power spectral density
SAP - Simple Aperture Photometry
SED - Spectral energy distribution
SIMBAD - Set of Identifications, Measurements and Bibliography for Astronomical Data
SNR - Signal-to-noise ratio
Spitzer - Spitzer Space Telescope
SPOC - Science Processing Operations Center
SRON - Netherlands Institute for Space Research
STScI - Space Telescope Science Institute
TESS - Transiting Exoplanet Survey Satellite
TIC - TESS Input Catalog
TOI - TESS Object of Interest
TPC - TESS Project Candidate

TPF - Target pixel file
UV - Ultraviolet
VISPhot - Visible Photometer

$^{\circ}$ - Decimal degree
$'$ - Minute of arc
$''$ - Second of arc
Δf - Frequency spacing
Δt - Sampling period
δ - Impulsive function
μ - Mean
σ and σ_{μ} - Standard deviation from the mean
σ_n - Instantaneous uncertainty of the data points in a time series
σ_{Mdn} or σ_m - Dispersion around the median
χ^2 - Chi-squared
χ^2_v - Reduced chi-squared
$\mathfrak{F}$ - Continuous Fourier transform
$\mathfrak{F}_d$ - Discrete Fourier transform
Å - Angstrom
A - Amplitude
$A(f)$ - Power amplitude spectrum
amp - Arbitrary amplitude unit
C_F - Calibration factor(s)
d - distance
Dec - Declination
dex - Decimal exponent
e^- - Electron
F_d and f_d - Flux density
FSR - Flux stability requirement
f - Frequency
f_s - Sampling frequency
$g(t)$ - Continuous signal
H - H wavelength band
Hz - Hertz
$H\ mag$ - Magnitude in the H band
h - Hour
I_C - Johnson−Cousins I_C filter
i - Imaginary unit
J - J wavelength band
$J\ mag$ - Magnitude in the J band
Jy - Jansky
K - Kelvin
K_f - Degrees of freedom
K_s - K_s wavelength band
$K_s\ mag$ - Magnitude in the K_s band
$Kep\ mag$ - Kepler magnitude
$log(g)$ - Surface gravity
Mdn - Median
Mdn_i - Running median

m - Meter
mag - Magnitude
mas - Milliarcsecond
min - Minute
N - Total number of samples
n - Iterative index or multiplicative factor
n_s - number of samples
$P_S(f)$ - Classical periodogram's power spectrum
P_g - Fourier's power spectrum
P_{max} - Maximum power amplitude peak of LSP
pc - parsec
$pmdec$ - Proper motion in Declination direction
$pmra$ - Proper motion in Right Ascension direction
ppm - Part per million
R_C - Johnson−Cousins R_C filter
R_{Sun} - Solar radius
RA - Right Ascension
S_{back} and s_{back} - Background-subtracted signal
SE - Standard error
s - Second
T - Total time length of the signal
T_{eff} - Effective temperature
T_{mag} - TESS magnitude
t - Time
V - Johnson−Cousins V filter
W - Watt
x_n - Data points in a time series
Z - Metallicity
z - Sloan Digital Sky Survey z filter

Abstract

Ariel is an ESA space telescope planned for launch in 2029. It is aimed at observing ∼1,000 exoplanets using transit spectroscopy at infrared wavelengths between 0.5-7.8μm, to characterise the chemical composition and thermal structure of their atmospheres. To detect minute flux variations and confidently extract the exoplanet signal, Ariel requires good instrumental flux stability. Moreover, the instrumental response must be monitored through time, and instrumental effects and stellar noise must be estimated and minimised to get as close as possible to the photon noise limit. To do this, Ariel must be calibrated in-flight by observing a set of stellar calibrators, stars whose spectral energy distribution is assumed to be well-studied and constant. This project is aimed at building the stellar calibrator catalogue for Ariel and looking for possible correlations between stellar properties and flux stability. This can also serve as a base study for future space missions other than Ariel. A starting candidate sample mainly including G dwarfs that were observed by TESS is filtered with a three-step procedure using the Lomb-Scargle periodogram and the reduced chi-squared statistic. Assuming an Ariel 3σ flux stability requirement of $100ppm$ over 6 hours, the defined Ariel Catalogue comprises 581 stars. These are fairly homogeneously distributed over the sky, without periodic variability, and without a large flux excess dispersion. It is found that when selecting the correct effective temperature ranges ($5,000 < T_{eff} < 6,300K$), the probability of finding stable stars is almost constant at 30%. Only the dimmer stars seem to have a higher probability of passing the selection, which is possibly due to the higher noise in the data which makes variability more difficult to identify. Note, however, that the catalogue defined in this study still comprises calibrator candidates. The noise in TESS light curves is probably too high to effectively analyse the stability within the current requirement. So, follow-up observations of these stars with a lower noise are recommended, for instance with PLATO, to greenlight them as final Ariel calibrators.

1

Introduction

The Atmospheric Remote-sensing Infrared Exoplanet Large-survey (Ariel, Tinetti et al. 2021) is a European Space Agency (ESA) M-class mission that will be launched in 2029. Ariel will study the physical processes affecting planetary atmospheres as well as their composition and evolution. It will target around 1,000 exoplanets distributed over the sky with a wide range of properties, conditions, and host star types. The main goal is to study exoplanet populations. The observation strategy makes use of three different techniques: Transit, secondary eclipse, and phase variation, all in infrared (IR) spectroscopy, specifically in the wavelength range between 0.5-7.8μm. During a transit, a small part of the radiation coming from the star is transmitted through the exoplanet's atmosphere. During a secondary eclipse instead, the thermal radiation and reflected starlight from the exoplanet are blocked. Lastly, the phase variations are minute changes in flux caused by observing different sides of the planet as it orbits about its host star. These techniques enable us to infer some properties of the planetary atmospheres like their composition, thermal structure, cloud properties, and others.

To reach its objectives, Ariel must be able to detect minute features in the spectra of exoplanetary transits and eclipses. In the IR, the difference between the stellar and planetary signals can be very small, with a planetary contrast in the orders of $1,000$ to $10ppm$ (parts per million) during secondary eclipses for Jupiter-sized to super-Earth-sized planets (Waldmann et al. 2013). Sometimes, many transits are required to get a sufficient signal-to-noise ratio (SNR). This allows us to confidently extract the planetary signal. A good instrumental photometric (or flux) stability is crucial for individual transits and for combining different observations. Relative flux stability is the degree to which an instrument maintains its flux measurement accuracy over a period of time. This is a driving requirement for the Ariel mission since a certain level of accuracy is needed to identify the desired minute variations in flux originating from the planet. Particularly, it is desired that Ariel's observations reach the lowest noise possible, i.e. they are photon noise-dominated. Any other noise contributions, instrumental effects, or other systematics should be removed or minimised. The required flux stability is tested during the instrument design process. However, the instrument's performance must be monitored over time once the spacecraft is in flight. Moreover, instrumental effects and stellar noise must also be estimated and minimised in the observations. In order to do this, Ariel must be calibrated in flight by using external flux references.

In astronomy, it is a general practice to observe stellar calibrators, stars whose spectral energy distribution (SED) is assumed to be well-studied and constant through time. By observing a

set of calibrators with different properties, like their brightness, the signal contribution coming from the star and unwanted instrumental effects can be derived and later compensated. So, flux calibration is essential to confidently extract the planetary signal and reach the scientific objectives of the mission.

The stars that have been historically employed as calibrators for space-based missions are mainly white, A, and G dwarfs. This is because these spectral classes are known to generally be photometrically stable. A set of stars should always be specifically selected for every new mission, following its characteristics. For Ariel, this holds more than ever, as stars used in previous missions do not possess suitable properties that are desired for Ariel calibrators. For example, Ariel will possess a wide sky coverage, so calibrators should ideally be distributed homogeneously over the sky so that at least one of them is observable at any time (Petralia, Micela, et al. 2017). Moreover, they need to be stable within Ariel's flux stability requirement, and in the correct brightness range. For its importance, the construction of a stellar calibrator catalogue is of high interest within the Ariel community. The first selection of the candidates started more than a decade ago but was never completed. Based on the findings of the photometric variability population study using data from the Kepler mission (Koch et al. 2010) by Ciardi et al. 2011, it was assumed that statistically, many stable stars would be available to photometrically calibrate Ariel. In particular, the study by Ciardi et al. 2011 claims that the most stable group of stars is G dwarfs: around 80% of them are stable. Moreover, it was found that around 50% of both F and K dwarfs also are stable. This project is aimed at selecting stars that are photometrically stable, so as to build the stellar calibrator catalogue for the Ariel mission. While doing this, a flux variability population study of the calibrator candidate sample will be carried out with the goal of looking for possible correlations between stellar properties and different levels of flux stability. It is worth pointing out that, although this project is based on the Ariel mission's characteristics and requirements, it can also serve as a base study for future space missions in need of selecting stellar calibrators.

Chapter 2 clearly states the research objectives of this study and their motivation, as well as the questions that need to be addressed in order to reach those objectives. Chapter 3 describes the scientific background at the base of this study. This includes absolute and relative flux calibration, flux stability, stellar photometric variability types, a historical overview of stellar calibrators, and the findings of the flux variability population study by Ciardi et al. 2011. Chapter 4 focuses on the Ariel mission, going more in-depth into its objectives, observation strategy, photometric performance and requirements, as well as into the first Ariel flux calibration efforts. Chapter 5 describes in detail the databases, datasets, and mathematical and statistical tools that are used to obtain the results, which are instead presented in Chapter 6. The results include the selection procedure to determine the most promising Ariel calibrators and a photometric variability population study that is compared with the results from Ciardi et al. 2011. Moreover, a discussion section contains considerations and caveats on the final results, before closing with recommendations for future work. Finally, the work is summarised in Chapter 7.

2

Research objectives

In-flight calibration employing stellar calibrators as an external flux reference is crucial to reach the lowest noise and best stability possible in the observations and monitoring that the instruments are working correctly through time. The definition of Ariel's stellar calibrator catalogue has been ongoing for more than a decade, but a stability analysis of the candidates was never performed due to the lack of data. The construction of a catalogue comprising promising stellar calibrators is of utmost importance to ensure that Ariel's results can be trusted.

In this Chapter, the research objectives of this project are defined in detail, as well as the questions that must be answered to reach them. The primary objective (Section 2.1) is focused on the Ariel mission and the relation of stability with stellar properties, while the secondary one (Section 2.2) is more generally related to different types of variability and stability levels.

2.1. Primary research objective

The primary research objective of this project is:

To select stars to build the stellar calibrator catalogue of ESA's Ariel mission.

In order to reach this objective, the project will address the following questions:

1. *Which stars possess the required characteristics to be Ariel stellar calibrator candidates?*
2. *Which correlations are there between stellar flux stability and stellar properties, if any?*

Answering these questions will allow us to select the Ariel stellar calibrators properly. The selection must be performed starting from the Ariel instruments' characteristics and requirements, and stellar properties. These include:

- Properties of each single star - limit magnitudes following Ariel's faintest (M5V star with a K_s magnitude of 8.8, equivalent to GJ 1214) and brightest (K0V star with a K_s magnitude of 4.0, equivalent to 55 Cnc) observable targets (Tinetti et al. 2021).
- Properties of the whole group of stars - total number of stars and their distribution over the sky. The ideal goal is to have a significant number of calibrators spread in various sky directions so that Ariel can observe at least one of them at any time to monitor the instrumental response and minimise slewing (Petralia, Micela, et al. 2017).
- Flux stability requirement - currently defined as an overall photometric stability of $100ppm$ (3σ) or better over a nominal integration time of 6 hours (length of an Ariel observation).

The first step to undertake is the creation of a starting calibrator candidate sample comprising stars with desired properties, like magnitude and spectral type, that make them possible Ariel calibrators. Afterwards, the data and the mathematical tools needed to analyse the flux variability of the stars must be selected and how to properly employ them must be understood. So, a selection procedure must be defined for the identification of the stars among the candidates that do not appear to be photometrically variable. Finally, an analysis of their stellar properties, total number, and distribution in the sky is made to evaluate if enough calibrators are available and explore their characteristics. The findings from the analysis of the most stable stars can also be useful for missions other than Ariel, to identify possible correlations between stellar characteristics and flux stability. The most interesting properties to explore include effective temperature, magnitude, surface gravity, and metallicity as a proxy for age.

2.2. Secondary research objective

The secondary research objective is defined as:

To perform a flux variability population study of the stellar calibrator candidate sample.

During the analysis of the sample and the selection of the most promising candidates for Ariel, a general flux variability population study can be performed. The goal is to study different types and levels of flux stability in the sample, also relating them to stellar properties. This goal can be translated into the following secondary research questions:

1. *What percentages of the sample consist of stars with different variability physical origins?*
2. *What percentages of the sample consist of stars with different variability levels?*

A categorisation can be made into different physical origins of flux variability. These include different behaviors of the stellar flux, like constant, variable, dominated by intrinsic properties of the star (rotational modulation, pulsations), or dominated by extrinsic properties of the stellar system (eclipsing binaries, transiting exoplanets, or other types of sporadic variability like stellar flares and coronal mass ejections (CMEs)). Moreover, a categorisation into different variability levels of the calibrator candidates (variability fractions) can be performed while analysing the starting sample. The findings can be compared with estimates from the relevant previous population study by Ciardi et al. 2011, corroborating or refuting its claims about the expected percentage of stable stars per spectral class. This will allow us to refine our expectations of the typical stability of calibrators and deepen our understanding of them. The considerations stemming from the population study can set a base for future stellar stability studies.

3

Background

This Chapter contains the background knowledge needed to properly understand the rest of the study. First, the transit and eclipse spectroscopy technique is described (Section 3.1). For the success of this technique, flux calibration is needed, which is presented in detail in Section 3.2. This includes a differentiation between absolute and relative flux calibration. Flux stability is instead described in Section 3.3. Afterwards, an overview of stellar variability and stellar calibrators is given (Section 3.4). This includes information on stellar photometric variability by spectral class and a historical outline of the stellar calibrators employed by past and current space telescopes as external flux references. Section 3.5 shows the most important results of the stellar flux variability population study by Ciardi et al. 2011. To conclude, Section 3.6 gives an overview of the recent mission Transiting Exoplanet Survey Satellite (TESS, Ricker et al. 2010) whose characteristics make it a relevant mission for stellar variability analysis.

3.1. Transit and eclipse spectroscopy

Transit and eclipse spectroscopy is a widely employed method for exoplanet characterisation. It enables us to infer something about the bulk composition of planets and their atmospheres. This technique is extremely useful because it does not require spatially resolving the radiation from the planet separately from the one coming from the star. Instead, the variations in the flux coming from the full system are exploited. Figure 3.1 shows an example of the observed total flux as the planet orbits about its host star. Naturally, for spectroscopic observations, the flux is also resolved in wavelength.

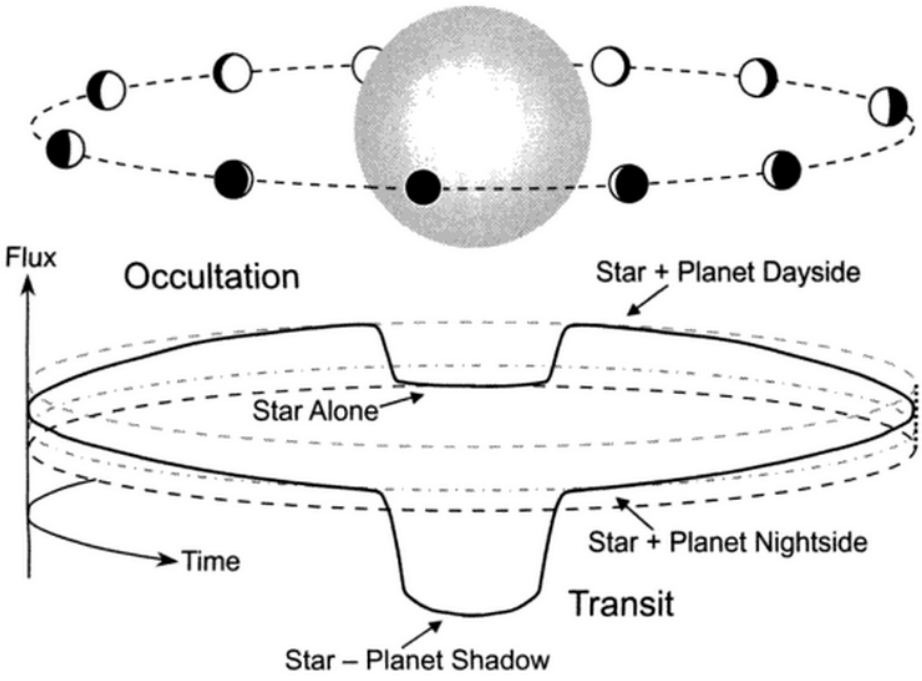


Figure 3.1: Observed flux from a planetary system as an exoplanet is orbiting about its host star. *Credit: Winn 2010.*

During a transit, the planet is moving in front of its host star with respect to the observer, blocking some of the radiation coming from it. This is transmitted through the atmosphere of the planet. So, by comparing the observations in and out of transit, it is possible to obtain the absorption spectrum originated by the planet, enabling us to study the composition of its atmosphere. In particular, transit spectroscopy probes the high-altitude regions of the atmosphere (Tinetti et al. 2012).

During a secondary eclipse (also called occultation), the planet moves behind the star, and the reflected and emitted radiation coming from it is blocked. In this case, the difference between observations in and out of eclipse provides the spectrum of the planet's day side, i.e. the side that is facing the star. This method probes lower altitudes of the atmosphere (Tinetti et al. 2012).

Additionally, as the planet orbits about its host star, varying parts of the different sides of the planet are observed. Measuring the minute flux variations as a function of the orbit can

allow us to study the dynamics and redistribution of energy between the planet's day and night sides. This technique is also used for non-transiting planets. Finally, transit, secondary eclipse, and phase curve observations can be combined to derive the spectrum of the night side of the planet (Tinetti et al. 2012).

Since the flux variations can be extremely small ($10-1000ppm$, depending on the ratio between the planetary and stellar radii (Waldmann et al. 2013)), multiple orbits are sometimes needed to improve the SNR and confidently extract the planetary signal from the total one. Stellar activity is especially limiting as it introduces noise. So, a good instrumental flux stability is needed for these types of observations. Flux stability is a driving requirement for the instrument design of transit missions. However, the required level of stability should not only be reached but also guaranteed in flight and through time. Monitoring of the instrumental performance, removal of unwanted instrumental effects, and minimisation of the stellar noise are all crucial factors for the success of the mission. All of this is performed by flux calibration.

3.2. Flux calibration

Calibration is defined as the process of comparing the measurements given by a device with those provided by a calibration standard of known accuracy. Different types of calibration can be defined, including photometric (or flux), astrometric, radiometric, and wavelength. In particular, flux calibration is crucial for the success of spectrophotometric measurements and the monitoring of instrumental behavior through time. It provides a reference to which observed quantities can be compared, a means of translating the variables from instrumental to physical units, and a way to minimise or remove instrumental effects or other error sources in scientific observations.

When talking about flux calibration, one should distinguish between absolute and relative. Absolute flux calibration relates stars to absolute standards. It is mainly aimed at converting the measurements from instrumental to physical units, and it allows us to derive calibration factors that are later useful to relate sources together (Section 3.2.1). Relative flux calibration can instead be seen both as a way to relate and compare stars between each other (Section 3.2.2), and each star with itself through time (flux stability, Section 3.3).

In astronomy, flux calibration is usually performed by observing a set of stellar calibrators that are used as external flux references. Stars are the most used for this purpose. Therefore, it is necessary that the stars used as stellar calibrators are stable within the stability requirement of the instrument, so as to calibrate it properly.

3.2.1. Absolute flux calibration

Answering most science questions of space-based telescopes requires knowledge of the flux density and surface brightness of the target. To obtain this, stellar calibrators are observed. Stellar calibrators are stars whose SED is assumed to be well-studied and constant through time. Commonly, the relation between the expected flux value and the actual flux measurement of a calibrator star is used to convert the measurements between instrumental and physical units. This is done through the calculation of calibration factors, C_F (Bohlin et al. 2014):

$$C_F = \frac{F_d}{S_{back}} \tag{3.1}$$

where F_d is the expected flux density and S_{back} is the background-subtracted observed signal. Calibration factors are computed for multiple calibration stars with different magnitudes to create a relationship between instrumental units (e.g. electrons per second $[e^-/s]$ at the detector) and physical units of spectral flux density (e.g. Janskys $[Jy]$ or $W/(Hz\ m^2)$). Calibration factors later allow us to convert between units for any target source, and to compare and relate sources between each other. To translate the signal observed in instrumental units into flux density for any target source it is possible to use:

$$f_d = C_F s_{back} \tag{3.2}$$

where f_d is the flux density of the target source and s_{back} is its background-subtracted observed signal.

The knowledge of the SED is mainly limited by the precision of stellar atmosphere models, which are the tools used to predict it (Bohlin et al. 2014), and it is usually known in a limited wavelength range. Stellar atmosphere models are nowadays fairly accurate at modelling stellar spectral types that are mainly dominated by radiative rather than convective transport (Gordon et al. 2022). The atmospheres of pure-hydrogen white dwarfs are simple to model at optical wavelengths, so these stars have been historically used as standard calibrators, for example for the Hubble Space Telescope (HST). Models of these types of stars are accurate to 1-2% in the wavelength range between ultraviolet (UV) and infrared (IR) (Deustua et al. 2012). However, although models can be quite good at modelling the general characteristics of a spectral type, it can be more difficult to accurately model the behavior of a specific star belonging to that type.

As an example of absolute flux calibration efforts, the recent James Webb Space Telescope (JWST, Gardner et al. 2006) Absolute Flux Calibration Program (Gordon et al. 2022) of Cycle 1 (first year of observations) can be taken above all. Its aim was to enable a robust and coherent flux calibration both between the different instruments of JWST and between observation modes within each instrument. To do so, measurements of flux densities from a network of calibration stars were performed and then compared to their predicted flux densities resulting from stellar atmospheric models. The absolute flux prediction accuracy was set to 2%, with the goal of pushing the limit to 1%. As the goal also included the empirical quantification of both statistical and systematic uncertainties in the calibration process, following Gordon et al. 2022, the required calibration observations included:

- Multiple calibration stars, to obtain an empirical measurement of statistical accuracy.
- Calibration stars of different spectral types, to test for systematic biases that might arise from the stellar atmospheric models because of an incomplete understanding of stars.
- Calibration stars with different flux densities, to test the calibration at different flux density levels and the performance of cross-calibration between instruments, as they possess different sensitivity ranges.

Averaging measurements of multiple calibration stars of different stellar types is required because the stellar atmosphere models used are fairly good at modelling classes of stars, but generally show lower accuracy at modeling any specific star belonging to the class (Bohlin et al. 2011). To reach the goal of 1% accuracy per type of star, Gordon et al. 2022 estimated

that there is a need for five calibrator stars of each selected spectral type to account for issues that can arise from individual stars and estimate the systematic errors by type. This number also allows for one star out of the five to be found not suitable with the new observations, and so be discarded from the analysis later.

3.2.2. Relative flux calibration

First, relative flux calibration is concerned with the comparison of different stars with each other. Since the calibration factors carry within them a certain level of uncertainty, it is sometimes easier to compare the flux of different sources with each other to relate any target star with absolute standards. To do this, the relative flux difference between the sources that have not been calibrated in an absolute way and the ones that were is used.

Second, relative flux calibration can also be applied to each star with itself, to monitor its flux variations through time. This is referred to as flux stability. The importance of instrumental flux stability for transit spectroscopy and its function is described more in detail in the next Section.

3.3. Flux stability

Stability defines the degree to which an instrument maintains its metrological characteristics or, on the contrary, changes its measurement accuracy over a period of time. A change in average flux is usually called “drift”, and is defined as the variation in the measurement due to variations in the metrological characteristics of the instrument. So, flux stability is concerned with the capability of an instrument to distinguish variations in flux of at least a desired value through time.

Usually, for transit spectroscopy, a certain requirement is set on the flux stability that the instruments need to reach. This is generally expressed in the form of the smallest flux variation over a desired integration time that the instrument must be able to distinguish. This is very important because the signal coming from the transiting planets can be extremely small. This means that any source of noise should be minimised as much as possible, to get to photon noise-limited observations. Photon noise is the randomness in signal associated with the particle nature of photons collected by the instrument and is therefore a physical limit, representing the lowest noise reachable.

There are three main factors that can limit flux stability for spectrophotometric space-based missions (Waldmann et al. 2013):

- Pointing stability
- Thermal stability
- Stellar noise and other external temporal noise sources

To ensure that an instrument is photometrically stable within its requirement, the main drivers for its design are its pointing and thermal stability. These can be achieved by keeping the jitter to a minimum and cooling down the instrument, especially the detector, to reduce the thermal noise. Stellar noise and other temporal sources of flux instability that need to be corrected are, however, beyond the control of the instrument design (Waldmann et al. 2013).

To deal with those, stellar calibrators are observed. Such stars are assumed to possess a stable flux within the desired requirement, i.e. within a defined amplitude and time scale. So, assuming that a stellar calibrator's flux is stable and known, the instrumental response can be monitored in time and modelled through calibration observations. This is possible by comparing the observed flux with the expected one predicted by the models.

Calibration observations include observing a stellar calibrator at different times but under the same conditions, and observing a set of different stellar calibrators with diverse properties.

- The first type of observation is used to monitor the instrumental response in time, by checking how much observations of the same flux reference differ through time. If two measurements of the same calibrator star differ, it is assumed that the origin is instrumental, and not physical. This is of course a simplification since calibrators are in reality not perfect flux references. But this allows us to estimate the contribution to the signal coming from instrumental effects and stellar noise. Afterwards, this can be removed from the other scientific observations.
- The second type of observation is instead used to draw a relation between stellar properties, observed flux, and instrumental response. For example, first, a stellar calibrator is observed, which gives a certain flux value at the detector. Later, another calibrator that is twice as bright is observed. Ideally, the flux value at the detector should be twice the one from the first calibrator. However, this is not always the case due to a variety of reasons, which include differences in the instrumental response in relation to different stellar properties. So, the response can be modelled and then estimated for any target star with diverse properties.

To conclude, all calibration observations are aimed at modelling the instrumental effects and stellar noise so as to remove their contributions from the scientific observations of any other target. Moreover, to increase the SNR, multiple observations of the same transiting planet are performed, and flux stability is also critical to successfully combine different observations. Calibration measurements ensure that the detected flux variations are of a physical nature and come from the process that it is desired to observe. Instrumental effects, other temporal noise sources, and stellar noise are instead removed or minimised, to get as close as possible to photon-noise-limited observations that can be trusted.

3.4. Stellar calibrators

Stellar calibrators are the most commonly employed astronomical sources that provide an external flux reference to perform flux calibration of space-based telescopes. It is assumed that their flux is stable and well-studied. Naturally, such stars are actually often variable, but at time scales that are longer than the transits' time scales, or with an amplitude that cannot be detected by the instruments (Deustua et al. 2012). Moreover, their predicted flux value is usually obtained through stellar atmosphere models, which will be affected by a certain error.

To understand which stars are good calibrators, first, it is important to take into consideration concepts like spectral type, stellar structure, and stellar evolution, as these play a large role in defining photometric variability in stars. Aside from these intrinsic properties, external conditions can also play a role, like the type of system the star belongs to (for example, single, binary, or multiple). This is presented in Section 3.4.1. Some spectral types and some specific stars have been employed in previous calibration efforts for both ground and space-based

missions thanks to their characteristics which make them stable. A historical overview of them is presented in Section 3.4.2. Learning from them and striving to look for more and more stable calibrators is essential to ensure that future missions will be provided with suitable stars for their flux calibration.

3.4.1. Stellar photometric variability

Stars whose magnitude changes in time are termed "variables". Variables have been widely studied in the literature and their number has grown more and more together with the increase in the observations' precision. In fact, all stars are variable, but some types of variability might be too small to be picked up by instruments. Moreover, as the structure of the star changes as it evolves, the stellar flux behavior can change in time depending on the evolutionary stage the star is in.

Variables are classified based on the shape of their light curves and their spectral type or class (Karttunen et al. 2007). A light curve is the brightness variation as a function of time, whilst the spectral class is a short code made of letters and numbers indicating the ionization state of the star, and it is strictly related to its effective temperature. Figure 3.2 shows the position of the main variables on the Hertzsprung–Russell (HR) diagram.

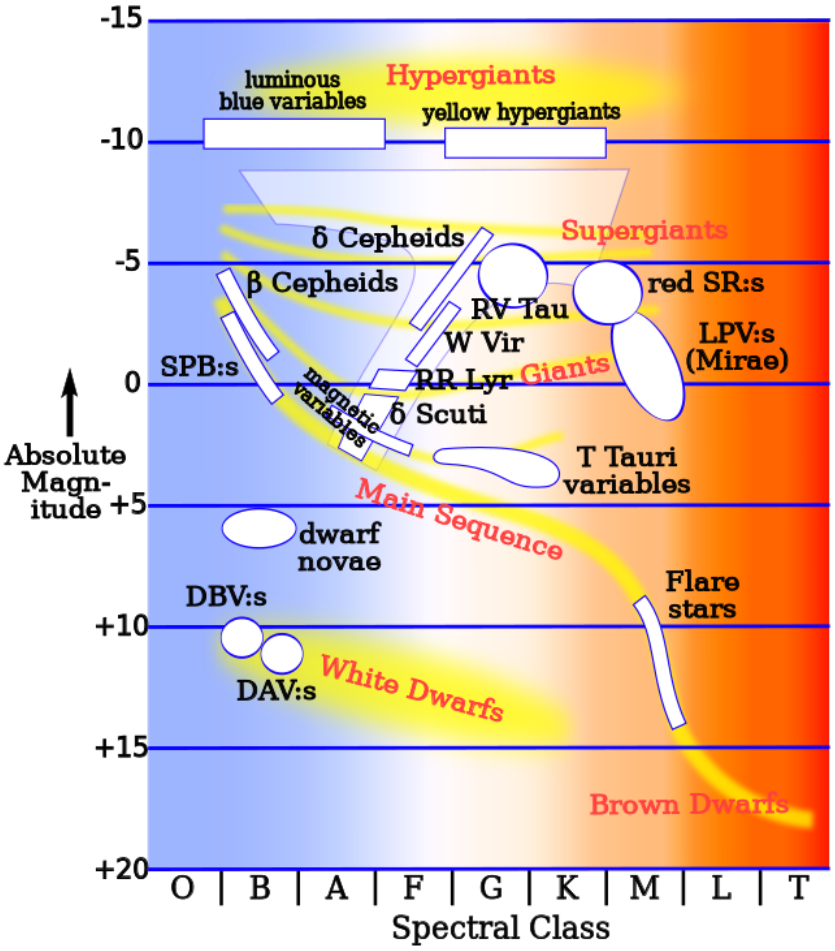


Figure 3.2: Hertzsprung–Russell diagram with instability strips labelled. *Credit: Rursus, Wikimedia Commons.*

Variables can be divided into four main groups. For the first three, the brightness variations are due to the intrinsic properties of the stars, including temperature variations, rotation, or layers' instabilities.

1. *Pulsating variables*

Pulsations in stars are caused by a periodic, alternate expansion and contraction in the outer layers. These variables can be giant and supergiant stars in an unstable stage of their lives (for example, Cepheids, RV Tauri, and RR Lyrae stars). Although the diameter of the star can double in size during the pulsation, this effect is not the main driver for the change in brightness. The flux variation is mainly due to the periodic variation of the surface temperature of the star. In fact, the luminosity is proportional to the fourth power of the effective temperature. Stars with a surface temperature between $6,000$ and $9,000K$ can oscillate due to this instability. An important sector of the HR diagram where pulsators can be found is called the "Cepheid instability strip". Here, δ Scuti (also referred to as dwarf Cepheids, Breger 2000) and γ Doradus variables (Kaye et al. 1999) can be found, usually belonging to the A or F dwarf classes (Uytterhoeven et al. 2011). δ Scuti variables pulsate with frequencies between 18 minutes and 8 hours in pressure modes and sometimes mixed with low-order gravity modes. γ Doradus variables pulsate instead in non-radial high-order gravity modes, generally with periods of 1 day (Grigahcène et al. 2010). Some stars can also be hybrid pulsators between these types.

2. *Rotating variables*

In rotating variables, the brightness variations are caused by an uneven temperature distribution on the stellar surface which is filled with cool or hot spots (stellar spots and plages). These spots are darker or brighter than the rest of the surface and appear and disappear from view as the star rotates about its axis, modulating the light curve (Karttunen et al. 2007). The shape given by spots rotating in and out of view is usually a fairly periodic semi-sinusoid (Ma et al. 2022). Stars belonging to this category are mainly stars with strong magnetic fields, which give rise to star spots, for example, magnetic A stars (Balona 2017). The period of rotation varies from about 1 to 25 days (Karttunen et al. 2007). The Sun itself is a weak rotating variable due to spots and, generally, most stars are probably affected by stellar spots. Moreover, the younger the star, the faster it will rotate, and the stronger its activity will be. So, stars tend to become quieter and more stable as they get older, and their activity, including spots on their surface, stellar wind, and magnetic activity, decreases.

3. *Eruptive and cataclysmic variables*

Eruptive variables are generally faint stars ejecting mass, mostly belonging to close binary systems in which mass is transferred between the stars. Their variability is not regular or periodic, with sudden outbursts of mass in the corona. Usually, these stars are surrounded by gas or interstellar matter which also participates in the eruption. Flare stars are part of this group (Karttunen et al. 2007). Cataclysmic variables are similar but the eruptions are due to nuclear reactions in the surface or in the interior of the star. If the explosion is so violent that it destroys the star, we talk about novae-like stars or supernovae.

4. *Eclipsing variables*

Differently from the other types of variability that are caused by intrinsic properties and physical changes of the star, the variability of eclipsing stars is due to external characteristics

of the system. Eclipsing variables are part of a binary or multiple stellar system where one star periodically passes in front of the other with respect to the observer, causing a change in the flux measurement. The shape of these light curves includes regular dips that are usually quite deep, although the size depends on the relative size between the companions.

3.4.2. History of stellar calibrators

Calibrator stars that are normally used from the ground can guarantee stability only in the order of 1%, while most space-based telescopes now require two orders of magnitude better. The stars that are used from the ground for absolute flux calibration are called primary irradiance standards. Although the precision of laboratory reference standards has reached a higher accuracy than what is needed for astronomical research, it is difficult to transfer these results to stars around the whole sky and in a wide range of wavelengths. Since stars are used as surrogate flux standard sources, they would ideally need to be calibrated against laboratory standards at the full range of wavelengths. However, only a few sources underwent this process, and only for specific wavelengths (Deustua et al. 2012). Alpha Lyrae (Vega), for example, has been calibrated at a wavelength of 5,556Å and its recommended SED between 3,300 and 10,500Å has been compiled (Oke and Schild 1970, Hayes and Latham 1975). Following discussions on the suitability of Vega as an absolute flux standard, due to evidence that shows it is a pole-on rotator with a dust disk, Sirius has also been studied. Alpha Canis Majoris, or Sirius A, is almost four times brighter than Vega. It also has a faint companion, Sirius B, which is a white dwarf. The Sun can finally be used as an absolute calibrator for near-infrared (NIR) and longer wavelengths, as its spectrum and flux have been deeply studied by many space missions and ground-based experiments (Deustua et al. 2012).

Usually, it is desired to observe stellar calibrators in specific parts of the sky or all around it. So, it is important to define some criteria that help identify possible good calibrators. These include (Gordon et al. 2022):

- A mixture of stellar properties such as (1) being a single star (not part of an eclipsing binary or multiple system), (2) belonging to a spectral type that is straightforward to model (for instance, modeling a stellar atmosphere dominated by radiative transport rather than convective transport is easier), (3) not possessing or possessing a small detectable photometric variation, (4) not having a circumstellar disk, (5) not possessing a rapid rotation and not being very active (with few stellar spots and low stellar wind), which also means that older stars are preferred over younger ones;
- A low or well-modelled line-of-sight dust extinction;
- The use of stars belonging to calibration programmes of previous missions, which means that observations are already available.
- Stars that are not too far away, as this obviates the problem of interstellar reddening (Deustua et al. 2012).

Following these characteristics, the best spectral types to be used in calibration efforts are hot stars (white dwarfs and OB stars), A dwarfs, and solar analogs (late F and early G dwarfs). These spectral types are the most quiescent ones and have been employed in previous space missions in need of being calibrated.

- White dwarfs possess simple atmospheres dominated by radiative transport that are fairly easy to model with stellar atmosphere models. Moreover, some white dwarfs have been

used extensively for the flux calibration of HST (Stone 1996), and bright OB dwarf stars were used to calibrate the International Ultraviolet Explorer (IUE, Wamsteker 1993).

- A dwarfs also possess simple radiative transfer-dominated atmospheres and have mainly been used in the past for the flux calibration of the Spitzer Space Telescope (Spitzer, Werner et al. 2004).
- Solar analogs are among the most stable stars (Ciardi et al. 2011) and are similar to the Sun, which is the most studied and observed star. Moreover, they constitute the primary calibrators for many ground-based NIR observations.

All these three types are included in the absolute flux calibration plan of JWST (Gordon et al. 2022).

With the need for more and more stable calibrators arising due to the advancements in instrumental capabilities, a recent study has demonstrated that new, higher-accuracy missions like TESS, detected variability in some Spitzer calibrators (Mullally et al. 2022). Such variability was below the detection capability of Spitzer itself, i.e. a stability of around 1% (Reach et al. 2005), but it is now visible with improved instrumental sensitivity. Other modern space-based missions like Convection, Rotation and planetary Transits (CoRoT, Auvergne et al. 2009) or Kepler can reach high precision only after instrumental effects have been removed. So, it is important to exploit the knowledge coming from past, present, and future observations to monitor the best candidate stars and be able to select the most suitable calibrators for each different mission and application.

Variability VS stability

Recently, TESS observations, distributed over a large area of the sky, have enabled new transiting exoplanet as well as stellar variability studies (Fetherolf et al. 2022). An extremely recent and interesting study aimed at looking for variability in candidate calibrator stars of JWST was carried out by Mullally et al. 2022. In this study, TESS light curves are analysed to make sure that the candidates do not show a flux variability larger than the desired one. However, studies aimed at identifying stable stars are rare. Many more are instead focused on a search for periodic variability or for transiting exoplanets, rather than on a search for stability in stars. These include, but are not limited to: the analysis of rotational modulation caused by stellar spots to recover stellar rotation periods (McQuillan et al. 2014, Claytor et al. 2022), the study of stellar pulsations in particular variables (Gautschy and Saio 1995), and the identification of eclipsing binary systems (Kirk et al. 2016, Pra et al. 2022).

Studies addressing stellar flux variability population studies have been carried out so far only for small portions of the sky, for example using Kepler data. The work from Ciardi et al. 2011, although more than ten years old, is still used as a reference when talking about statistics in stellar stability. The most important results of this study are presented in the next Section.

3.5. Past flux variability population studies

Ciardi et al. 2011 is a preliminary study on the overall flux variability statistics in stars that made use of the first quarter early-release Kepler data. The analysis was carried out based on light curve observations with a length of 33 days and a sampling time of 30 minutes. The stars in the sample were categorised into dwarfs and giants through discrimination of their surface gravity and effective temperature, resulting in around 126,000 dwarfs and 17,000 giants. These

two groups were further divided into the spectral classes A, F, G, K, and M for dwarfs, and G, K, and M for giants, through a temperature binning. The number of M and A-type stars is relatively small with respect to the other classes.

The statistical quantities used to characterise the photometric variability of the sample are the dispersion around the median and the reduced chi-squared. The first quantity is defined as:

$$\sigma_{Mdn} = \sqrt{\frac{\sum(x_n - Mdn)}{N}} \tag{3.3}$$

where σ_{Mdn} is the dispersion around the median, x_n represents the data points in the time series (flux, for light curves), Mdn is the median, and N is the total number of points in the time series. Mind that in the figures from Ciardi et al. 2011 the dispersion about the median is referred to as σ_m.

The second quantity is defined as (Andrae et al. 2010):

$$\chi^2_v = \frac{\chi^2}{K_f} = \frac{1}{N-1}\sum_{n=1}^{N}\left(\frac{x_n - Mdn}{\sigma_n}\right)^2 \tag{3.4}$$

where χ^2_v is the reduced chi-squared, χ^2 is the chi-squared, and σ_n is the instantaneous uncertainty of the data points in the time series (flux error, for light curves). Finally, K_f corresponds to the degrees of freedom i.e. the difference between the number of data points and the fitted parameters. In this case, the number of data points is N and the number of fitted parameters is 1 since the reference around which the reduced chi-squared is defined is a constant, i.e. the median of the time series.

χ^2_v is a unitless quantity that describes how much the data varies on top of the uncertainty in the measurements, which can be considered to be the instantaneous noise in the data points. The greater χ^2_v is with respect to 1, the more the star varies on top of the instantaneous noise and its flux dispersion with respect to the measurement uncertainty increases. Following Ciardi et al. 2011, it can be assumed that a star is very variable if $\chi^2_v \geq 100$ (excess dispersion of ~ 10 times the uncertainty), significantly variable if $10 \leq \chi^2_v < 100$ (excess dispersion of ~ 3 times the uncertainty), just barely variable if $2 \leq \chi^2_v < 10$ (excess dispersion of ~ 1.5 times the uncertainty), and stable if $\chi^2_v < 2$.

The most important results of this study are shown in Fig. 3.3 and 3.4. Figure 3.3 displays the dispersion around the median as a function of the Kepler magnitude ($Kep\ mag$), for the different spectral classes. As also shown in other studies (Gilliland 2008), all giant classes show many variable stars with large dispersion, with an increase in variability for cooler stars. M giants are very few in the sample but they show the largest dispersion in the giants' group. K giants mainly have dispersions in the range of 0.1-10$mmag$ and G giants between 0.1-1$mmag$.

As expected and previously demonstrated in other studies (Gilliland 2008, Jenkins et al. 2010), the dwarf stars as a whole are quieter than the giants. Moreover, all spectral classes show a

bimodal distribution. The first group of stars is clustered in the low-dispersion region of the plot and follows pretty well the trend of the data's median uncertainty (solid gray line). The second group possesses instead a higher dispersion. The lower-dispersion region of the plot is composed of more stars than the higher-dispersion one. Light curves of the stars with lower dispersion are generally dominated by "stochastic" white noise, whilst those belonging to the higher-dispersion group display periodic variability at one or more main frequency values.

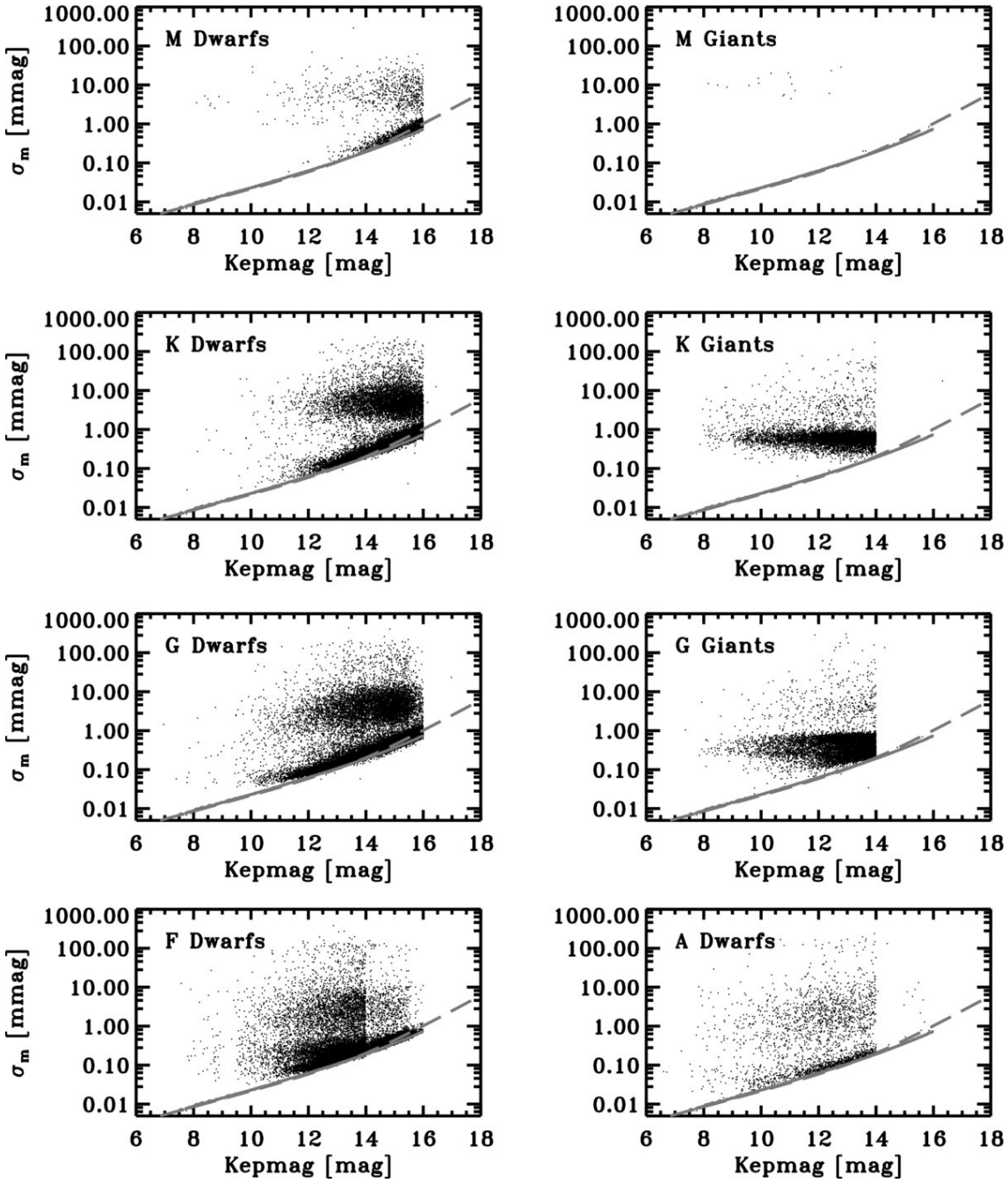


Figure 3.3: Figure 6 from Ciardi et al. 2011. Photometric dispersion around the median for the different spectral classes as a function of the Kepler magnitude. The dashed gray curve is the uncertainty upper limit empirically estimated by Jenkins et al. 2010, and the solid gray curve is the median uncertainty value of the Kepler data.

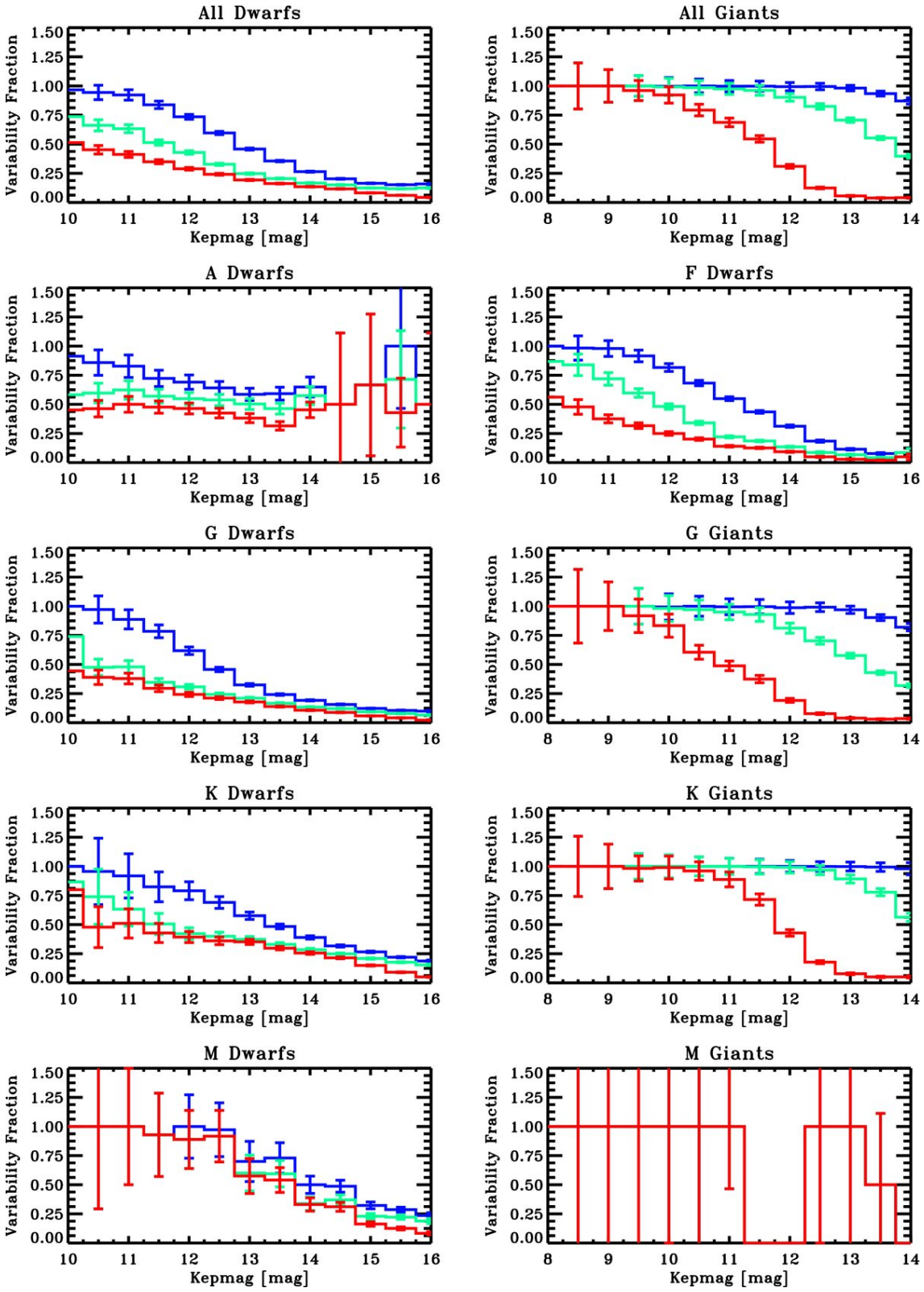


Figure 3.4: Figure 10 from Ciardi et al. 2011. Variability fractions as a function of the Kepler magnitude. The blue curves represent the fractions of stars with $\chi^2_v > 2$, the green curves the fractions of stars with $\chi^2_v > 10$, and the red curves the fractions of stars with $\chi^2_v > 100$.

The reduced chi-squared takes into account the uncertainty in the data. Fig. 3.4 displays the variability fractions of the stars, depending on the reduced chi-squared values of their light curves. These are shown as a function of the Kepler magnitude, for the different spectral classes.

The variability fractions are dependent on the magnitude of the stars because the instrumental photometric precision decreases as the star becomes dimmer. This is due to the fact that the noise is larger for dimmer stars. So, as the stars become fainter, the variability fractions decrease, which is related to the degraded instrumental precision. On the contrary, for all spectral types, the variability fraction of bright stars ($Kep\ mag \lesssim 11$) with $\chi^2_v > 2$ approach unity, as the uncertainty in the data is lower, and it is easier to detect variability above it.

Nearly all the giants are variable, particularly 94%, 99%, and 100% of them for G, K, and M giants, respectively. Among the dwarfs, the G class is the least variable, with around 80% of the sample being stable. Even when the Kepler magnitude is restricted to only include the brighter stars ($Kep\ mag < 14$), 70% of the G dwarfs are estimated to be stable. K and F dwarfs are comparable, with around half of them being stable with the restricted magnitude $Kep\ mag < 14$. Finally, M and A dwarfs are more variable, with only around 30% of them being stable, for the same restricted magnitude.

So, the most important conclusion from the study by Ciardi et al. 2011 is that G dwarfs are are most stable spectral class in the sample, with around 80% of them being considered stable. Particularly, they were found to not vary on top of the instantaneous noise in the light curve. G dwarfs spend about 90% of their lifetimes on the main sequence of the HR diagram, so they are quite stable, and they possess an effective temperature between around $5,000$ and $6,000K$. Other classes of stars that are not very variable are F and K dwarfs, with around 50% of them being considered stable.

In this Section, the most important results from the stellar variability study from Ciardi et al. 2011 were summarised, which showed the utility of Kepler data in this type of application. A mission that provides similar data is TESS. Kepler and TESS light curves are similar and can be exploited for the same purposes. TESS is also more relevant, as it has the advantages of being more recent and providing a much larger sky coverage than Kepler. So, its characteristics make it a relevant mission for stellar variability analysis. TESS is presented in the next Section.

3.6. TESS

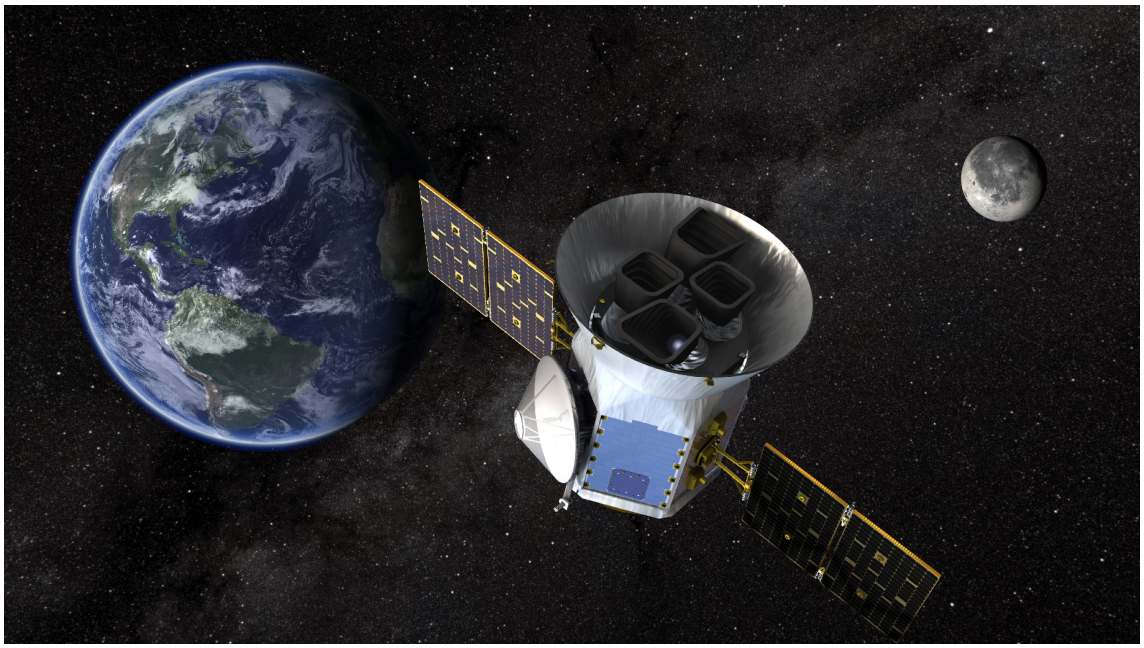


Figure 3.5: Illustration of the TESS spacecraft in orbit around the Earth. *Credit: NASA's Goddard Space Flight Center.*

The full description of the TESS spacecraft and its specifications can be found in Ricker et al. 2010. The information in this section is based on Ricker et al. 2010 unless stated otherwise.

TESS is an all-sky survey mission performing wide-field photometry with the goal of discovering new exoplanets around more than 200,000 main-sequence stars through transits. TESS's target planets are mainly mini-Neptunes and super-Earths. It was launched in April 2018 and orbits a highly elliptical orbit with a period of 13.7 days around the Earth. In July 2020, TESS completed its two-year-long primary mission, which covered about 75% of the sky, an area about 400 times bigger than the one covered by Kepler. The extended mission is still ongoing, collecting new observational data and revisiting some regions of the sky, especially towards the ecliptic poles, which have the longest observing intervals. At the time of writing, according to the NASA Exoplanets Archive[1], TESS has identified thousands of exoplanet candidates among which 350 were confirmed to be planets, and around 1,800 have been proven to be false positives. The confirmation of the candidates is still in progress, counting more than 4,000 yet to be confirmed.

The TESS payload consists of a data handling unit and a single plate on which four identical cameras are mounted at different angles, as shown in Fig. 3.6 so that their fields of view (FOV) are aligned to create a combined rectangular FOV of $24 \times 96°$ or $2300(°)^2$ in the sky.

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

Figure 3.6: Illustration of the orientations of TESS payload's four cameras and mounting plate. *Credit: Ricker et al. 2010.*

The TESS scanning strategy is shown in Fig. 3.7. Each ecliptic hemisphere is covered by 13 sectors, with an instantaneous combined FOV covering the celestial sphere from the ecliptic pole to an ecliptic latitude of 6°. Some regions have partial overlapping between sectors and others are not covered. Each sector has an observation time length of two TESS orbits, i.e. about 27.4 days. After the end of the observing time of one sector, the FOV is shifted eastward to observe the following one. It takes one year to observe a full hemisphere.

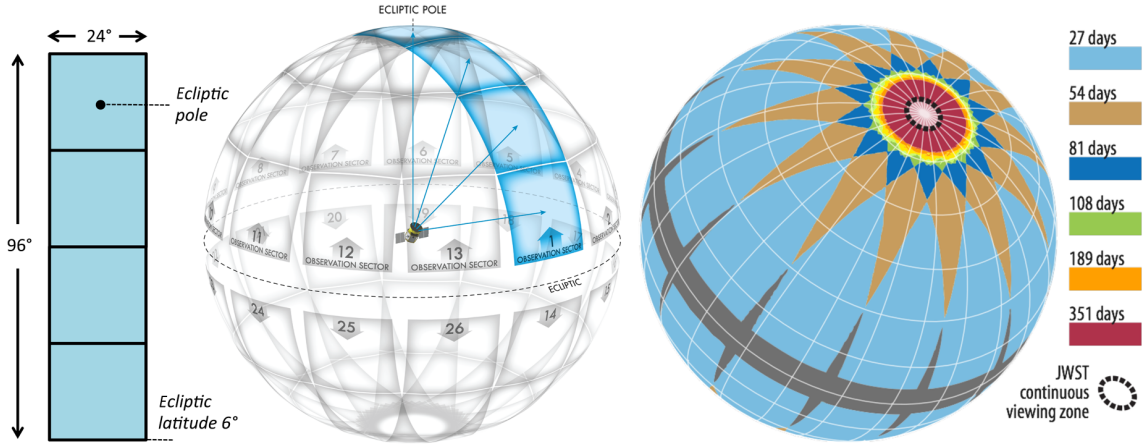


Figure 3.7: From left to right: The combined FOV of the four TESS cameras; the division of the celestial sphere into 13 sectors per ecliptic hemisphere; and the duration of the observations on the celestial sphere. *Credit: Ricker et al. 2010.*

In celestial coordinates, the performed observation sectors are shown in Fig. 3.8. The region not covered by observations is the galactic center.

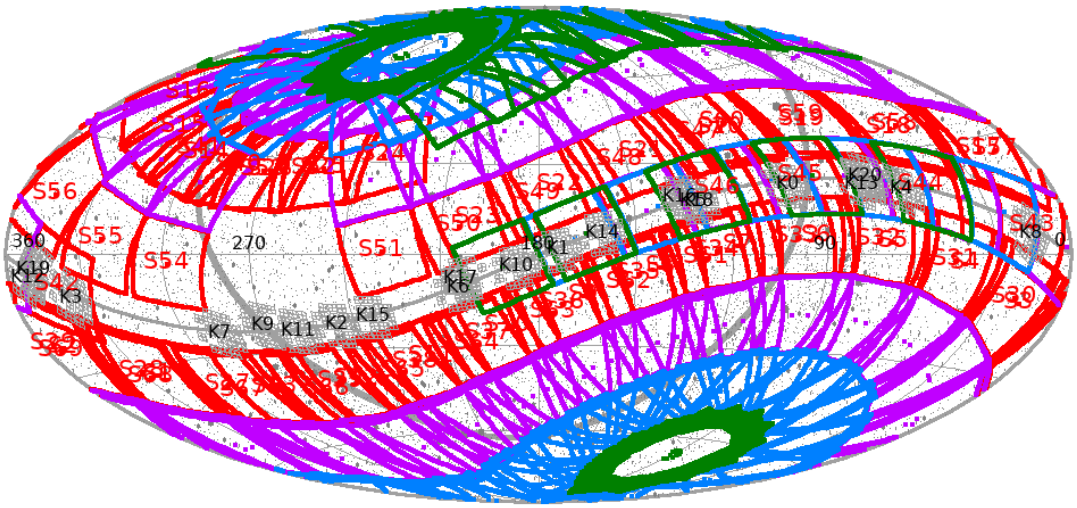


Figure 3.8: TESS sectors observed so far in ICRS.[a]
[a] https://tess.mit.edu/observations/

Since transits usually last a few hours and the ingress and egress would last several minutes for the type of planets targeted by TESS, the desired cadence of the observations is a few minutes or less. Flux measurements of stellar targets of interest are recorded with a cadence of 2 minutes or 20 seconds to meet the desired requirement and allow asteroseismology studies, i.e. the study of stellar oscillations through their photometric variations, as a tool to define stellar mass, radius, and internal dynamical processes. A Full Frame Image (FFI), instead, has a cadence of 30 minutes. FFIs are used to extend the search for planets to any star in the FOV that is bright enough, even if this was not previously selected as a TESS Object of Interest (TOI) or does not have a TESS Input Catalog (TIC, Stassun et al. 2019) ID. Moreover, they enable the observation of other types of scientific targets including supernovae, Solar System asteroids, binaries, and others.

4

Ariel

The Atmospheric Remote-sensing Infrared Exoplanet Large-survey (Ariel, Tinetti et al. 2021) is the fourth M-class mission in ESA's Cosmic Vision programme, to be launched in 2029. A drawing of the Ariel spacecraft is shown in Fig. 4.1.

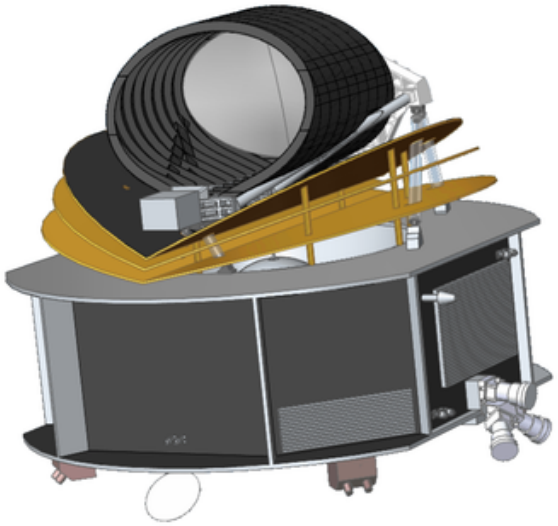


Figure 4.1: Illustration of the Ariel spacecraft. *Credit: ESA.*

In this Chapter, an overview of the Ariel mission is given, first focusing on its goals, observation strategy and instruments (Section 4.1, mainly based on Tinetti et al. 2021), continuing with considerations on its photometric performance estimated with the Ariel radiometric model ArielRad and its flux stability requirement (Section 4.2, mainly based on Mugnai et al. 2020). At the end of the Chapter, a description of the flux calibration efforts that have been carried out so far is given (Section 4.3).

4.1. Overview

The scientific questions that Ariel aims to answer comprise the understanding of the physical processes determining planetary atmospheres, their composition, and the formation and evolution of planetary systems. The mission will determine the chemical composition and thermal structure of a large number of atmospheres, by surveying planets with a wide range of masses and radii with different properties, conditions, and host star types. By studying this, Ariel can find evidence for planetary migration, determine the energy budgets of planetary atmospheres (including albedo, temperature structure, weather and temporal variations), constrain the chemical processes that play a role in the atmospheres (thermochemistry, photochemistry, transport), investigate the interactions between stars and planets, and constrain properties of clouds in exoplanets and their distribution. The observation strategy consists of observing around 1,000 exoplanets, ranging from gas giants to rocky planets orbiting stars from A to M spectral types in the warm and habitable zones. The target selection is currently ongoing (Edwards and Tinetti 2022) and will also take into consideration inputs from the scientific community. The survey will focus on relatively close systems with bright stars and will have a large sky coverage. Observations will be performed in the IR, between 0.5-7.8μm, through transit, eclipse, and phase-curve spectroscopy.

Although Ariel's targets are still under selection, the observational strategy has already been defined and consists of four tiers. First, all exoplanets will be observed for a basic initial characterisation of their atmospheres with a low spectral resolution. Around half of them will then be observed again in tier 2 for a more in-depth study of their atmospheres and the possibility of improving the spectral resolution and SNR. The third tier will be dedicated to the best targets to be observed multiple times, to track temporal variations in chemistry, weather, and cloud coverage. The fourth and final tier will include phase-curve observations and all other observations that were not possible in the previous tiers including targets of special interest, for example, smaller planets that might host secondary atmospheres (atmospheres that form from internal and surface processes or accumulation of material rather than by accretion of gaseous matter from the accretion disc of the host star). With this approach, the mission will address both population studies and detailed analysis of selected planets.

Ariel's telescope is an off-axis Cassegrain with an elliptical primary mirror with a diameter of $1.10 \times 0.73m$. Ariel's payload will comprise the following instruments:

- Ariel InfraRed Spectrometer (AIRS) - scientific instrument providing low-resolution spectroscopy in the two IR channels, Channel 0 (1.95-3.90μm band) and Channel 1 (3.90-7.80μm).
- Fine Guidance System (FGS), Visible Photometer (VISPhot) and Near-IR Spectrometer (NIRSpec) - it ensures the centering, focusing, and guiding of the spacecraft. Additionally, it ensures high-precision photometry in the visible wavelength band and also low-resolution spectrometry in the NIR. So, the functions of the FGS are combined with scientific capabilities. Observations of the host stars and their light are used to determine the pointing fluctuations and later attenuate them by stabilising the spacecraft. There are four spectral bands needed to meet the guiding and scientific goals, FGS1 (0.60-0.80μm), FGS2 (0.80-1.10μm), VISPhot (0.50-0.60μm), and NIRSpec (0.60-0.95μm). These are provided by using a series of dichroic mirrors in the interior of the FGS's optical module.

Ariel will orbit the Lagrange point L_2 of the Earth-Sun system, to provide high thermal stability and a large field for the observations.

4.2. Photometric performance and requirement

To reach its objectives, Ariel must be able to detect minute features in the spectra of exoplanetary transits, so a good instrumental photometric stability in the time scales of the transits (up to 10 hours, Pearson et al. 2022) is crucial for the success of the mission (Pascale et al. 2018). It is desired that Ariel's observations are photon noise-limited, which means that the dominant noise component in the observations should be the photon noise, and any other noise contributions should be removed or minimised as much as possible.

To estimate the photometric performance of Ariel, the software ArielRad is used (Mugnai et al. 2020). ArielRad is a radiometric simulator of the Ariel payload which simulates Ariel observations by implementing its instrument design, photometric uncertainties, and systematics both of instrumental and physical origin. The simulations are aimed at studying the effects of the noise processes in Ariel's observations and their correlation in space and time, in order to evaluate the photometric and spectroscopic performance of Ariel and demonstrate it complies with the requirements.

The noise components that are taken into account and simulated by ArielRad are static noise sources including readout noise, gain noise, dark current, photon noise, zodiacal background, instrument thermal emission, and payload noise floor. Finally, noise coming from the jitter of the line of sight is not simulated but manually introduced in the model. Figure 4.2 shows the noise contributions and total noise budget for two typical Ariel targets, a faint and a bright one.

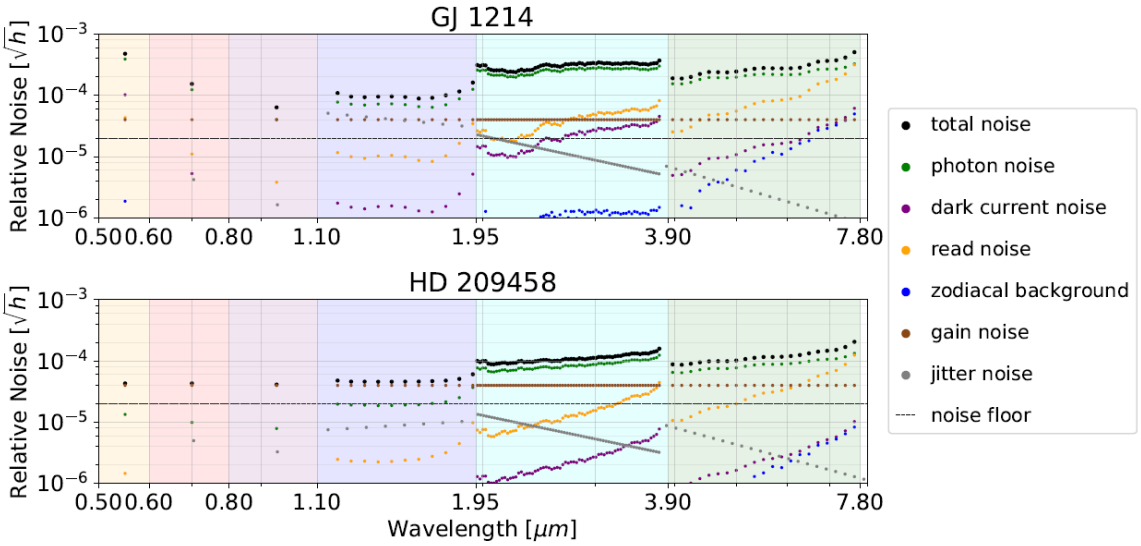


Figure 4.2: ArielRad noise budget at 1 hour integration time for a typical faint Ariel target (GJ 1214, in the upper panel) and a typical bright Ariel target (HD 209458, in the lower panel). *Credit: Mugnai et al. 2020.*

This shows that uncertainties in Ariel's observations are dominated by photon statistics, since the total noise (black line in Fig. 4.2) is very close to the photon noise (green line in the same figure), for both typical faint and bright targets.

However, these simulations do not account for other non-static noise sources. In particular, an important component is the noise coming from the stellar activity, which will play a role in real

Ariel observations and should be removed. Moreover, other unwanted instrumental effects and non-linearities should also be modelled and removed during Ariel's in-flight measurements, and Ariel's instrumental capabilities should be monitored in time. In other words, it is pivotal to make sure that any temporal variation in the calibrated signal must be attributable to changes in the signal coming from the exoplanet, rather than being the result of the stellar signal or the instrument response.

To make all of this possible, Ariel needs in-flight flux calibration. This was explained in detail in Sections 3.2 and 3.4, in astronomy, it is a general practice to observe stellar calibrators as an external flux reference for this purpose (Deustua et al. 2012), and Ariel will also use this method (Pearson et al. 2022).

The current definition of Ariel's flux stability requirement is a maximum flux dispersion of $100ppm$ (3σ) over a nominal 6-hour-long integration time (Petralia, Micela, et al. 2017):

$$FSR_{3\sigma}(6h) = 100ppm \tag{4.1}$$

So, it is necessary to create a stellar calibrator catalogue with stable stars following this stability requirement to perform Ariel's flux calibration.

In general, it is good practice to always perform new studies to define which stars are suitable for a new mission following its specific requirements and characteristics. Although it is good to use stars that were already employed in previous missions as a starting point for the analysis, one should also keep in mind the advancements in technology: Stars that seemed stable with previous instruments might not look stable anymore. This is the case for Spitzer, which made use of a set of A dwarf stars that seemed stable but now look too variable to more precise instruments (Mullally et al. 2022). The stellar calibrators usually employed from the ground are also not suitable, as they are typically stable only to the order of 1%. Modern space-based photometric missions can reach the required accuracy after instrumental effects are removed but are generally not suitable for Ariel's purposes. First, Ariel requires stellar calibrators distributed over the whole sky since it will observe almost the entirety of it, and it is desired that at least one calibrator is visible at any time (Petralia, Micela, et al. 2017). Stars used by CoRoT and Kepler would reach the required stability after systematics are removed but only cover a limited portion of the sky. Second, Ariel can only observe stars that are quite bright, and most of the calibrators used by HST and JWST, although they could also reach the required stability levels, are generally too faint.

4.3. First Ariel flux calibration efforts

4.3.1. EChO list (2013)

The first Ariel calibrator candidates list dates back to 2013 in the first stages of the Ariel mission, which was still called Exoplanet Characterisation Observatory (EChO, Tinetti et al. 2012). The list was defined by J.P. Beaulieu and A. Moneti (private communication). From now on throughout the report, this list will be called "EChO list".

This list has been drawn up starting from the Two Micron All Sky Survey (2MASS, Skrutskie et al. 2006). 537 stars distributed over the sky were added to the list by cutting at magnitude

values in order to get G, early K, and late F dwarf stars, which resulted in being the quietest stellar classes from the study by Ciardi et al. 2011. No other criteria were set for the EChO list, and the variability of the candidates was not analysed any further. Following the findings of Ciardi et al. 2011, it was thought that most G dwarfs (around 80%) would be stable, so the selection of the actual calibrators was not performed all the way until the end. It was just assumed that enough suitable calibrators would be found among those spectral types. However, the study by Ciardi et al. 2011 was restricted to the small coverage area of the Kepler mission and no further investigation into the reliability of those results was performed. So a new study should address again the problem using data with a larger sky coverage to understand if the percentage of stable stars is the same. Moreover, even if the percentage of stable stars is high, a study on the candidates should be performed to discard the unsuitable ones.

If the selection of the specific calibrators was not a priority a decade ago, now the interest in them is rising. This is because the scientific targets are in the definition phase (Edwards and Tinetti 2022), and there is the desire to schedule their observations in the most efficient way possible to maximise the scientific output (Pearson et al. 2022). The observations of calibrators must also be included in the schedule, so information on which stars are going to be calibrators and their positions in the sky is of high interest.

The available information for the stars in the EChO list only includes an arbitrary identifier, the Right Ascension (RA) and Declination (Dec) in epoch J2000, and the magnitudes in the J, H, and K_s bands. Note that the 2MASS ID is not present, nor any other ID referable to any other space mission or survey, but only the stellar position in the sky. So, it is not completely straightforward to identify which stars are on this list. More detailed information on the available columns of this catalogue and their units can be found in the Appendix Section A.1.

4.3.2. Catalogue 1 (2022)

After many years, it was necessary to revise the filtering parameters to define a new, refined candidate sample from scratch. In 2022 this updated candidate list was drawn up by J.P. Beaulieu and L. Bernard from the Institut d'Astrophysique de Paris (IAP) (private communication). From now on, this list will be called "Catalogue 1".

Instead of 2MASS, the new Gaia Data Release 3 (Gaia DR3, Vallenari et al. 2022) from the Gaia mission (Prusti et al. 2016) was used by filtering stars with some desired parameters. Gaia DR3, released in June 2022, contains up-to-date information compared to 2MASS (from 2006), and on many more astronomical sources (more than a billion in total). Following the same arguments of the flux stability of K, G, and F dwarfs, Catalogue 1 was created to only include G dwarf stars and, specifically, stars that have characteristics similar to the Sun. The Sun is the most well-known and studied G dwarf star because it is the star of our Solar System. It has a photospheric temperature of around $5,800K$ and surface gravity of $log(g) \approx 4.5 log(cm\ s^{-1})$. Therefore, the filtering criteria in Gaia DR3 are a temperature between $5,000$ and $6,000K$ and surface gravity between 4.0 and $4.5 log(cm\ s^{-1})$. Additionally, the maximum distance between the Earth and these stars is set to $50pc$. This is because when retrieving objects from Gaia DR3, only the first 2,000 objects matching the querying criteria are shown, so, it was necessary to reduce the number of compatibilities in some way. Moreover, by using stars that are not too far away, the problem of interstellar reddening is mitigated (Deustua et al. 2012). Catalogue 1, created in this way, is finally composed of 1907 stars. Thanks to Gaia, this list contains much more information on the stars than the first EChO list. More detailed information on

the available columns of this catalogue and their units can be found in the Appendix Section A.2.

5

Methodology

In Chapter 2 the research objectives and questions of this study were accurately formulated. Afterwards, Chapter 3 and 4 gave a more extensive overview of the important background knowledge to understand this project. This included the transit spectroscopy technique, flux calibration, stellar calibrators, and the Ariel mission with its past calibration efforts.

In this Chapter, the data and different tools employed to reach the objectives of this study are presented, as well as the selection process to build Ariel's stellar calibrator catalogue. First, the two astronomical archives used in this study are described (Section 5.1), followed by a description of the TESS data used for the analysis of the photometric variability of the candidate calibrators (Section 5.2). The process of unification of the two preliminary candidate lists previously described in Section 4.3 is then presented (Section 5.3). The mathematical algorithm used to investigate the power spectrum of the light curves, the Lomb-Scargle periodogram, is presented afterwards (Section 5.4), followed by useful statistical quantities to perform the selection of the calibrators (Section 5.5). The selection criteria to define the Ariel stellar calibrators are then presented (Section 5.6). The Chapter ends with a recap of the workflow (Section 5.7).

5.1. SIMBAD and MAST

Two tools that are widely used in the astronomic community are "Set of Identifications, Measurements and Bibliography for Astronomical Data" (SIMBAD) and "The Barbara A. Mikulski Archive for Space Telescopes" (MAST). In this study, both SIMBAD and MAST are employed particularly in the first part of the project concerning the unification of the first candidate calibrator lists. MAST is also used for the retrieval of observational data of interest.

SIMBAD[1] (Wenger et al. 2000) is operated at the Centre de Données Astronomiques de Strasbourg, France. It is an astronomical bibliographic database containing information on millions of astronomical objects of interest, excluding Solar System bodies. It is a useful tool because it provides data and measurements, cross-identifications of objects between different missions, and a wide range of bibliographies, documentation, and papers about astronomical objects. SIMBAD can be queried by object name, coordinates, or various other criteria, either for single objects or a list containing multiple ones. It is also possible to query SIMBAD directly from a Python script by using the package `astroquery.simbad` from `astropy`[2] (Price-Whelan et al. 2022).

MAST is an online astronomical data archive to search multiple collections of astronomical datasets and files, including observations and catalogues. Some MAST-supported missions include JWST, HST, TESS, Kepler, and others. It is developed by the Space Telescope Science Institute (STScI). Its goal is to maximise the scientific accessibility of astronomical data and the reproducibility of results. MAST includes many different search tools, learning resources including sample notebooks and documentation, and catalogues. An efficient way to search multiple astronomical collections in one place is by querying the MAST Portal[3]. The MAST Portal stores images, spectra, time series, calibration, and other supporting files in a single place. It is simple to query, as it can be done manually online by searching for a single object's identifier, position in the sky, or other criteria, or by uploading a target list containing multiple objects. Moreover, it is possible to select a specific collection to query or look for observations in any MAST-featured mission. Any target list can also be cross-matched with other astronomical catalogues to get observations or information on the targets from a specific mission. Finally, the MAST Portal can be also queried directly from a Python script by using the package `astroquery.mast` from `astropy`.

5.2. TESS data

The characteristics of the TESS mission and its observation strategy were explained in Section 3.6. Here, the motivation for using TESS for the flux variability study of Ariel calibrator candidates is presented. Moreover, the characteristics of its light curve data are shown.

5.2.1. Motivation

To analyse the stellar flux variability to build Ariel's stellar calibrator catalogue, data on their photometry is needed. Since stars spread all around the sky are desired, a space mission with a wide coverage must be used. Moreover, the more recent the data is, the more reliable the information on the current and near-future stellar behavior. TESS is suited for the purpose of this study since it has a wide sky coverage and provides recent observations, being active since 2018. With Gaia, a star is seen as a point source integrated over a long time, and a

[1] https://simbad.cds.unistra.fr/simbad
[2] https://www.astropy.org
[3] https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html

high-precision absolute flux calibration is reached. Differently, TESS is a photometric mission looking for exoplanet transits, so it is more concerned with the relative flux variation and flux stability, like Ariel. TESS retrieves photometric time series of stars, with the goal of hunting for planets. This means that TESS is surveying hundreds of thousands of stars, and the photometric time series, called light curves (LCs), can be exploited for this study.

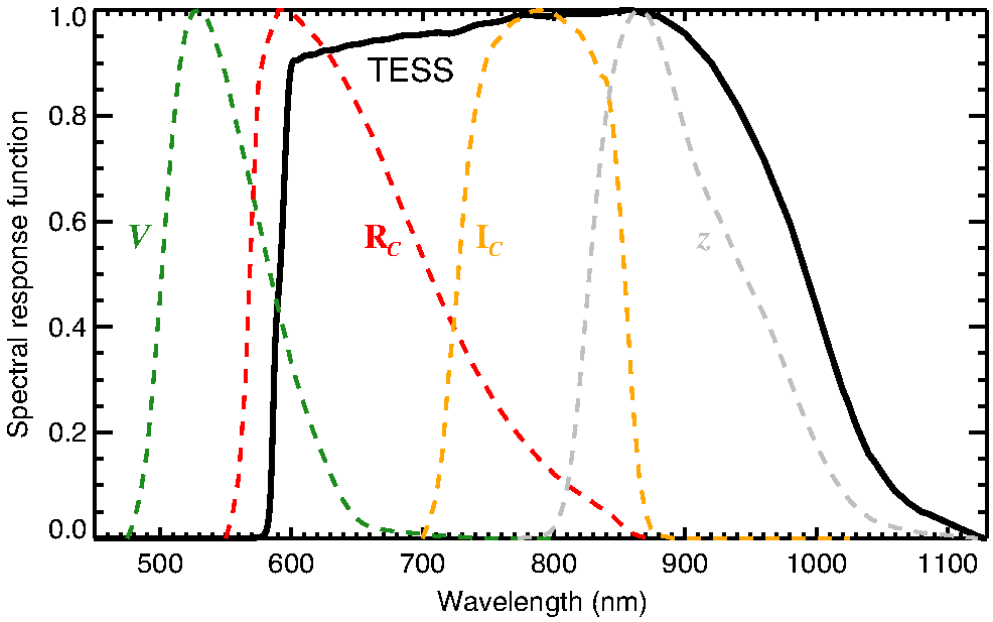


Figure 5.1: The TESS spectral response function (black line) compared to the Johnson–Cousins V, R_C, and I_C, and the Sloan Digital Sky Survey z filter curves. Their maximum value is scaled to unity. *Credit: Ricker et al. 2010.*

Figure 5.1 shows that the spectral response function of TESS covers the wavelength range between around 0.6 and 1.5μm. The goal of this study is to check the photometric stability of stellar calibrator candidates for the Ariel mission, which will observe in the IR part of the spectrum (0.5-7.8μm). Although TESS observes in a narrower wavelength window than Ariel, which is shifted more towards the optical part of the spectrum, it is suitable to study the stability of stars in the desired IR range. In fact, stellar spectra roughly look like black-body curves, and the energy of solar-type stars peaks toward visible wavelengths, due to their temperature. Since the energy in the IR will always be lower, it is possible to assume that a flux variation of a certain percentage in visible wavelengths means that there is the same variation or lower in the IR. So, TESS data, although with some noise limitations, is suitable for this study.

Fig. 5.2 shows the TESS 1σ photometric precision with respect to the magnitude of the target source, in units of *ppm* over an integration time of 1 hour. The Ariel stability requirement in the same units would become:

$$FSR_{1\sigma}(1h) = \frac{100}{3} \times \sqrt{6} \approx 81.65ppm \tag{5.1}$$

This means that TESS data is precise enough to study stability for Ariel as long as the star

is brighter than an apparent magnitude in the I_C band of around $7.5mag$. For dimmer stars, TESS observations have a total noise that is larger than the Ariel flux stability requirement.

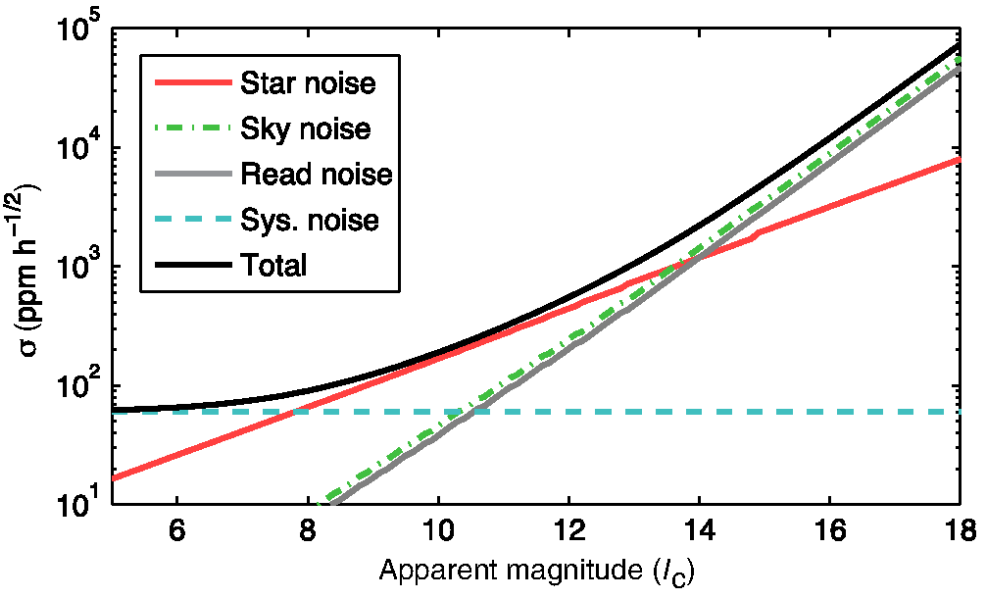


Figure 5.2: TESS expected 1σ photometric precision with respect to the stellar apparent magnitude in the I_C band on hourly timescales. *Credit: Ricker et al. 2010.*

5.2.2. Light curves

TESS LCs are used for the photometric analysis of stars in this study. The process to create LC files is shown in Fig. 5.3.

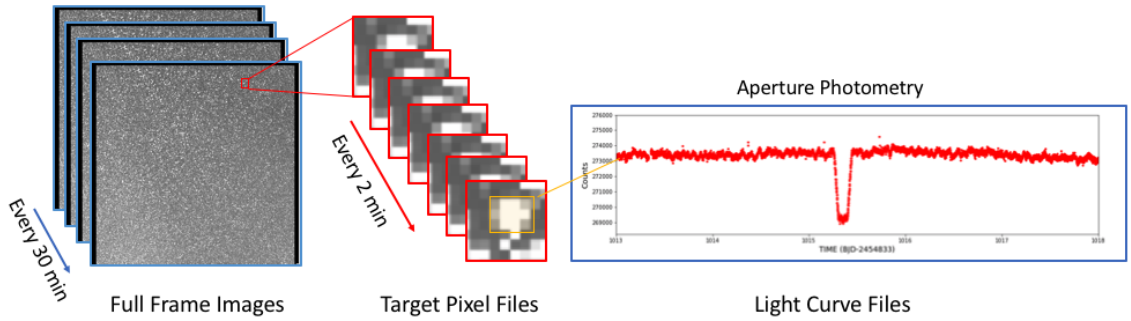


Figure 5.3: Graphical view of TESS photometric data products (*Credit: TESS Archive Manual*[a]).
[a] https://outerspace.stsci.edu/display/TESS/2.0+-+Data+Product+Overview

FFIs are recorded every 30 minutes and cover the entirety of the Charge-Coupled Device (CCD) detector. Some groups of pixels around selected targets are downloaded with a higher cadence (every two minutes) and are called Target Pixel Files (TPF). Afterwards, the flux recorded in each of these pixels is combined to obtain the total flux received from the target star. TPFs also include information on the astronomical background contribution which is removed from the flux, and on the recommended aperture for extracting the flux. A sample TPF is shown

in Fig. 5.4. Following aperture photometry extraction, flux time series are created, which are known as LC files. All files are stored in the MAST Archive in *.fits* format.

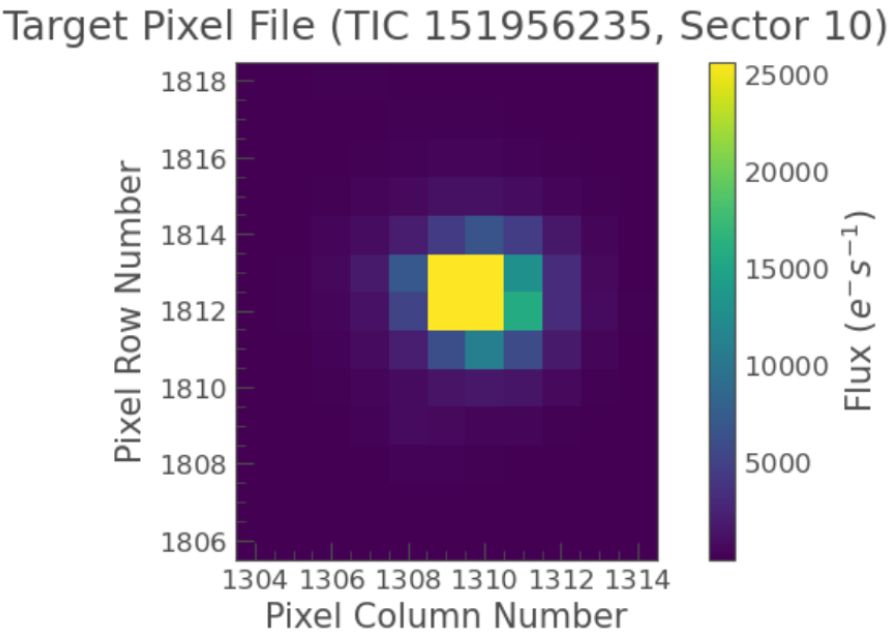


Figure 5.4: Example of a TPF, for the arbitrary target TIC 151956235, in Sector 10.

The extraction of the flux from the TPF can be made manually through Simple Aperture Photometry (SAP), which means that the values in the pixels belonging to a selected aperture are summed for each time stamp. With different shapes of the aperture mask, it is possible to filter the background or avoid contamination of nearby sources in the extracted signal. On MAST, it is possible to directly query LC files that already went through a photometric extraction process as well as through additional data cleaning steps depending on which pipeline was applied. For simplicity and consistency, this study only employs LC files that were automatically extracted with the Science Processing Operations Center (SPOC, Jenkins et al. 2016) data processing pipeline. Such files are downloaded directly from MAST. The SPOC pipeline is developed by the NASA Ames Research Center and it was chosen because it provides high-quality background subtraction and correction of obvious instrumental systematics. The flux stored in these files is called Pre-search Data Conditioned (PDC) SAP flux. It is produced starting from the SAP flux values after being nominally corrected for instrumental variations including changes in focus or pointing, radiation events on the detector, or others. This de-trending process of the time series is performed by the pipeline automatically by fitting vectors called Cotrending Basis Vectors (CBV) and subtracting them from the SAP flux data. These files represent the best estimate of the intrinsic flux variability of the target, with instrumental effects removed. Some research groups clean and de-trend the LC further. For example, Mullally et al. 2022 make use of a Savitzky-Golay filter (Savitzky and Golay 1964) with a window of a length equal to half the number of data points in the time series, to smooth the data. Figure 5.5 shows an example of the difference between the SAP and the PDC SAP flux for an arbitrary target.

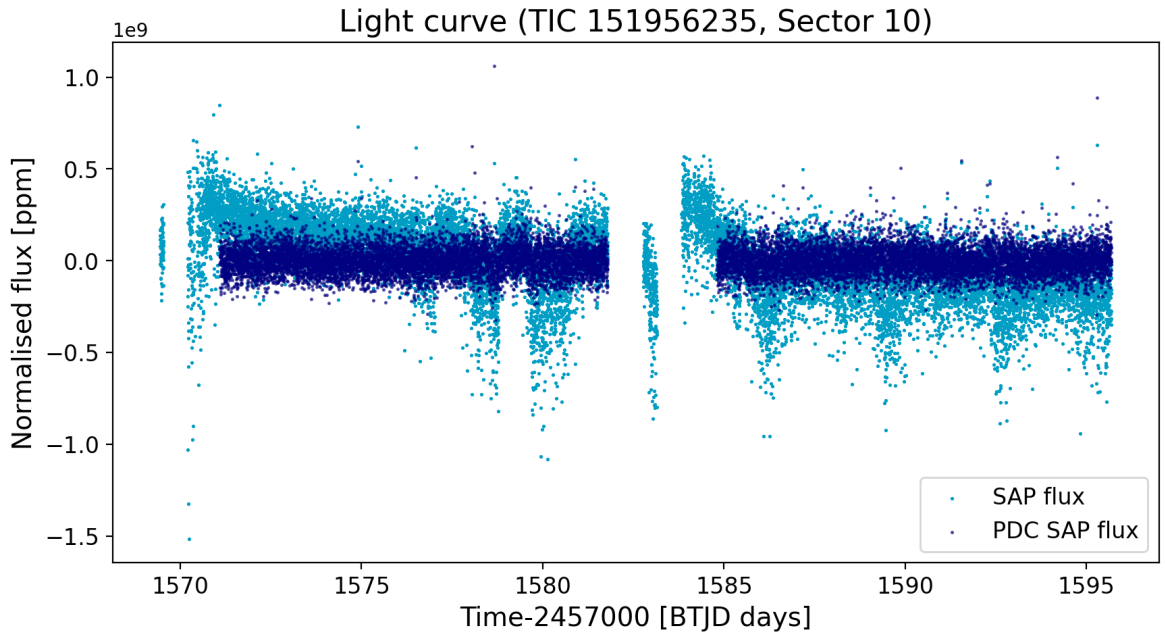


Figure 5.5: Example of the difference between the SAP flux and the PDC SAP flux, for the arbitrary target TIC 151956235, in Sector 10.

The x-axis represents the time, which is by default in units of TESS Barycentric Julian Day (BTJD), corresponding to a Julian day minus 2,457,000.0. Time defined like this is not affected by leap seconds and it is corrected to the arrival times at the barycenter of the Solar System (Tenenbaum and Jenkins 2018). The y-axis displays the normalised SAP and PDC SAP fluxes. The SAP flux clearly shows a drift in time and other errors that were found to be caused by instrumental effects and were therefore corrected by the pipeline.

The LC files processed by the SPOC pipeline can be downloaded directly from a Python script through the `lightkurve`[4] Python package (Cardoso et al. 2018). `lightkurve` offers user-friendly open-source tools to retrieve, inspect and analyse time series data from NASA's Kepler and TESS missions, and is supported by a large range of tutorials. It is possible to only retrieve data from a specific mission, TESS in the case of this study, and select the desired data cadence. For consistency, only 2-minute cadence files are used in this study. Moreover, if a target was observed in different sectors, multiple observations can be retrieved. However, for the purpose of this study, only the first one in chronological order is arbitrarily chosen for the analysis.

Following the recommendations of the TESS Archive Manual[5], the Cadence Quality Flags given by the pipeline are used to remove data points that might have been recorded during anomalous events. The meaning of the different flags is shown in Tab. 5.1. The points flagged with numbers 1, 2, 3, 4, 5, 6, 8, 10, 13, and 15 are removed as they likely possess lesser quality. The other flags are still included in the analysis as they only indicate that a correction was applied to that data point.

[4] https://docs.lightkurve.org
[5] https://outerspace.stsci.edu/display/TESS/2.0+-+Data+Product+Overview

Table 5.1: TESS Cadence Quality Flags[a].

Flag number	TESS Quality Flag description
1	Attitude tweak
2	Safe mode
3	Spacecraft in coarse point
4	Spacecraft in Earth point
5	Argabrightening event
6	Reaction wheel desaturation event
7	Cosmic ray in optimal aperture pixel
8	Manual exclude due to anomaly
9	Discontinuity corrected between cadences
10	Impulsive outlier removed before co-trending
11	Cosmic ray on collateral pixel
12	Stray light from Earth to Moon in FOV
13	Scattered light exclude
14	Planet search exclude
15	Bad calibration exclude
16	Insufficient targets for error correction exclude

[a] https://outerspace.stsci.edu/display/TESS/2.0+-+Data+Product+Overview

The unit of the PDC SAP flux is electrons per second $[e^-/s]$ at the detector, so it is an instrumental unit rather than a physical, calibrated one. To analyse the LC regardless of the brightness of the star, a normalised flux can be computed, to only focus on the relative variations:

$$F(norm) = \left(\frac{F(e^-/s)}{\mu(F(e^-/s))} - 1 \right) \tag{5.2}$$

where $F(norm)$ is the normalised flux with a value range between 0 and 1, $F(e^-/s)$ is the flux in electrons per second, and $\mu(F(e^-/s))$ is its mean. The normalised flux is usually employed in percentage, parts per million, or milli-magnitude. To convert the flux into milli-magnitudes ($F(mmag)$), the apparent magnitude equation can be used:

$$F(mmag) = -2.5 log_{10}(1 - F(norm)) \times 10^3 \tag{5.3}$$

Or, simply, the different units can be related as follows:

$$0.001 = 0.1\% = 1000ppm \approx 1.0863mmag \tag{5.4}$$

5.3. Calibrator candidate lists unification

As explained in Section 4.3, at the beginning of this project two calibrator candidate lists are available. The first one, the EChO list, was created in 2013 from 2MASS with coarse criteria and comprises 537 stars. The second one, Catalogue 1, includes 1907 stars and was refined in 2022 through the use of the modern Gaia DR3. To give continuity to the study of Ariel calibrators, the unification of the newest Catalogue 1 with the original EChO list must be performed. The resulting merged catalogue will represent the starting stellar sample for the analysis of this project. So, it is needed to verify if the stars in the EChO list are also present in Catalogue 1. As mentioned, the EChO list only contains an arbitrary name, the position of the star in the sky defined by RA and Dec in epoch J2000, and the JHK_s magnitudes. Since there is no identifier from an actual mission, it is not straightforward to match stars between lists. So, the cross-matching must be carried out by comparing the stars' positions in the sky.

First, the stars in Catalogue 1 that are not found when queried on SIMBAD and MAST are removed, since their LC data is not present in both archives. There are eight stars whose Gaia DR3 identifier could not be found.

To perform the matching, the stellar coordinates of the lists are converted into the same units. In the EChO list, RA is in time units, in the form $[h : min : s]$, while in Catalogue 1, RA is in decimal degrees [°]. So, the EChO list RA is converted into decimal degrees. Dec in the EChO list is in the sexagesimal form $[° : ′ : ″]$, while in Catalogue 1 it is in decimal degrees. Again, the EChO list Dec is converted into decimal degrees. Subsequently, the coordinates need to be defined at the same epoch. In the EChO list, following 2MASS, the epoch is J2000, whilst in Catalogue 1, following Gaia DR3, the epoch is J2016.0. For convenience, SIMBAD is used to retrieve the coordinates of the stars in Catalogue 1 in epoch J2000. J2000 is preferred because it is the "standard" epoch which is historically more used. The Gaia DR3 ID is queried in SIMBAD for each star of Catalogue 1 and their coordinates in epoch J2000 are stored. The positions in the sky of the stars of the two catalogues are shown in Fig. 5.6.

Note that it is not enough to change the epoch. Changing the epoch in which the coordinates are defined only accounts for the change in the definition of the reference system. However, the stars have a proper motion, which means that their position in the sky changes in time. This must be taken into consideration. The positions defined in Gaia DR3 are the ones in 2016, so the stars moved for 16 years with their proper motion with respect to their position in 2MASS (in 2000). Gaia DR3 contains information on the proper motion of the stars in both RA and Dec directions, $pmra$ and $pmdec$, respectively, and their errors. Those can be used to predict the new stellar positions in the sky. However, since most stars have small proper motions, for simplicity, in this study it is assumed that if stars belonging to different lists are positioned within $1'$ from each other, then they are the same star. Neglecting the spherical shape of the sky, if a star in Catalogue 1 has J2000 coordinates (RA_1,Dec_1) and a star in the EChO list has J2000 coordinates (RA_2,Dec_2), the distance d_{sky} that is needed to match the stars is:

$$\begin{aligned} d_{sky} &= \sqrt{(\Delta RA^2 + \Delta Dec^2)} < 1' \\ \Delta RA &= |RA_1 - RA_2| \\ \Delta Dec &= |Dec_1 - Dec_2| \end{aligned} \tag{5.5}$$

If multiple stars from one candidate list pass the distance requirement for one or more stars on the other list, only the closest star is considered to be the real match. The presence of these

duplicates is due to the fact that some stars belong to close binary or multiple stellar systems. As a result, 192 stars from the EChO list are found to also be contained in Catalogue 1.

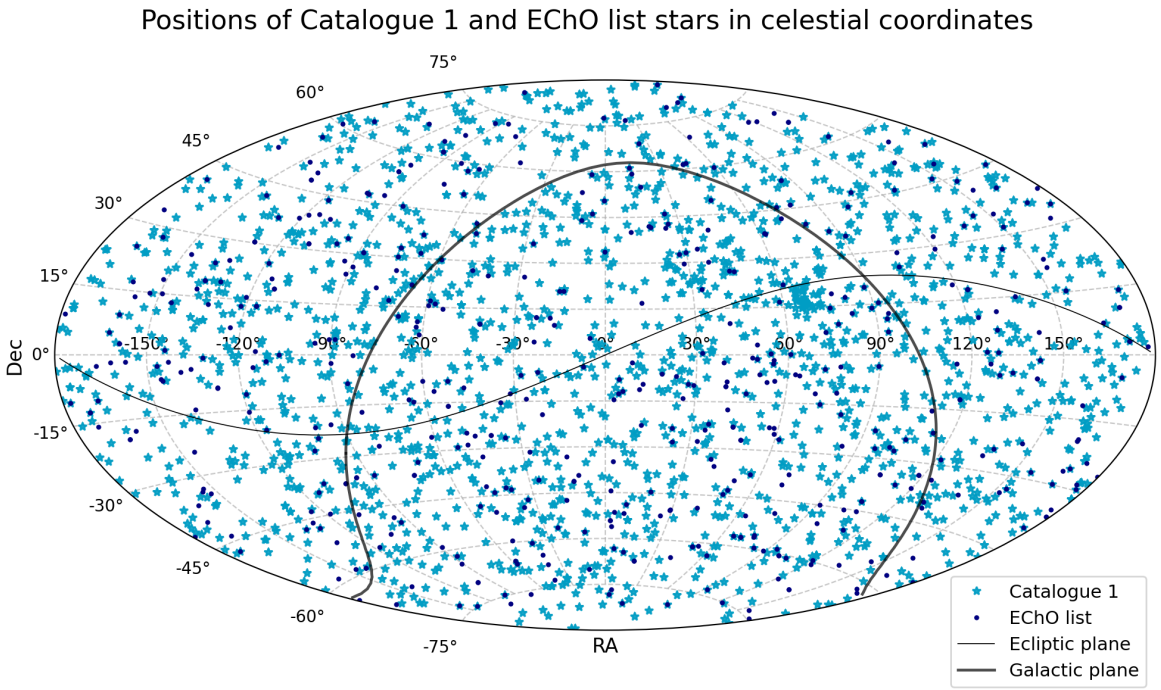


Figure 5.6: Positions in the sky of the stellar calibrator candidates from the EChO list and Catalogue 1, in International Celestial Reference System (ICRS) and epoch J2000.

On the MAST Portal, it is possible to upload a target list and cross-match it with other catalogues. So, the EChO list stellar positions are cross-matched with the TESS Catalogue and Gaia DR3, within a radius of 3″ (standard MAST value). The remaining EChO list stars are identified with stars that are not included in Catalogue 1, but their Gaia DR3 or TIC IDs or both are found and stored. Finally, 11 stars in the EChO list are removed completely since their position is not close enough to any star, from Catalogue 1 and from the MAST cross-matches, so no ID can be assigned to them.

The result of this unification leaves a list composed of 2233 stars. However, since the variability study of these stars is carried out with TESS data, only the stars that have TESS observations can be part of the sample. So, first, the stars that do not have a TIC ID are removed, even when they have a Gaia DR3 ID. These stars are only 49. Moreover, being assigned a TIC ID does not guarantee that a star was observed, but only that it is a desired target. Some stars with a TIC ID might be observed in future sectors, or their observation was planned but failed for some reason. So, the TIC ID of each star is queried on MAST to look for their data files. If the retrieved list of observations is not empty, the star was observed at least once, so its variability can be studied and it can be included in the sample. 247 stars have a TIC ID assigned to them but were not observed in any sector so far.

So, the final resulting unified candidate calibrator sample includes 1937 stars. Throughout the report, this list will be called “Catalogue 2”. More detailed information on the available columns of this catalogue and their units can be found in the Appendix Section A.3. The histograms in Fig. 5.7 show the distributions of selected stellar properties of the starting

sample Catalogue 2, and Fig. 5.8 shows the positions of such stellar candidate calibrators in the sky.

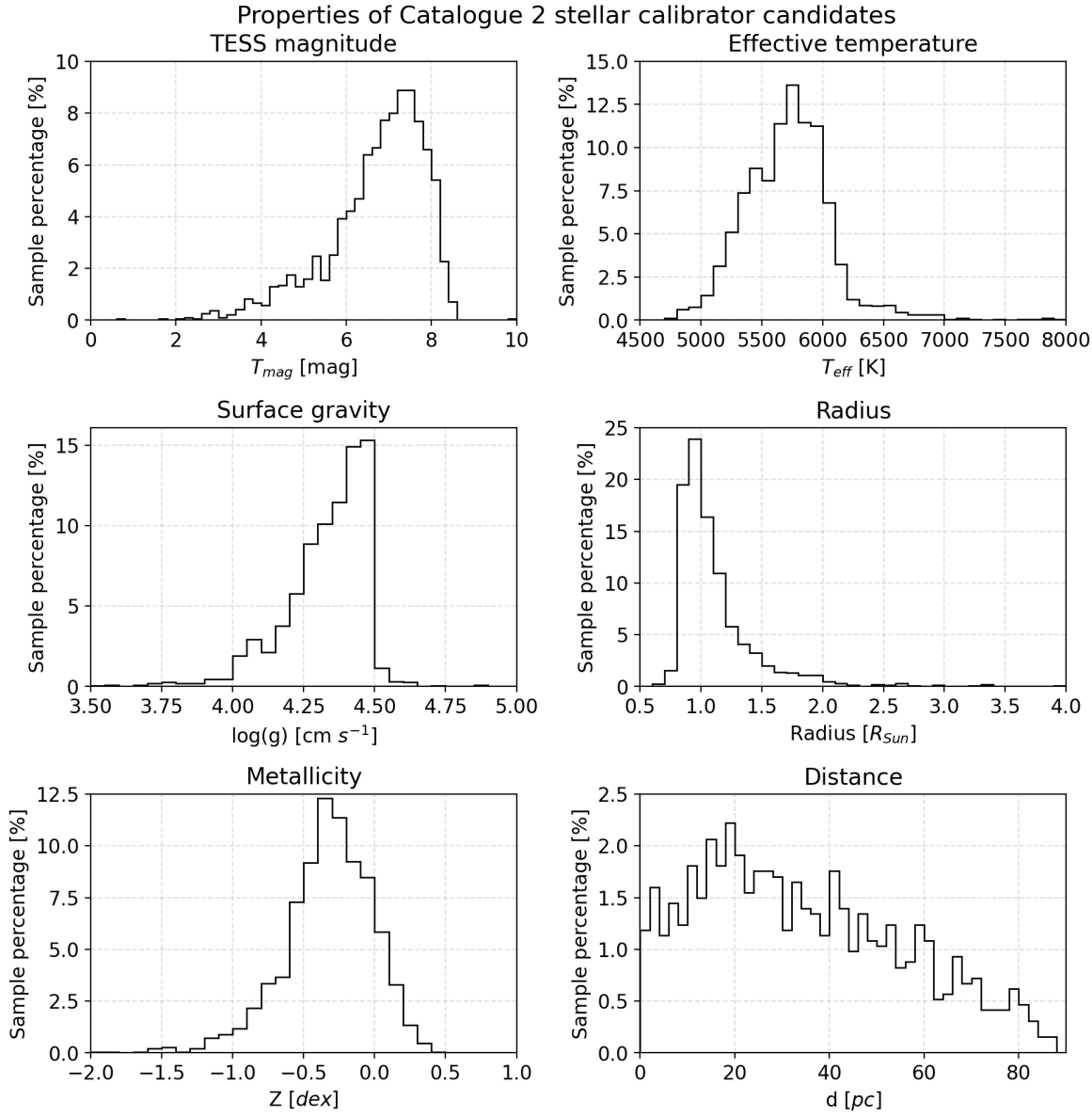


Figure 5.7: Properties of the Catalogue 2 stellar calibrator candidate sample, including TESS magnitude, effective temperature, surface gravity, radius, metallicity, and distance distributions in the sample.

These stellar properties have been selected for analysis for different reasons. Knowing the magnitude of the star is important because the LCs have different noise levels with respect to the brightness of the star. Moreover, the magnitude, together with the temperature, radius, and mass (or surface gravity), give us an indication of the stellar spectral class. The age of the stars is usually uncertain and not always available, but it is very interesting for flux variability purposes. This is because younger stars tend to be more active and variable, and they become more quiescent as they grow older. Iron abundance, or metallicity, can be used as a proxy for age, assuming that the higher the metallicity, the younger the star is. Finally, distance can help us understand if effects like absorption can play a role in the flux observations.

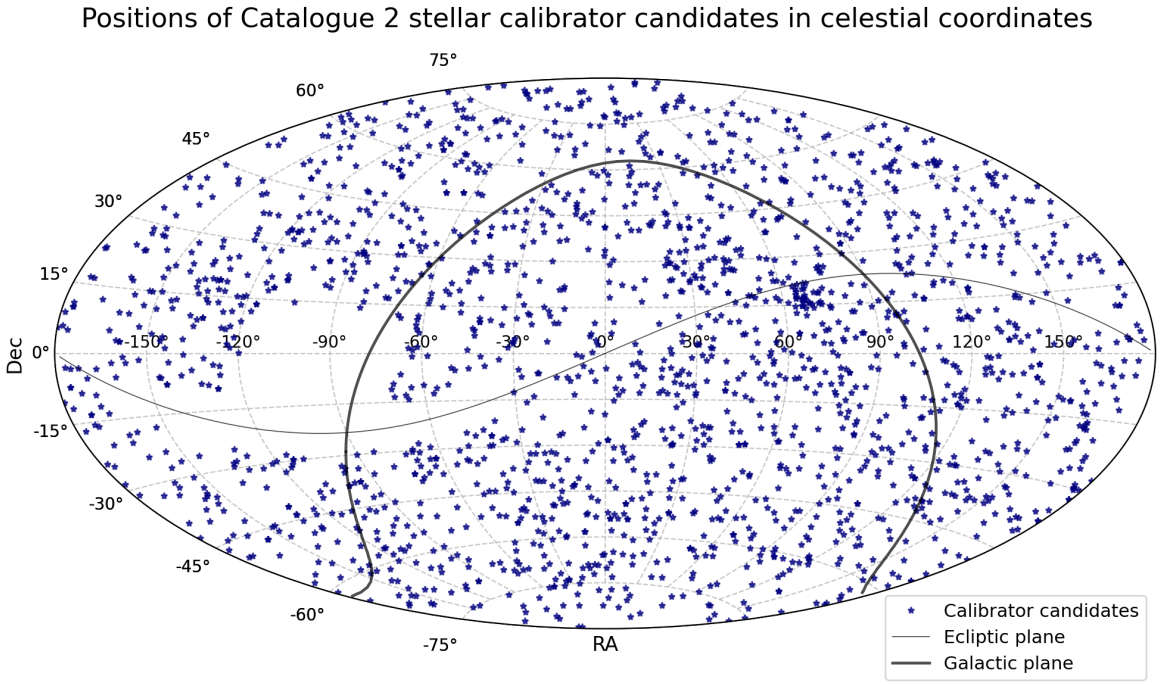


Figure 5.8: Positions in the sky of the stellar calibrator candidates from the unified Catalogue 2, in ICRS and epoch J2000.

5.4. Lomb-Scargle Periodogram

For the photometric variability analysis of the LCs, a useful mathematical tool is the Lomb-Scargle periodogram (Lomb 1976, Scargle 1982). From now on, in this report, we will refer to it as LSP, for simplicity. Note that the information in this Section is based on VanderPlas 2018 unless stated otherwise.

5.4.1. Intuitive description

The LSP is a well-known algorithm widely used by astronomers to detect periodic variability in time series data that is not evenly sampled. It can be roughly thought of as a Fourier transform for unevenly-spaced data. In fact, even if TESS LCs are nearly uniformly spaced with a 2-minute cadence, the data points can be not exactly spaced in time (there can be differences in the order of milliseconds due to reasons related to the processing of the data). Moreover, sometimes the observations have gaps of several minutes due to a variety of causes including the presence of the Earth in the field of view, re-orientation of the spacecraft, bad data that was discarded due to a cosmic ray hit, or others. Therefore, a discrete Fourier transform cannot be computed for these time series. The LSP enables the estimation of a Fourier-like power spectrum for this type of data to study the frequency content of the signal and detect potential periodic variability. Aside from the Fourier-like interpretation, the LSP can additionally be interpreted in a least-squares sense. The LSP fits simple sinusoids at different candidate frequencies to the data points and estimates how well the model fits the data by minimising the residuals. So the Fourier-like power spectrum is estimated from the goodness-of-fit of the sinusoidal model to the data at the different frequencies. The LSP approach has been adopted in different studies on stellar rotation or pulsations, like Ma et al. 2022, thanks to the fact that it models and fits variations that are of semi-sinusoidal shape to the data efficiently, also when the noise is large.

5.4.2. Mathematical derivation

To get to the formulation of the LSP, the continuous form of the Fourier transform can be used as a starting point:

$$\mathfrak{F}\{g\} = \int_{-\inf}^{\inf} g(t) e^{-2\pi i f t} dt \tag{5.6}$$

where $g(t)$ is a continuous signal, $\mathfrak{F}\{g\}$ is its continuous Fourier transform, $i = \sqrt{-1}$ is the imaginary unit, f is the frequency, and t is the time. Figure 5.9 shows the continuous Fourier transform of a sinusoidal signal.

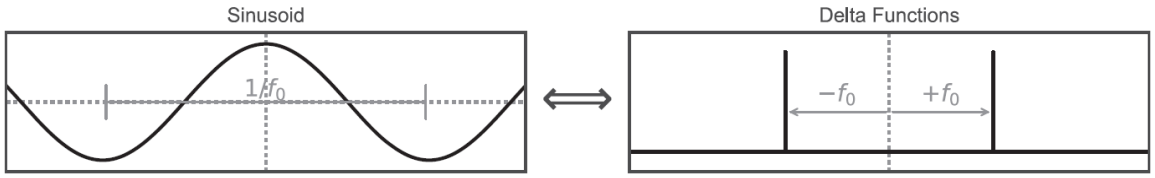


Figure 5.9: Visualization of a sample sinusoid and its Fourier transform. *Credit: VanderPlas 2018.*

It is interesting to note that a signal with a characteristic scale in the time domain will have its inverse as a characteristic scale in the frequency domain. For instance, a sinusoid with a period $1/f_0$ has a characteristic scale $\pm f_0$ in the Fourier domain.

The squared amplitude of the Fourier transform is known as power spectral density (PSD) or power spectrum P_g and is a real-valued function that describes the contribution of each frequency to the signal:

$$P_g = |\mathfrak{F}\{g\}|^2 \tag{5.7}$$

First, the consequences of working with real signals (finite and discrete) rather than with ideal ones (infinite in time and continuous) have to be taken into account:

- Effect of the signal lasting a finite time: Convolution of the continuous Fourier transform with a rectangular window. This means that each δ function is replaced by a sinc function. The wider the observing window, the narrower the spread about the peak value.
- Effect of the signal being discrete: Convolution of the continuous Fourier transform with a Dirac comb. This means that the transform becomes a long sequence of its aliases with spacing T^{-1}, where T is the total time length of the signal. So, evaluating the transform in the frequency range between 0 and T^{-1} is enough to get all available information about the signal content, as at higher frequencies the transform would just repeat in shape. Moreover, if a signal is sampled with a sampling rate f_s, it is possible to fully recover frequency information on the signal only if it is band-limited to a frequency of $f_s/2$, which is also called Nyquist frequency.

Following this, the discrete form of the Fourier transform can be written as:

$$\mathfrak{F}_d\{g\} = \sum_{n=0}^{N} g(n\Delta t)e^{-2\pi i f n\Delta t} \tag{5.8}$$

where $\mathfrak{F}_d\{g\}$ is the discrete Fourier transform, Δt is the regular spacing in time at which the signal is observed, $g(n\Delta t)$ is a continuous signal sampled at the different sampling times, and N is the total number of samples. Its power spectrum can again be computed with Eq. 5.7.

The classical periodogram $P_S(f)$ is defined by Schuster 1898, and is equivalent to Parseval's theorem, given as the power spectrum (Kjeldsen and Frandsen 1992):

$$P_S(f) = \frac{1}{N}\left|\sum_{n=1}^{N} g(n\Delta t)e^{-2\pi f n\Delta t}\right|^2 \tag{5.9}$$

This formulation only differs by a factor of N^{-1} with respect to the discrete Fourier power spectrum with uniform sampling. In fact, it is worth noting that the periodogram is an estimator of the power spectrum: The periodogram is the statistics computed from the data, while the power spectrum is the underlying continuous function of interest (Scargle 1982).

The last step to get to the LSP is accounting for non-uniform sampling of the signal. The major effect of non-equispaced sampling is the introduction of non-structured peaks in the transform.

The generalized form of the LSP is given by:

$$\begin{aligned} P_{LS}(f) =& \frac{1}{2}\left(\sum_n g(n\Delta t)cos(2\pi f[(n\Delta t)-\tau])\right)^2 / \sum_n cos^2(2\pi f[(n\Delta t)-\tau]) \\ &+ \frac{1}{2}\left(\sum_n g(n\Delta t)sin(2\pi f[(n\Delta t)-\tau])\right)^2 / \sum_n sin^2(2\pi f[(n\Delta t)-\tau]) \\ \tau =& \frac{1}{4\pi f} tan^{-1}\left(\frac{\sum_n sin(4\pi f n\Delta t)}{\sum_n cos(4\pi f n\Delta t)}\right) \end{aligned} \tag{5.10}$$

The LSP reduces to the classical form of the periodogram in the case of equally spaced data. For non-uniform data, it differs only to the extent that the denominators in Eq. 5.10 differ from the factor N^{-1} in Eq. 5.9. The LSP can also be obtained by fitting simple sinusoids at different candidate frequencies to the data and constructing the power spectrum from the goodness of fit. This least-squares formulation of the LSP can be written as:

$$\begin{aligned} \chi^2(f) &= \sum_n (g(n\Delta t) - g(n\Delta t; f))^2 \\ P_{LS}(f) &= \frac{1}{2}[\chi_0^2 - \hat{\chi}^2(f)] \end{aligned} \tag{5.11}$$

where $g(n\Delta t; f)$ is the sinusoidal model to be fitted at each candidate frequency, $\chi^2(f)$ is the chi-squared function to minimise at each frequency, $\hat{\chi}^2(f)$ is its minimum value, and χ_0^2 is the constant reference model. In this view of the periodogram, the power peak height at a certain frequency is related to how well a sinusoid of that frequency fits the data.

Important considerations must be made for the limiting frequencies and the grid spacing. $1/T$ should be set as the minimum frequency so that a signal with that frequency can complete one full oscillation. The maximum frequency should instead be set as the Nyquist frequency. However, the uneven sampling affects the definition of the Nyquist frequency, which becomes large but not infinite, and dependent on the precision of the time measurements. The non-uniform Nyquist limit is a debated topic in the literature so, in this study, it is assumed that it is the same as for uniformly-spaced data. This is a fair assumption since the TESS data is nearly uniformly spaced. One must keep in mind that this "limit" will be imperfect owing to the uneven spacing, and so that there might be a nearly-aliased region. So, to summarise, the frequency limits used in this study are:

$$\begin{aligned} f_{min} &= \frac{1}{T} \\ f_{max} &= \frac{f_s}{2} \end{aligned} \tag{5.12}$$

Finally, in order to not miss power peaks, it is essential to determine how finely the frequency range just defined should be sampled. With a too-fine frequency grid, the computation time can become long, while if it is too coarse the periodogram might miss narrow power peaks falling between the grid points. So, the spacing must be at least equal to the expected widths of the periodogram peaks or smaller. A signal of length T will translate into sinc-shaped peaks of width around $1/T$ in the periodogram, so the frequency spacing Δf should be:

$$\Delta f = \frac{1}{n_0 T} \tag{5.13}$$

where the factor n_0 is at least 1, or larger. Most computational methods automatically optimise the frequency grid once the minimum and maximum frequency limits are passed to the function.

5.4.3. Units and normalisations

It is useful to be able to use the peaks in the power spectrum to infer the amplitude of the signal oscillations. There are different types of normalisations that can allow this. In the

unnormalised form (also called PSD normalisation), the values on the periodogram can be interpreted as the squared amplitudes of the signal's Fourier components at each frequency. So, if the signal has an amplitude of arbitrary units $[amp]$, the unit of the power spectrum is $[amp^2]$. Kjeldsen and Bedding 1994 however, point out that for a finite signal, the unit of the power spectrum would be $[amp^2/Hz]$. In this case, to derive the signal's amplitude oscillations from the power peaks, also the length of the signal T must be used. In particular, the longer the signal, the more the power peak increases. An oscillation of amplitude A in the LC at a certain frequency will translate into a power peak proportional to A^2T at that frequency in the LSP. Finally, for a signal that is finite and discrete, also the sampling time Δt plays a role and should be accounted for. In particular, a higher sampling rate translates into a higher power peak. Therefore, an oscillation A in the LC at a certain frequency translates into a power peak proportional to A^2N at that frequency in the periodogram, where, naturally, $N = T/\Delta t$.

For reading simplicity, it is desirable to be able to directly read the amplitude oscillations of the LC at each frequency on the periodogram, so the power spectrum should be scaled to power amplitude spectrum. The proportion between the power spectrum and the power amplitude spectrum is:

$$P_S(f) = \frac{N}{4} A(f)^2 \tag{5.14}$$

where $A(f)$ is the power amplitude spectrum. This relation can be derived from the equations in Kjeldsen and Frandsen 1992 and is also described in the documentation of the Python package `scipy`[6] (Virtanen et al. 2020), in the function `scipy.signal.lombscargle`[7].

5.4.4. Motivation

The LSP is a powerful tool that allows the identification of periodic signals in unevenly-spaced time series. Since different types of periodic variability can be found in stars due to rotational modulation given by spots on the surface, or pulsations in different modes, it is important to recognise which stars among the candidate calibrators possess these features and remove them from the Ariel catalogue. These types of variability usually persist for a long duration of the life of a star. So, identifying a star that is variable in a semi-periodic way through recent data, means that it will most likely continue to be variable in the near future when space missions like Ariel are launched. Similar considerations can be made for stars that do not show a clear periodic variability in TESS data: They will most likely be quiescent also in the next years. An example of a study that employs the LSP to look for variability in candidate calibrators is Mullally et al. 2022.

So, after all the considerations of this Section, it can be concluded that the LSP can be used as an effective tool for this study as well.

5.5. Statistical quantities of interest

Some useful statistical quantities that are computed to help in the characterisation of the sample include the mean μ, median Mdn, standard deviation around the mean σ_μ, and dispersion

[6] https://docs.scipy.org
[7] https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.lombscargle.html

around the median σ_{Mdn} of the LCs. The definition of the latter was given in Eq. 3.3. The reason for computing the dispersion around the median is that LC data might be dominated by outliers, that could originate from instrumental effects, corrected data points due to cosmic ray hits, or other reasons. So, the median value is much more representative of the "average" value of the time series than the mean, making it the most suitable reference level from which to compute the dispersion of the data. The larger σ_{Mdn}, the more the data points are spread out far away from the median. Since the aim of this study is to analyse the stability of stars, an ideal calibrator's LC should have a dispersion as close as possible to the standard deviation of the noise.

However, the photometric dispersion is not enough to assess the variability of stars, as it is dependent on some stellar properties like the magnitude. A quantity that is effective in further studying the flux stability also taking into account the flux uncertainty in the LCs is the reduced chi-squared χ^2_v, as it was defined in Eq. 3.4. Its values are interpreted like in the study by Ciardi et al. 2011. This was explained in Section 3.5. So, for good stellar calibrators, it is desired that $\chi^2_v < 2$, i.e. the star can be considered stable, with an excess dispersion of less than ~ 1.5 times the instantaneous noise in the data.

A final useful quantity is the standard error (SE). The SE of a statistic is the standard deviation of its sampling distribution (Anthony 2003). This quantity can be related to the flux stability requirement. So, the desired statistic to analyse is the median calculated over integration times of 6 hours, in order to understand its behavior and infer something about the flux stability. The standard error of the sample median is defined as:

$$SE(n_s) = \frac{\sigma_{Mdn}}{\sqrt{n_s}} \tag{5.15}$$

where n_s is the number of samples in the desired sample time. So, for this application, $n = 180$ since, in first approximation, there are 180 data points in 6-hour time intervals with a 2-minute cadence.

The sampling distribution of the median of the LC can be generated by repeatedly sampling the LC creating groups containing a number n_s of random points and calculating the median of each group. The medians obtained have their own distribution and statistical characteristics. In this study, the standard deviation for 6-hour-long intervals is of interest. To get this, instead of performing the sampling, the SE for a sample size n_s can be used since this is mathematically equivalent to computing the standard deviation of the sample median distribution. The advantage of using the SE rather than performing the sampling is that it is less computationally expensive.

So the SE gives an indication of the behavior of the distribution of the median, computed in time intervals that are 6 hours long. It can be interpreted as an averaging of the dispersion of the points within the interval of interest. However, it is important to keep in mind that the formal condition for the SE to provide this type of statistics is that the data points must be taken randomly within the time series. In the case of this study, it is desired to select consecutive data points to study 6-hour-long observing intervals. The SE statistic is still representative of the behavior of the sample median only if the time series is not dominated by a signal with structure but by Gaussian noise. This is shown in Fig. 5.10 by analysing

synthetic signals resembling TESS LCs that have been constructed as the sum of Gaussian noise and an underlying single sinusoidal signal with a defined frequency and amplitude.

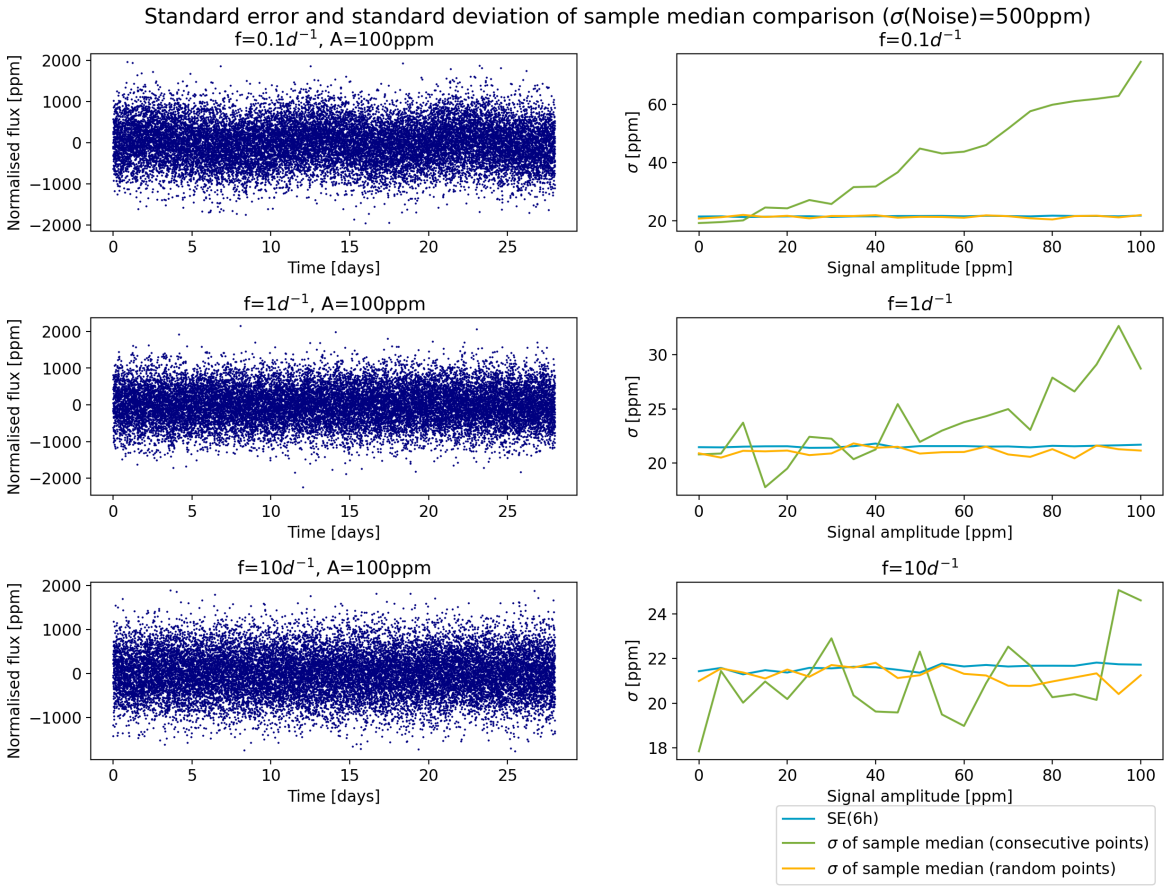


Figure 5.10: Comparison between the SE and the standard deviation of the sample median with consecutive and random point sampling, computed for three synthetic LCs. The panels on the left show LCs with different frequencies of the underlying sinusoid and an amplitude of 100ppm, with a white noise of standard deviation 500ppm. The right panels show the deviation between the SE and the sampling median for an underlying signal with a defined frequency and as a function of its varying amplitude.

Fig. 5.10 shows that the SE (blue line), following its definition, is equivalent to randomly sampling the LC multiple times and calculating the standard deviation of the sample median (green line). However, it is not always equivalent to calculating the standard deviation of the sample median when the samples are made of consecutive points (orange line). They are equivalent only when the underlying sinusoid is small in amplitude with respect to the noise or the frequency is large. The deviation is dependent on the number of points in the samples (in this study, the samples are set to 6-hour-long intervals) and the frequency of the sinusoid. With large frequencies (order of $10d^{-1}$) the statistics coincide. When the frequencies are smaller (orders of 0.1-1d^{-1}) the statistics coincide only when the underlying signal's amplitude is small (up to around 30ppm in this example). As the amplitude increases, the standard deviation using consecutive points increases and deviates from the SE. This means that it is important to make sure that the assumption of LCs without a large structured underlying signal is possible before using the SE, or the standard deviation of the sample median would be underestimated. If it is not possible to assume no structure in the signal, the behavior of the sample median for samples with consecutive points can only be known by performing the actual sampling, which

is computationally more expensive.

5.6. Selection criteria

To select suitable stellar calibrators for Ariel, the candidate list undergoes a cleaning process made up of three steps. The stars that pass all three criteria are good stellar calibrator candidates, while the others are discarded.

The criteria are, in order:

1. LSP criterion

The LSP power amplitude is lower than the 3σ *flux stability requirement of Ariel over 6 hours (*$100ppm$*), at each frequency.*

$$A(f) < 100ppm, \forall f \tag{5.16}$$

This first step is meant to perform a first cleaning of the candidate list by deleting all the stars showing photometric variability that is periodic and clearly identified by the LSP. This is mainly caused by rotational modulation of the LC or pulsations, as explained in Sec. 3.4.1. The limit is set to $100ppm$ because of the current definition of the Ariel flux stability requirement. In particular, it is not desired to have oscillation components of amplitude larger than $100ppm$ at time scales of 6 hours or lower ($f \geq 4d^{-1}$). For $f < 4d^{-1}$, the allowed single maximum peak amplitude would increase as the frequency decreases. However, stellar flux time series generally do not have a perfect sinusoidal shape at a specific frequency but rather possess multiple components. So, these might add up making the flux fluctuate more than $100ppm$. Moreover, the LC noise also adds spurious peaks in the power amplitude spectrum, which can happen at any frequency. So, this first criterion is extended to the whole frequency range of the LSP to perform a first skimming of all stars with a detected periodic variability.

2. Chi-squared criterion

The reduced chi-squared is lower than 2.

$$\chi^2_v < 2 \tag{5.17}$$

As explained previously and described by Ciardi et al. 2011, the larger the reduced chi-squared, the more variable the flux is on top of the uncertainty in the measurements. The LCs are assumed to have an excess dispersion lower than 1.5 times the instantaneous noise when the reduced chi-squared is lower than 2. It is assumed that this means that they are stable.

3. Host star criterion

The star does not host one or more confirmed or candidate exoplanets.

This step is performed through a cross-match with the NASA Exoplanet Archive's Planetary Systems Table (NExScI 2020). Removing host stars makes sure that the LC will not be affected by dips and other flux variations due to exoplanets. This is the least stringent requirement, as small planets or planets that are very far away from the star might not give a large contribution to the signal. To be safe, these stars are removed but, in the case of a shortage of stellar calibrators, very stable stars with transiting exoplanets that are not too big and not orbiting close to the star might still be taken into consideration. These stars might still be stable calibrators if observed outside of the transit times. However, in this case, a careful evaluation of each host star would be necessary.

After these three steps, the stars that are left do not show a clear, periodic variability in their fluxes, do not have an excess dispersion of more than 1.5 times the measurement uncertainties, and are not affected by known or candidate exoplanets. So, these stars do not vary above the noise of the LC and they are, in first approximation, mainly dominated by noise.

5.7. Summary

In this Chapter, the methodology used in this study was presented. After having understood how databases like SIMBAD and MAST work and the characteristics of the TESS LCs, the mathematical algorithms and statistical quantities employed for the flux stability analysis of the starting calibrator candidates were described.

The workflow of this project is organised as follows:

1. The first Ariel calibration lists, EChO list and Catalogue 1, are unified, as explained in Section 5.3, to obtain the starting calibrator candidate sample of this study, called Catalogue 2.
2. The TESS 2-minute-cadence LC files for each candidate are downloaded through the `lightkurve` Python package. If multiple observations of a single star are available, only the first sector is arbitrarily chosen for the analysis. The time and PDC SAP flux vectors are stored. All data used has been extracted and processed with the SPOC pipeline.
3. The LSP is computed for each LC using the `astropy.timeseries.LombScargle` Python package. The resulting frequency and power amplitude vectors are stored and, in particular, the maximum power amplitude peak and its corresponding frequency are recorded for each candidate.
4. Other statistical quantities of interest described in Section 5.5 including the mean, median, standard deviation, dispersion about the median, and reduced chi-squared are computed and recorded for each candidate.
5. The selection process is performed as described in Section 5.6.
6. The computed quantities are used to create plots and perform a stellar flux variability population study. The results and discussions are presented in the next Chapter.

6

Results

This Chapter presents the results of this study. First, an overview of the different types of LC shapes and corresponding Lomb-Scargle periodograms is given (Section 6.1). There, the sample is also roughly categorised into different types of stellar variability by physical origin (rotators, pulsators, etc). Afterwards, the selection process is performed and the most promising stellar calibrator candidates are found (Section 6.2). A sensitivity analysis is also conducted to understand how many stars would pass the selection for different values of the flux stability requirement. Then, a variability population study is performed in order to look for possible correlations between flux stability and stellar properties (Section 6.3). These results are compared with the previous variability study by Ciardi et al. 2011 that was presented in Section 3.5. Then a discussion and caveats on the results are presented (Section 6.4), and recommendations for future work are finally given (Section 6.5).

6.1. Candidate calibrators categorisation

Most LCs and their LSP look similar and can be grouped into different categories. The following figures (Fig. 6.1-6.6) display the typical shapes that are found throughout the sample, with possible explanations for their nature. Table 6.1 at the end of this Section recaps how many stars in the sample belong to each of the defined categories.

6.1.1. Transit events

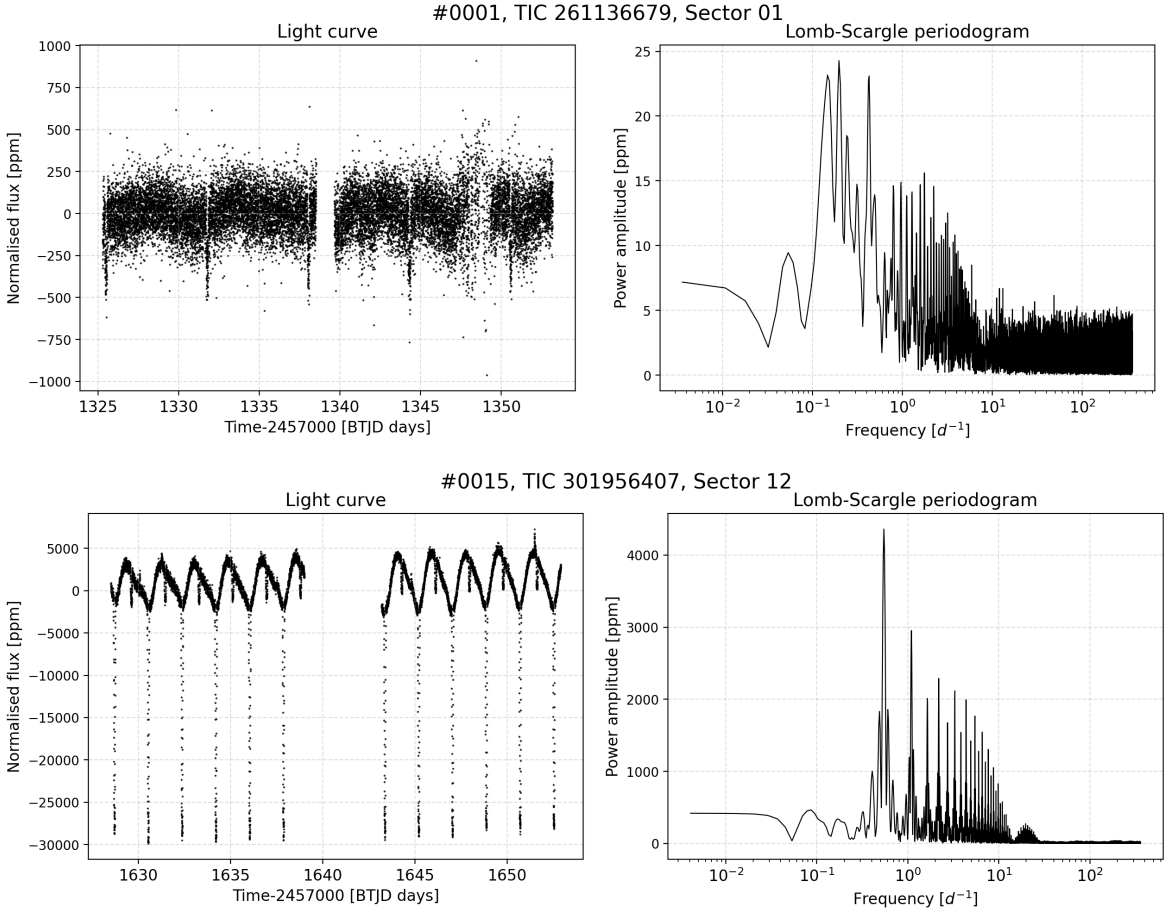


Figure 6.1: Examples of LCs with transit events and their associated LSP. In the top panels, the transits are caused by an exoplanet, while in the lower panels, the transits are due to a transiting companion star.

Fig. 6.1 shows transit events in the LCs. In the upper panels, multiple exoplanet transits are visible. This planet has an orbiting period of around 6 days. As it is known, the LSP is not constructed to efficiently detect this type of variability (sporadic flux dips) but rather periodic variability that resembles a sinusoid or sum of sinusoids. To identify which stars might have these dips it is necessary to cross-match the sample with the NASA Exoplanet Archive[1] Planetary Systems Table (NExScI 2020) to remove all the host stars, as it will be done in the third step of the selection procedure.

Similarly, the power panels show multiple transit and eclipse events caused by a stellar companion. The candidate list, in this case, could be cross-matched with a TESS eclipsing binary

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

catalogue (Pra et al. 2022), although it still does not cover the most recent sectors. However, this cross-match is not strictly necessary as the dips are extremely large. Due to their large flux variation, eclipsing binaries are identified in the LSP and can easily be filtered.

6.1.2. Pulsators (high-frequency variations)

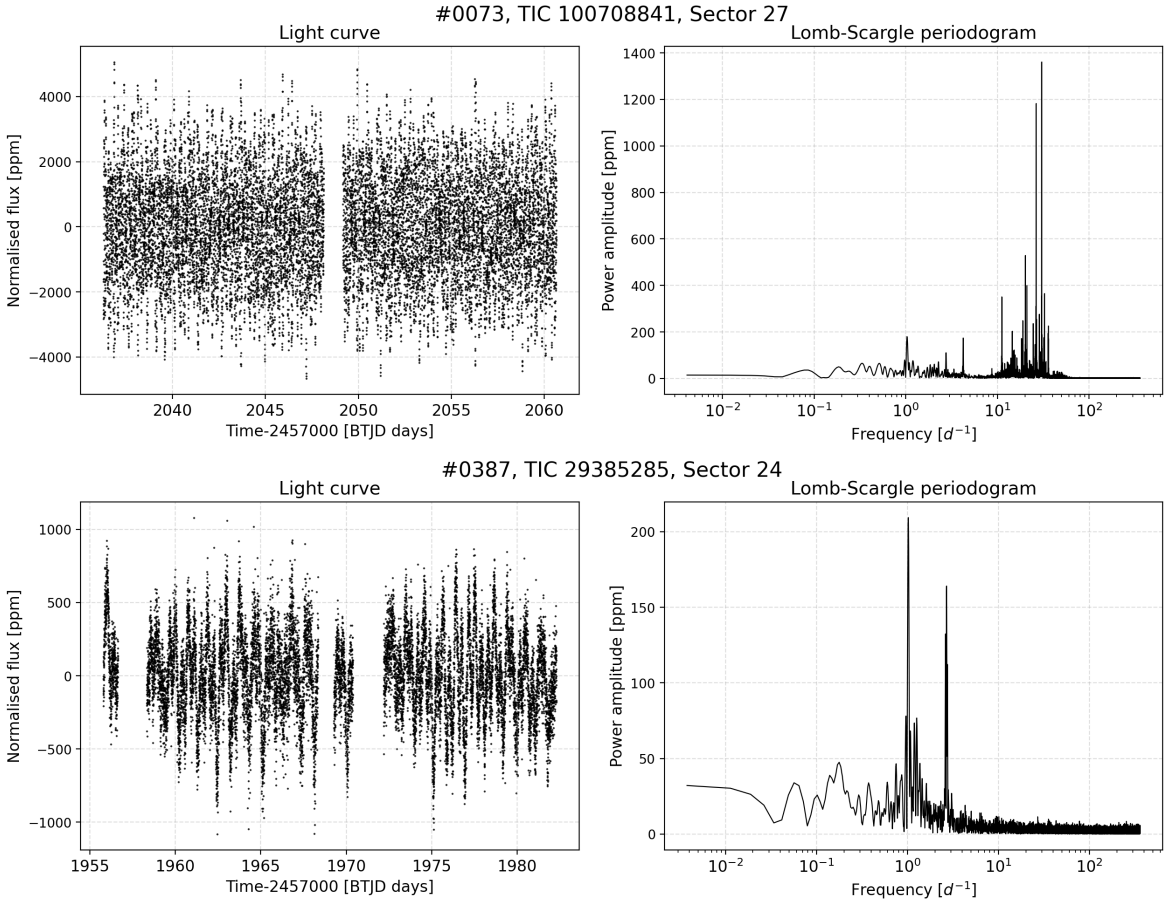


Figure 6.2: Examples of LCs of pulsating variables and their associated LSP.

The LCs of pulsating variables show high-frequency flux oscillations as displayed in Fig. 6.2, which can have a range of amplitudes. For the sake of the categorisation, it is assumed that stars dominated by "high-frequency" oscillations are those with the largest LSP power amplitude peaks at $f > 1d^{-1}$. This means that their oscillation periods are in the order of hours.

As previously explained in Section 3.4.1, pulsators are stars whose outer layers quickly expand and contract, and possess large surface temperature variations which cause the observed flux variation in their LCs. The LSP is good at detecting this type of variability as it is quite regular and periodic. So, the LSP presents clear peaks at the oscillation frequencies of the stars, of a value that is dependent on the amplitude of the oscillations.

6.1.3. High-amplitude rotators (low-frequency variations)

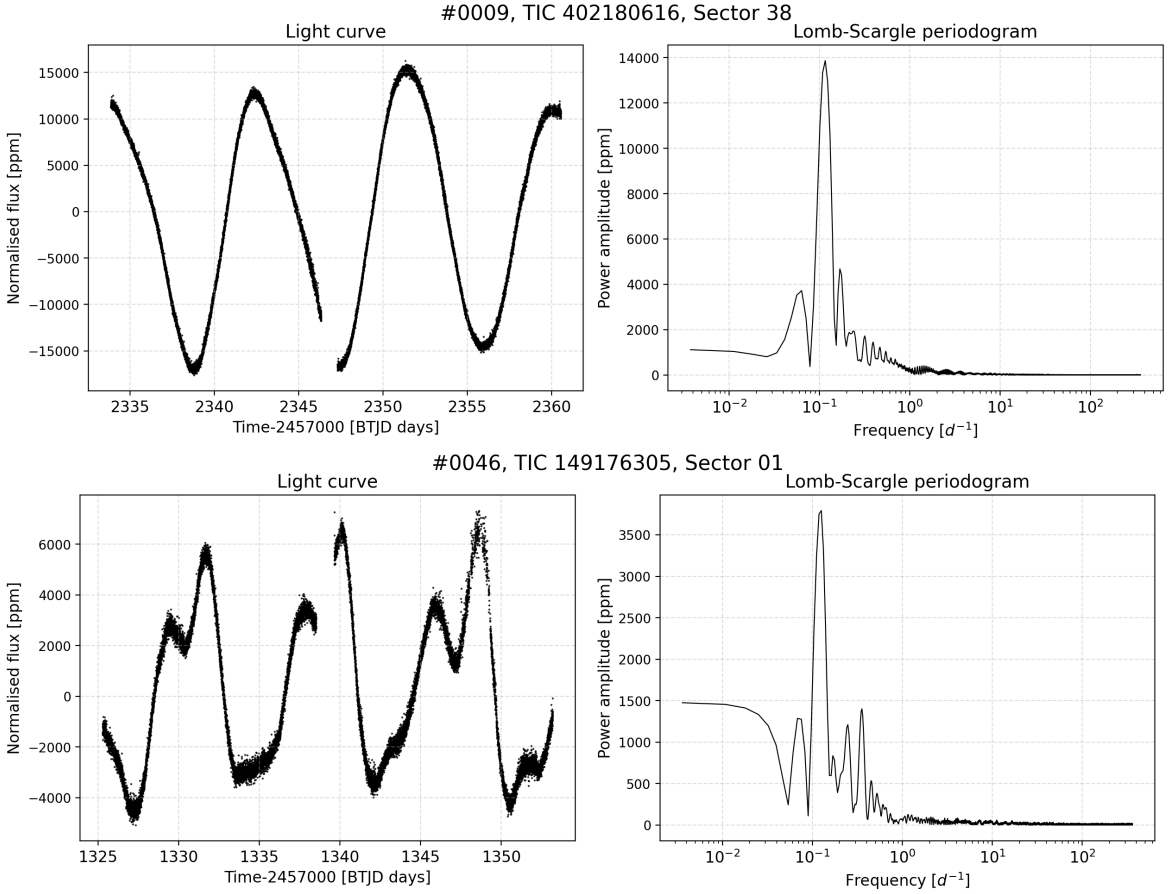


Figure 6.3: Examples of LCs of high-amplitude rotating variables and their associated LSP.

The LCs of rotating variables show periodic low-frequency oscillations, as shown in the examples in Fig. 6.3. Contrarily to the pulsators, it is assumed that stars dominated by "low-frequency" oscillations are those with the largest LSP power amplitude peaks at $f \leq 1d^{-1}$. So, their flux oscillation has a period in the order of days. Moreover, if they possess at least one power amplitude peak in the LSP larger or equal to $100ppm$ ($P_{max} \geq 100ppm$), then they are categorised as "high-amplitude" rotators.

As previously explained in Section 3.4.1, rotators are stars whose brightness variations are due to an uneven temperature distribution on the stellar surface originated by stellar spots and plages that rotate in and out of view as the star rotates about its rotational axis. High-amplitude rotators show a large and clear periodic variability that is clearly semi-sinusoidal in shape. So, the LSP is very effective in recognising the oscillation. In this case, looking at the dominant frequency in the LSP enables us to obtain information on the angular velocity of the star as it rotates about its axis. Some studies exploit this characteristic shape of the LCs to recover stellar rotation periods, like McQuillan et al. 2014.

6.1.4. Low-amplitude rotators (low-frequency variations)

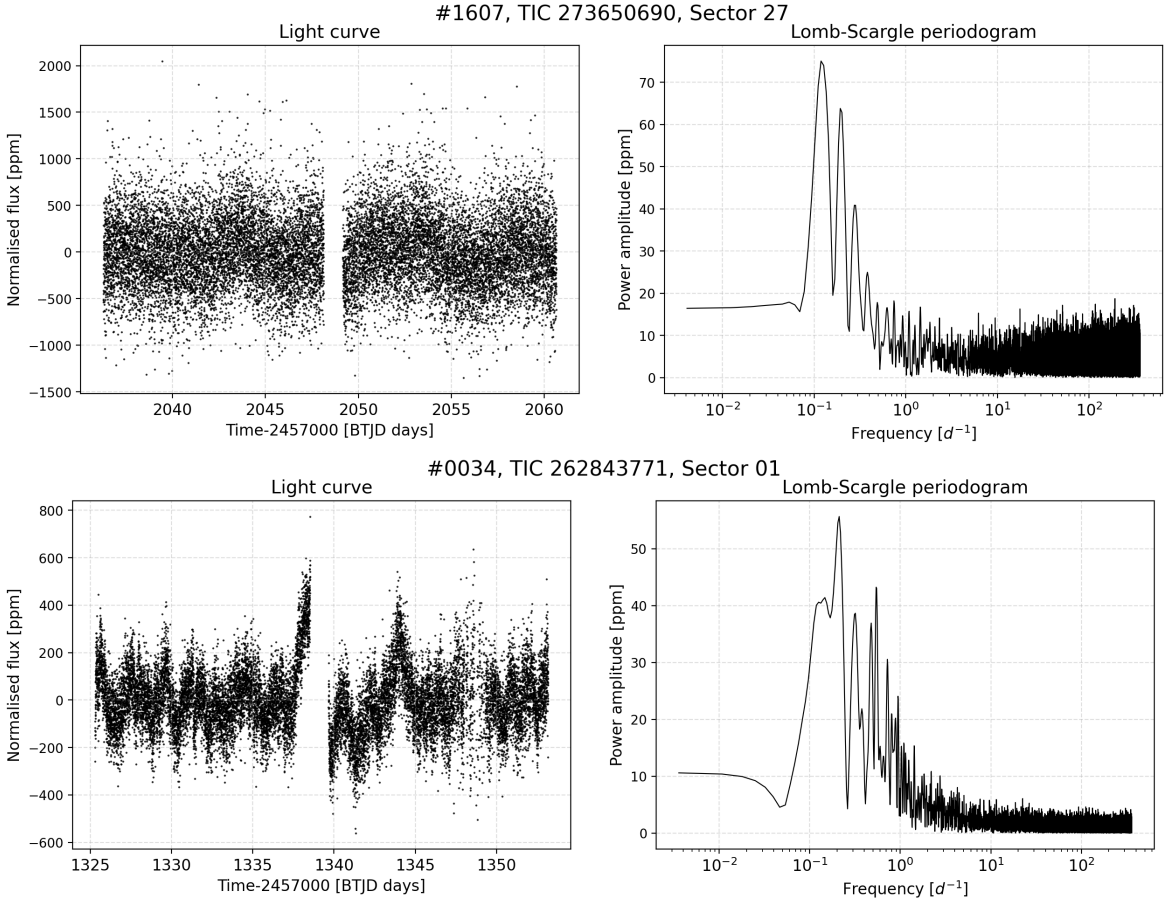


Figure 6.4: Examples of LCs of low-amplitude rotating variables and their associated LSP.

Similarly to the case of the high-amplitude rotators, the low-amplitude rotators possess LCs with a flux variability that resembles a semi-sinusoid or a sum of sinusoids. An example is shown in Fig. 6.4. "Low-amplitude" refers to stars with clear LSP periodogram peaks that are lower than $100ppm$ ($P_{max} < 100ppm$).

These flux variations can again be due to the rotation of the star about its axis, with stellar spots and plages on its surface rotating in and out of view with respect to the observer. However, as a star gets older, the rotation slows down due to the angular momentum loss and the stellar activity generally decreases, leading to lower-amplitude variations. So, it is possible that stars belonging to this category are older than the ones shown in the previous sub-section, and are therefore less active rotators. The flux variations have lower amplitudes in this case, so they start to approach the noise level of the LCs. The LSP is good at identifying structured periodic signals even when they are not clearly visible to the eye. So, some clear peaks are still visible in the frequency domain in these examples, way above the noise level. However, as the amplitudes get smaller and smaller, the peaks in the LSP can become too small and indistinguishable from the spurious, random peaks caused by the noise. In this case, the stellar variability would be too small and undetectable by the LSP.

6.1.5. Mixed stars

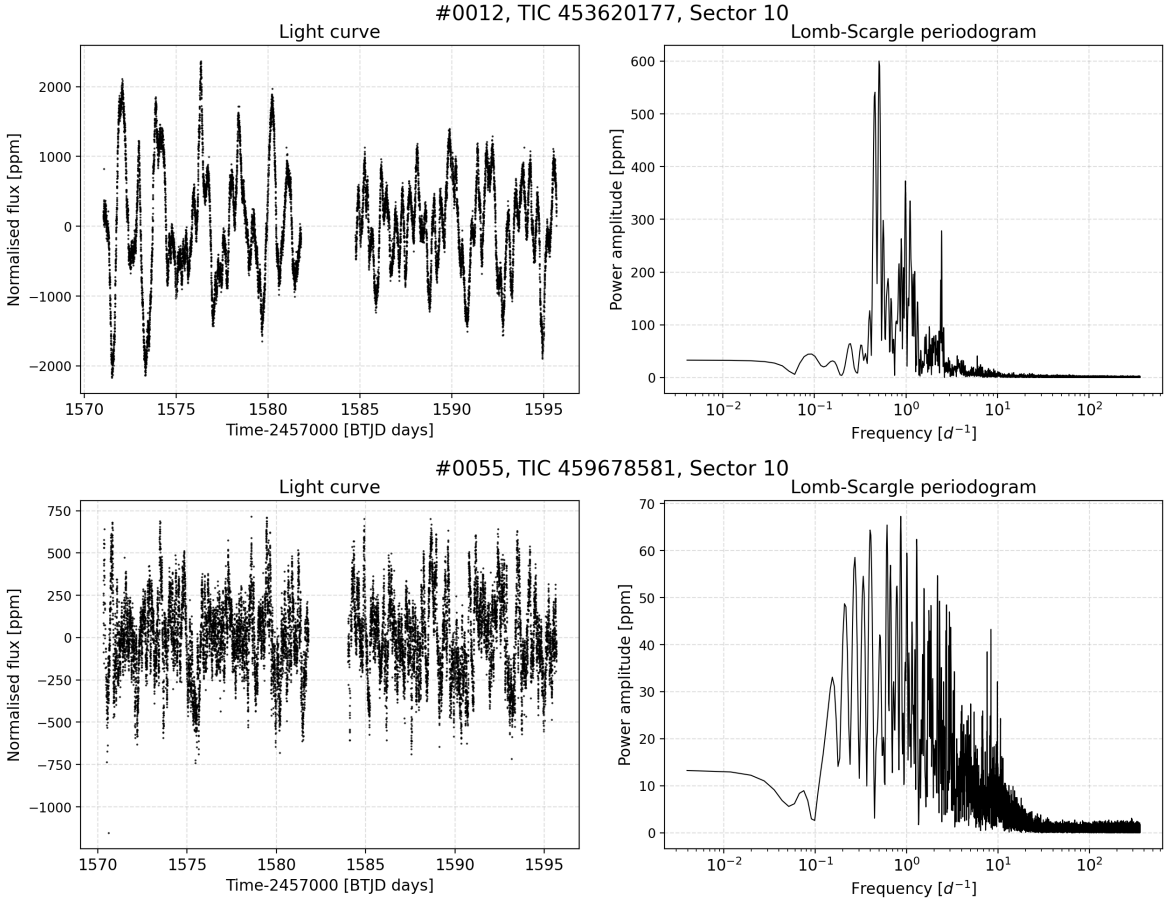


Figure 6.5: Examples of LCs showing a mixed flux variability, with both a high-frequency and a low-frequency oscillation component of comparable amplitudes, and their associated LSP.

As shown in Fig. 6.5, stars can show multiple oscillation components of a similar amplitude at both low and high frequencies (both in the order of days and hours). This gives origin to a mixed behavior in the LCs.

6.1.6. No periodic variability

Fig. 6.6 instead shows examples of stars that do not appear to vary in flux. Their LC, by eye, appears to only be noise, and the LSP does not display significant power amplitude peaks on top of the random noise. The stars that belong to this type are generally the most probable to be good calibrator candidates.

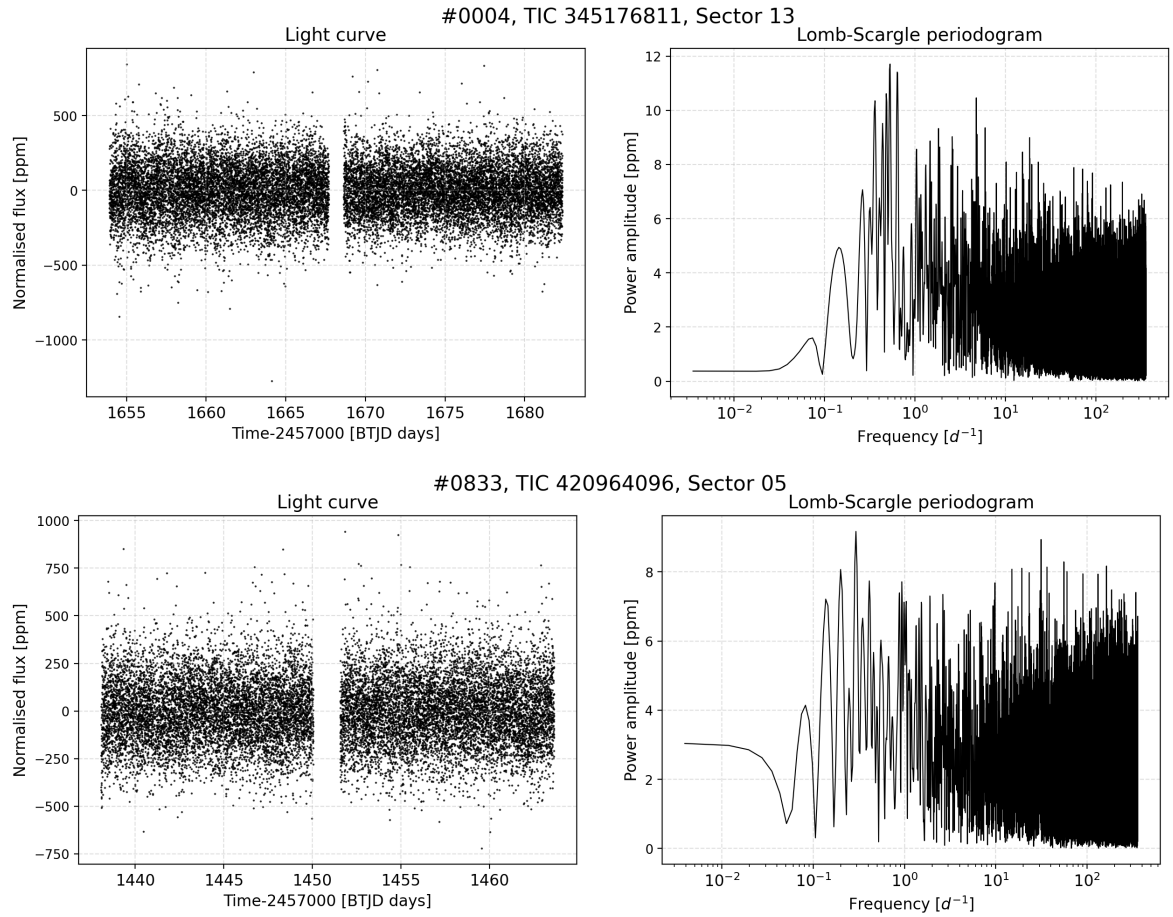


Figure 6.6: Examples of LCs of stars not showing a clear flux variability and their associated LSP.

6.1.7. Summary of categories

The categories, their definition, and the number of stars belonging to each are shown in Tab. 6.1. Note that the mixed stars are not categorised, but only the maximum power amplitude peak and its corresponding frequency are used. Moreover, as a first approximation, the stars without detected variability are assumed to have power amplitude peaks in the LSP lower than $50ppm$. The proper selection of stable stars is performed in Section 6.2.

The high-amplitude rotators make up the majority of the sample, with a percentage of around 41%, followed by the stars with no detected variability (31%), and the low-amplitude rotators (15%). It is important to note that the ratios between these last two categories are dependent on the maximum allowed LSP power amplitude peak in order for a star to be defined as stable or variable with a low amplitude. So, these percentages are rough estimates and somewhat arbitrary. In the next Section, the stable stars are defined with the formal selection steps described in Section 5.6. Finally, LCs with transiting events, either due to exoplanets or stellar companions, make up 11% of the sample, and pulsators are the smallest category (less than 2%).

Figure 6.7 depicts all the stars of the sample as a function of the maximum LSP power amplitude peak and its corresponding frequency. This gives a visual representation of the different categories.

Table 6.1: Number of stars belonging to each defined category. Mixed stars are not categorized and the no variability stars are estimated roughly as a first approximation by setting a maximum allowed LSP power amplitude peak of $50ppm$.

Category	$f(P_{max})$ $[d^{-1}]$	P_{max} $[ppm]$	Number of stars	Percentage of the sample
Transiting events	any	any	212	10.94%
Pulsators	> 1	any	33	1.70%
High-amplitude rotators	≤ 1	≥ 100	801	41.35%
Low-amplitude rotators	≤ 1	≥ 50 and < 100	282	14.56%
No detected variability	any	< 50	609	31.44%

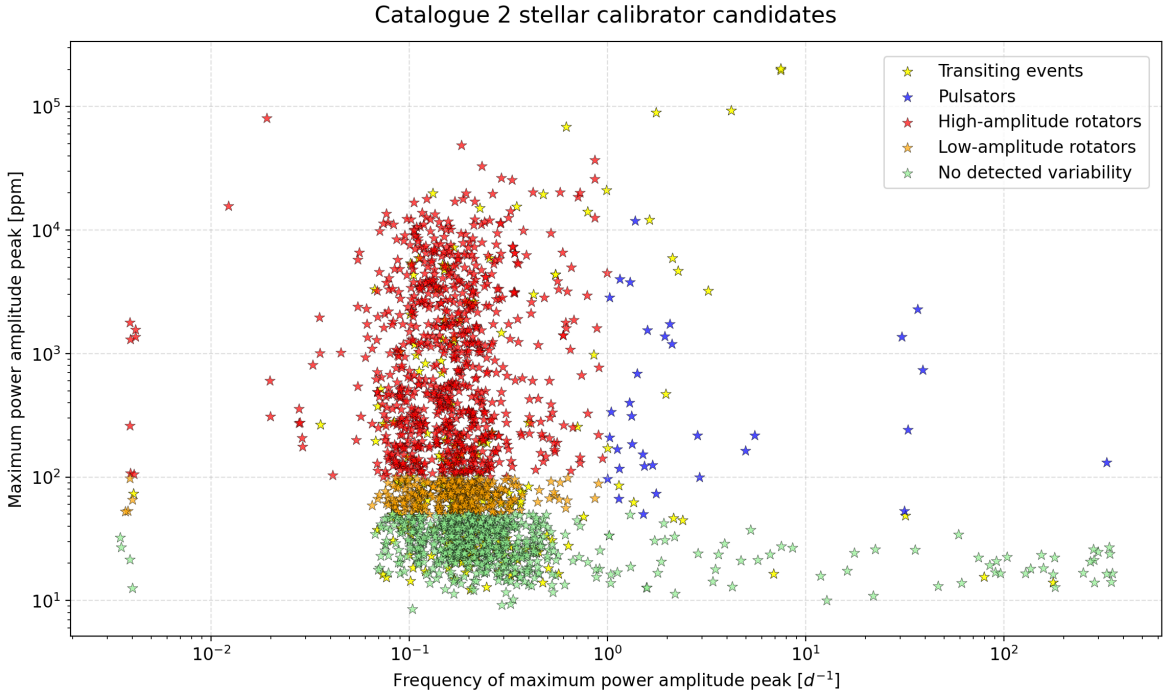


Figure 6.7: All calibrator candidates displayed with respect to their maximum LSP power peak amplitude and corresponding frequency and divided into categories.

The high and low-amplitude rotators and the stars with no detected variability are mainly clustered in a frequency range between $f = 0.07d^{-1}$ and $f = 0.5d^{-1}$, which corresponds to rotation periods of the stars about their axis between $2d$ and $14d$. It is important to note that the cluster seems to be divided into two groups centered at different frequencies. The groups are centered around frequencies of $f = 0.1d^{-1}$ and $f = 0.2d^{-1}$, corresponding to rotation periods of $5d$ and $10d$, respectively. These are roughly 1/3 and 1/6 of a TESS Sector length ($\sim 30d$). It is not clear if this peculiar distribution has a physical origin or is caused by windowing effects. The stars with no detected variability are spread also in the lower right part of the plot, where frequencies are higher. This is because, when noise becomes the dominant factor in the LC, the highest peak can more easily be found at any frequency as a sum of random noise components. So, this spread is expected.

Some stars with different values of maximum power amplitude peak are found at the left edge of the plot, at very low frequencies (around $f = 0.006d^{-1}$). These are LCs that show a drift,

which can either have a physical origin, like a flux variation with a very long period, or can be due to instrumental systematics that were not properly corrected in the LCs. Pulsators are only found in the right part of the plot, where peaks occur at high frequencies. Finally, the stars undergoing transit events, as expected, are spread everywhere in the plot without being correlated to maximum power or frequency values.

Figure 6.8 shows the distributions of some stellar properties of interest for three stellar categories that were just defined. These are compared to the distributions of the whole sample that were previously shown in Fig. 5.7.

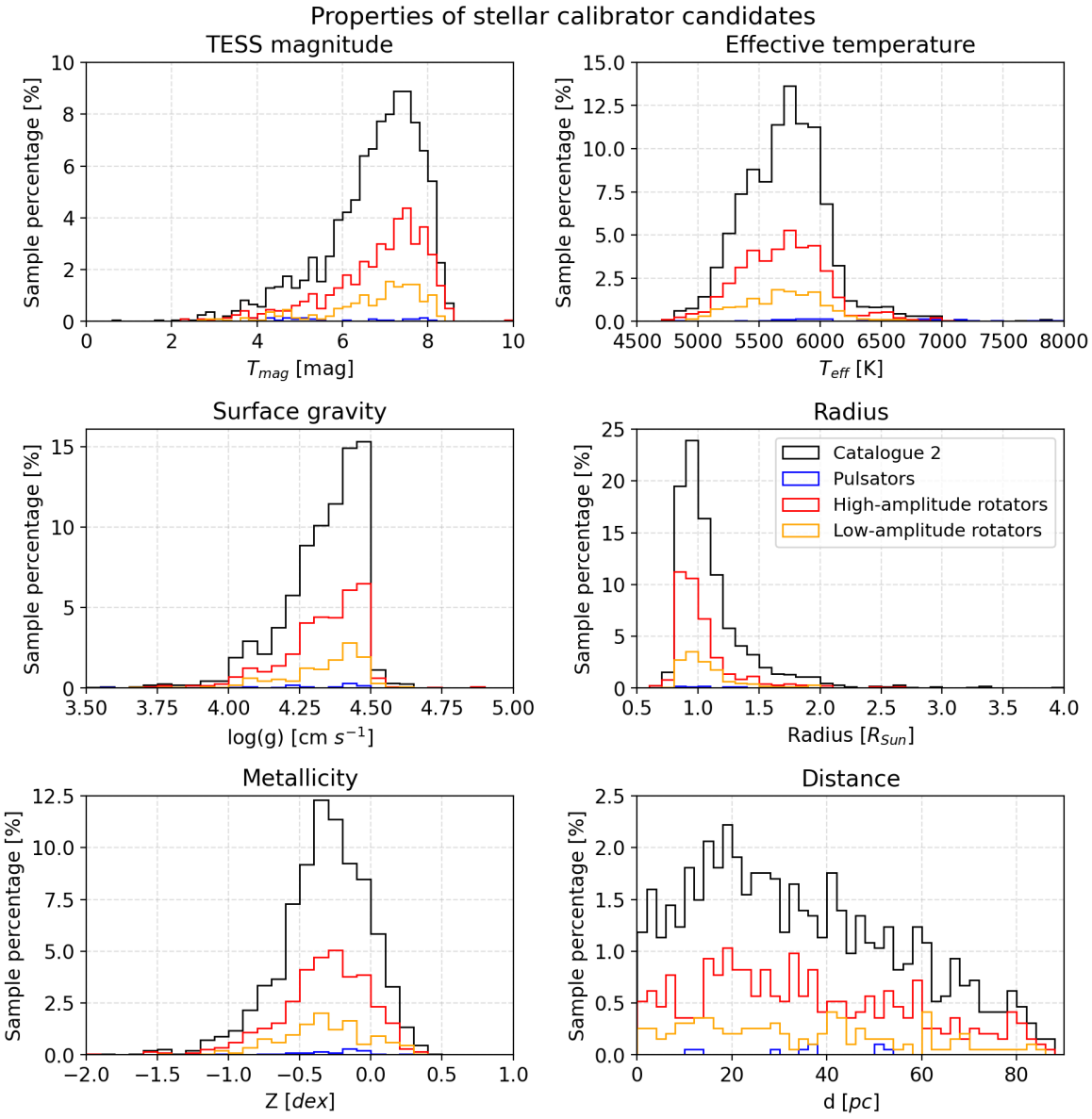


Figure 6.8: Properties of the original candidate sample Catalogue 2 and of the defined stellar categories including pulsators, low-amplitude, and high-amplitude rotators.

The high and low-amplitude rotators follow almost exactly the distribution of the whole sample for each of the properties shown in the histograms. This makes sense since these stars make

up the majority of the sample. They cover the full range of each of the parameters and have a similar distribution shape to the whole sample. The pulsating stars also generally follow the same distribution shapes. However, these seem to cover only a very specific temperature range (upper right panel): They are mainly found at the highest temperatures of the sample ($7,000$-$8,000K$), and some others around $5,500$-$6,000K$.

6.2. Calibrators selection

The selection criteria described in Section 5.6 are applied to the starting sample Catalogue 2 in a consecutive order. The results are shown in Tab. 6.2.

Table 6.2: Number of stars passing through the selection criteria. The starting sample is composed of 1937 stars (Catalogue 2). The criteria are applied in the order shown and in a consecutive way.

Criterion	Number of stars passing	Percentage of stars passing
1 (LSP)	1033	53.33%
2 (Chi-squared)	664	34.28%
3 (Host star)	581	29.99%

The first step filters out flux macro-variability, i.e. periodic flux variations with amplitudes that are too large for the Ariel stability requirement. So, the candidate calibrators that are surely too variable are deleted in this step thanks to the large power amplitude peaks displayed in the LSP. These bad candidates include most of the pulsators, the high-amplitude rotators, and also the eclipsing binary stars. Almost half of the sample is removed after the LSP criterion is applied.

The second step filters the stars with an excess dispersion in their flux larger than around 1.5 times the flux uncertainty. So, this step is concerned with the instantaneous noise in the LC and deletes the stars that show a significant variation with respect to their flux error bars. Another 19% of the starting sample is removed by the chi-squared criterion. These mainly include some low-amplitude rotators and mixed stars that possess multiple oscillation components with amplitudes lower than $100ppm$ but whose flux variations are significantly larger than the instantaneous noise in the LCs.

The third step filters out the stars that are known to host transiting exoplanets or that could possibly do (called TESS Project Candidates, TPC). This step ensures that sporadic dips in flux or effects of the phase curve of the planet in the LCs will not be present. The stars additionally removed by the "host star" step are 81 (of which 6 are TPC), around 4% of the starting sample. Again, the stars removed by this step might be good calibrators if observed outside of the transits and if the phase curves of the planets do not have a big influence. However, a careful evaluation of each host star should be performed separately. For simplicity, in this study, we remove all confirmed and possible host stars.

The remaining stars after the selection is performed are 581 ($\sim$30% of the original sample). These do not show a clear, periodic variability in their fluxes, do not have an excess dispersion of more than 1.5 times the measurement uncertainties, and are not affected by known or candidate exoplanets. So, as these stars do not vary above the flux uncertainty and they are mainly dominated by noise, they are assumed to be stable. However, there could still be underlying structured signals of amplitude up to $100ppm$, as well as systematic effects that

were not completely filtered by the pipeline. This makes them the most promising candidates to be good Ariel stellar calibrators but also means that further study on their variability is needed. So, they make up a new stellar calibrator candidate list that will be called "Ariel Catalogue" for the rest of this study. More detailed information on the available columns of this catalogue and its units can be found in the Appendix Section A.4.

Figure 6.9 shows the positions in the sky of the Ariel Catalogue stars. These appear to be distributed fairly homogeneously in the sky.

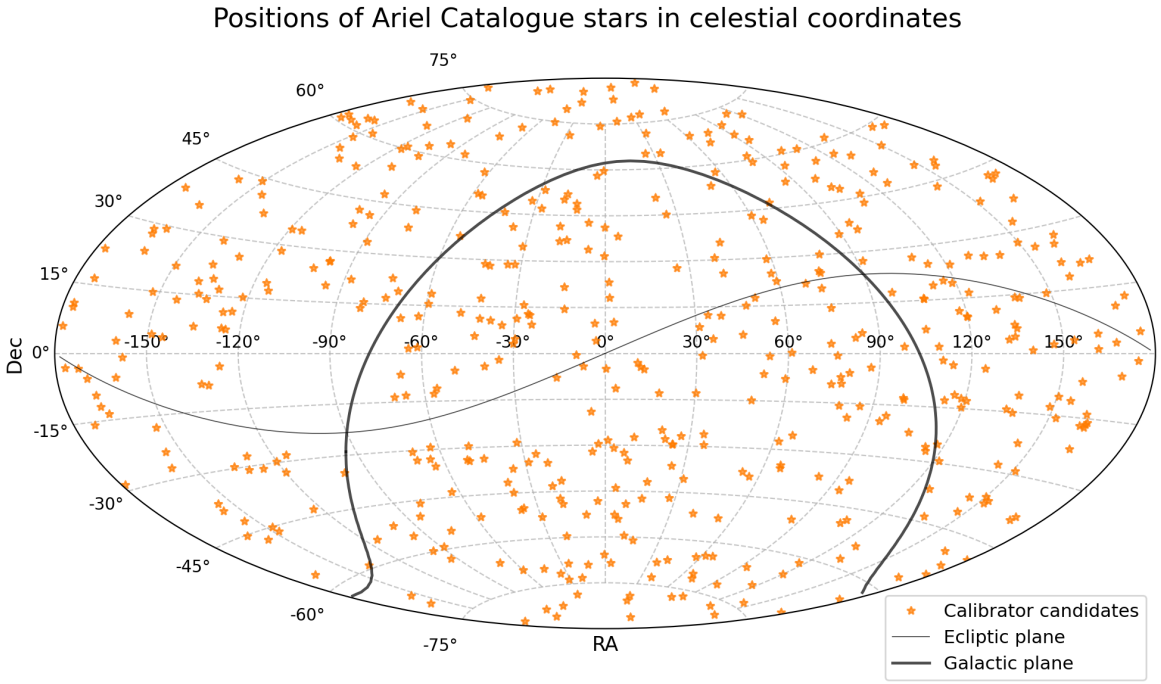


Figure 6.9: Positions in the sky of the Ariel Catalogue stars, in ICRS and epoch J2000.

Figure 6.10 shows the distributions of some stellar properties of interest for the Ariel Catalogue stars, with respect to the distributions of the whole sample that were previously shown in Fig. 5.7. A striking characteristic of the Ariel Catalogue distributions is that their shapes resemble, for each stellar property, the shapes of the distributions of the whole sample. This means that, once the correct value range for each stellar parameter is selected, the probability of finding a stable star inside that range remains constant and around 30%. When looking at the TESS magnitude (upper left panel) it can be noted that this is the only stellar property to show a relation with stability. In fact, many faint stars are selected, which is possibly due to the noise level in the LCs getting higher as the star gets dimmer, making it more difficult to identify flux variability. In fact, the noise level can "hide" the variations. On the contrary, there are no selected stars with a brightness lower than $T_{mag} = 5mag$. This is probably due to the lower uncertainty in the LCs for bright stars that make it easier to pick up flux variability on top of the instantaneous noise and make the chi-square statistics increase. Therefore, this effect is probably not physical but related to the different noise levels and precision of the instrument in relation to the stellar magnitude. The effective temperatures (upper right panel) of the selected stars are instead found only in the range $5,000$-$6,300K$. For the surface gravity (central left panel) the range extends almost to the whole sample range ($4 < log(g) < 4.5log(cm\ s^{-1})$). For

the radius (central right panel) there are no stable stars with a radius larger than $1.8R_{Sun}$. Finally, the metallicity (lower left panel) and the distance (lower right panel) ranges for stable stars are the same as for the whole sample. As for the metallicity, it was expected that stars with low metallicities (assumed to be older, i.e. less active) would have a higher probability of being stable, but this does not seem to be identified.

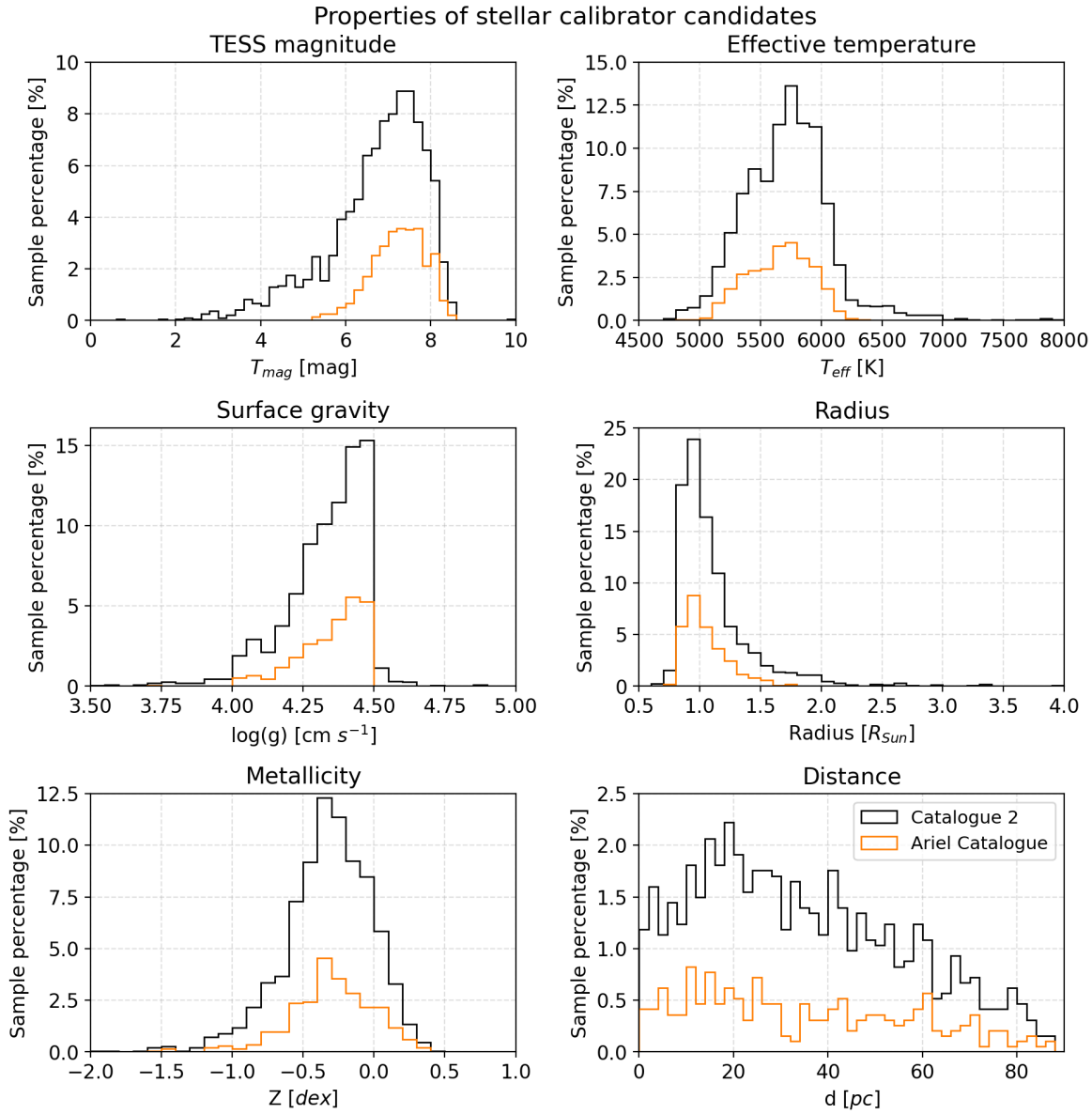


Figure 6.10: Properties of the original candidate sample Catalogue 2 and of the Ariel Catalogue stars, including TESS magnitude, effective temperature, surface gravity, radius, metallicity, and distance distributions.

Figure 6.11 shows a zoomed-in version of Fig. 6.7 where only the stars passing the selection are included. They are displayed also in relation to their effective temperature to look for a possible correlation between a specific temperature in the range and a power or frequency value. However, no clear correlation is found.

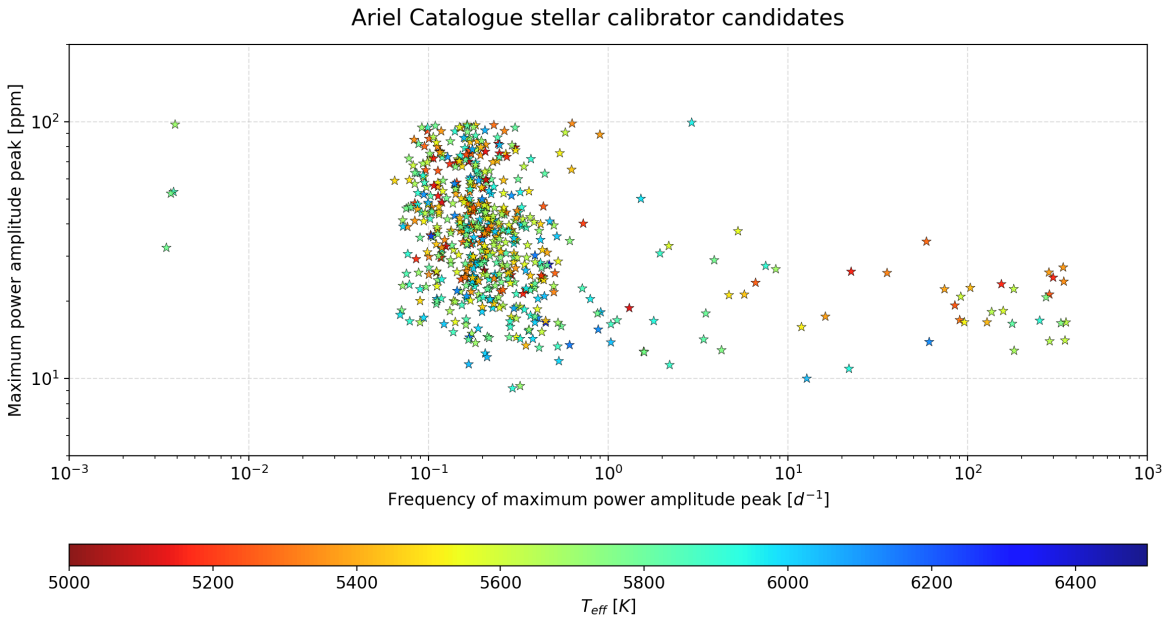


Figure 6.11: Ariel Catalogue stars displayed with respect to their maximum LSP power peak amplitude and corresponding frequency, with an indication of their effective temperatures.

6.2.1. Sensitivity analysis

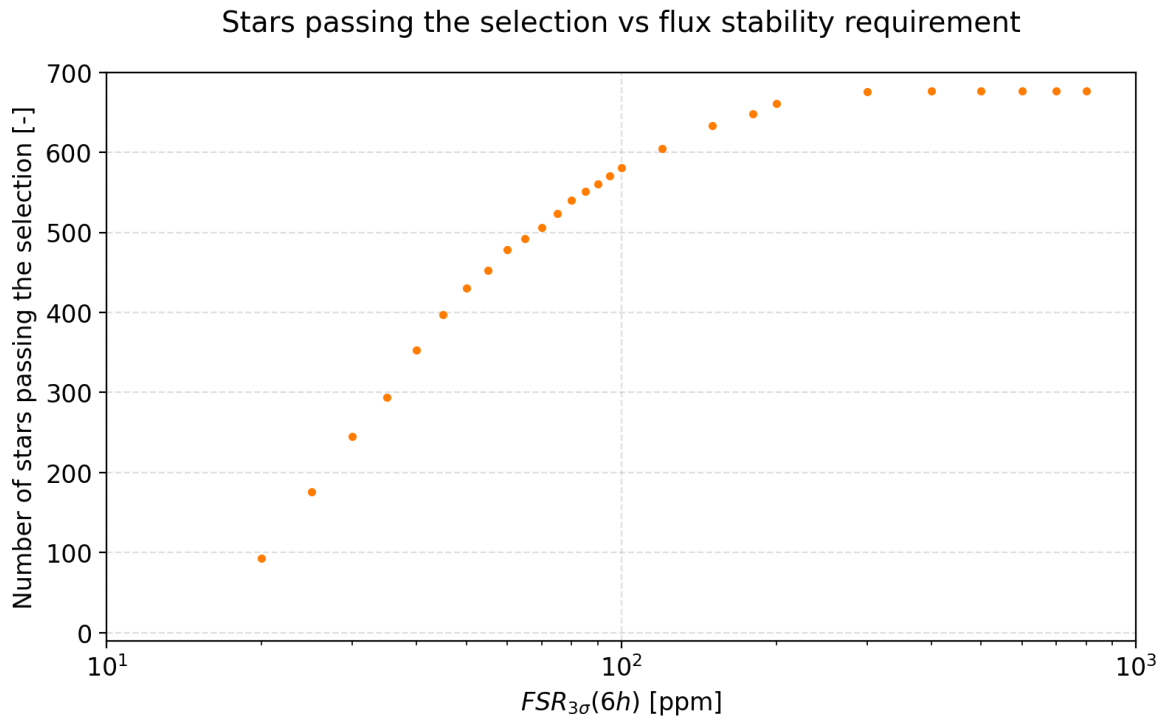


Figure 6.12: Number of stellar calibrators passing the selection procedure for different values of the flux stability requirement.

The sensitivity analysis is aimed at investigating how many calibrators would be selected from the sample for different values of the stability requirement. By just tweaking the limiting value

in the first selection step (LSP), it is possible to make an estimate of how many stars would be stable for different stability requirements, either less or more stringent than the current one. This is important to explore because the Ariel stability requirement is still not set in stone and could change in the future. Although this is applied to Ariel, it also gives insights into the different levels of stability of the sample that can also be applied to other missions.

The second and third selection criteria are not changed, whilst the first one is run with different values of the flux stability requirement ($FSR_{3\sigma}(6h)$). Figure 6.12 shows that the number of stars passing the selection procedure rises in a non-linear way as the requirement value increases. It finally reaches a plateau from $FSR_{3\sigma}(6h) = 200ppm$ onwards, counting around 700 stars passing the selection.

6.3. Variability population study

By using the statistical quantities that were calculated for each LC, it is possible to perform a variability population study of the candidate sample, with the aim of looking for possible correlations between stellar characteristics and flux stability. The plots in this Section are inspired by the ones found in the paper Ciardi et al. 2011 whose main results were described in Section 3.5. Comparing the results of this study with those by Ciardi et al. 2011 is essential to understand if the stability fractions determined for the small region of the sky observed by Kepler are also representative of the whole sky.

Fig. 6.13 can be compared with the G dwarfs panel of Fig. 6 by Ciardi et al. 2011 (Fig. 3.3 in this document). It displays the dispersion around the median of the LCs as a function of the TESS magnitude for the whole Catalogue 2 sample, and the stars passing the selection are shown in a different color. A bimodal distribution is visible, with two clusters of points in the ranges $100 < \sigma_{Mdn} < 1,000ppm$ and $1,000 < \sigma_{Mdn} < 10,000ppm$. This is consistent with the figure by Ciardi et al. 2011. The Ariel Catalogue stars belong to the lower-dispersion group and follow the shape of a curve representing the uncertainty limit of the LC. As the star gets dimmer, the noise increases, making the dispersion of the LC also increase.

The same bimodal distribution is also seen in Fig. 6.14, which can be compared with the G dwarfs panel of Fig. 7 by Ciardi et al. 2011 and shows a good agreement. Here the percentage of the stars belonging to the two groups is more clearly visible. 80% of the Catalogue 2 sample belongs to the region with lower dispersion, whilst in Ciardi et al. 2011 that region is occupied by 90% of their Kepler sample. These percentages are slightly different, and this can be due to a bias related to the limited size of the Catalogue 2 sample with respect to the sample of Ciardi et al. 2011, or it can have a physical reason.

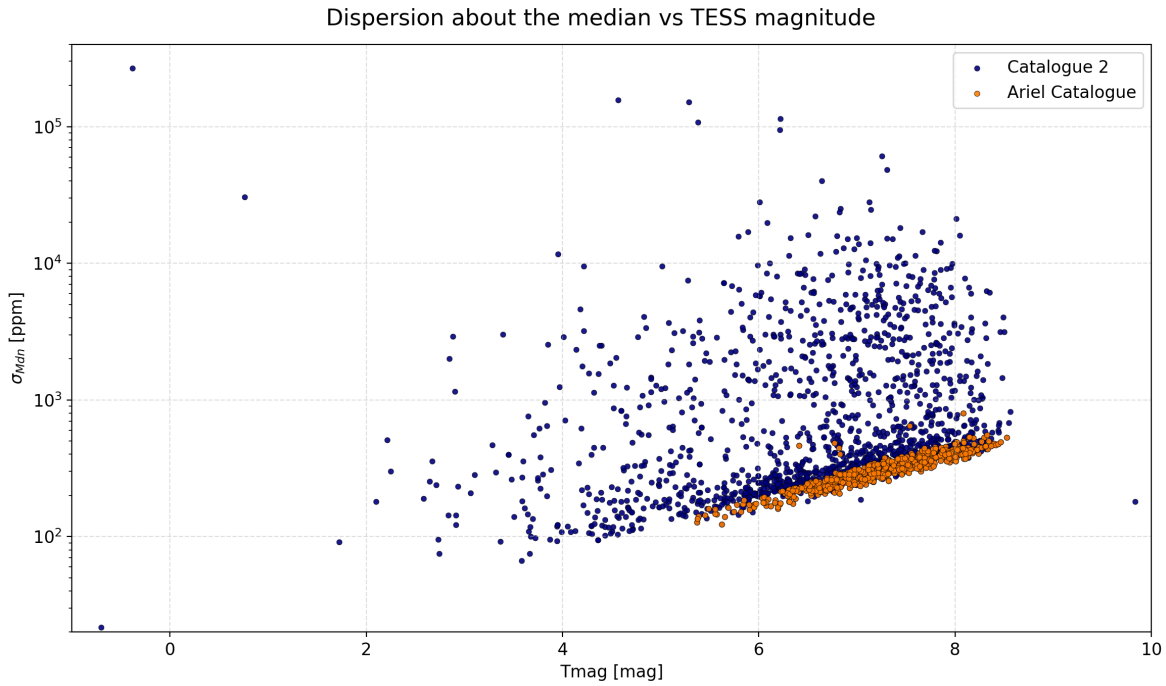


Figure 6.13: Photometric dispersion about the median calculated over the full LC as a function of the TESS magnitude.

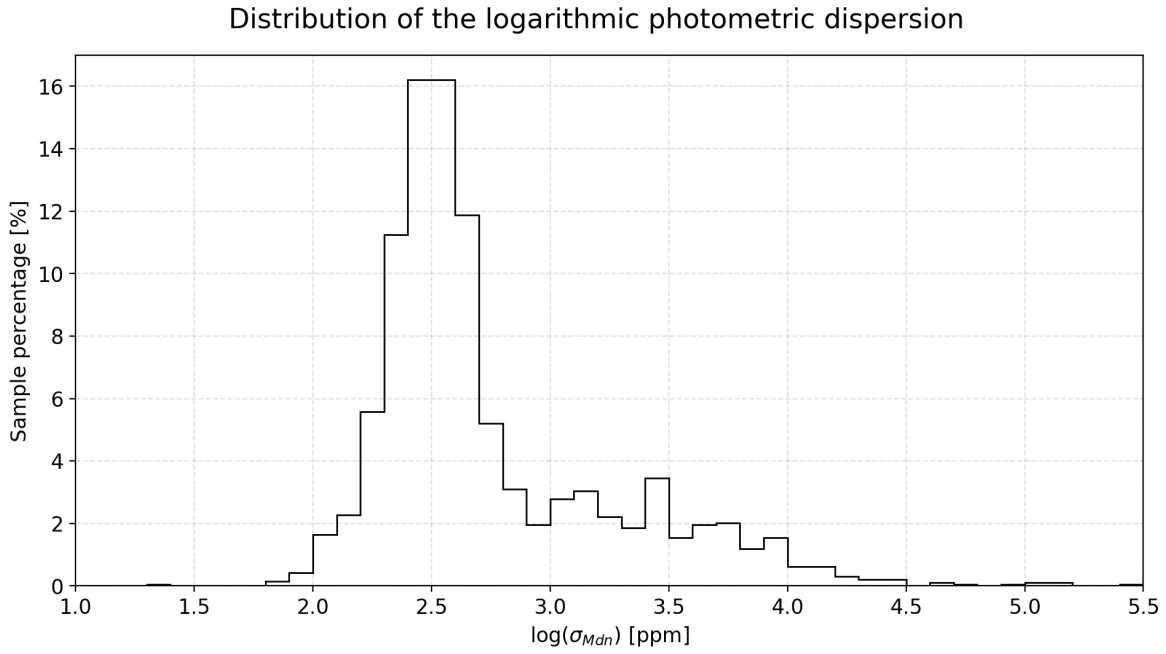


Figure 6.14: Distribution of the logarithmic photometric dispersion about the median calculated over the full LC.

Finally, Fig. 6.15 shows the variability fractions of the sample and can be compared with Fig. 10 by Ciardi et al. 2011 (Fig. 3.4 in this document). These fractions show a strong dependency with the magnitude: No stars brighter than $T_{mag} = 5mag$ are stable (the $\chi^2_v > 2$ curve

approaches unity) and, as the stars get dimmer, the variability fractions generally decrease accordingly. This decreasing trend is smooth in Ciardi et al. 2011 and caused by the degraded photometric precision of the instruments for dimmer stars. Here, the same result is also generally identified, but with some fluctuations. For instance, the curves for $\chi^2_v > 10$ and $\chi^2_v > 100$ show a very variable trend for stars with $T_{mag} < 5mag$. Moreover, all curves show a general decrease as the stars get dimmer for $T_{mag} \geq 5mag$ with some fluctuations. This can be due to a bias caused by the limited size of the sample used in this study, as well as to some physical reasons.

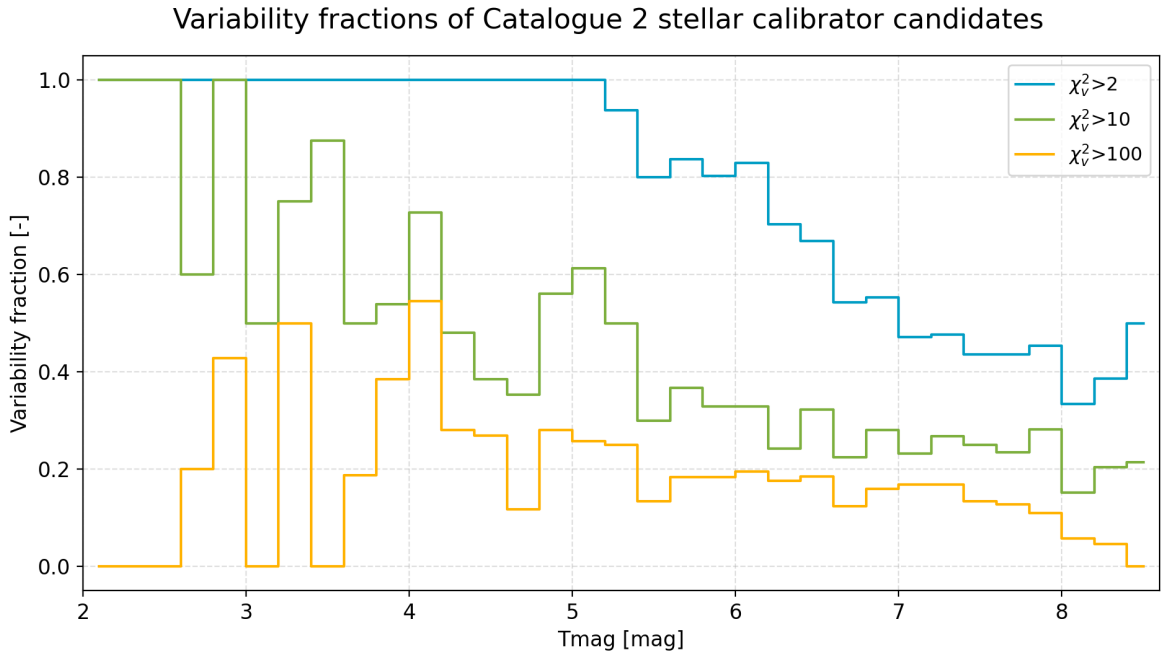


Figure 6.15: Variability fractions as a function of the TESS magnitude. The light blue curve represents the fractions of stars with $\chi^2_v > 2$, the green curve the fractions of stars with $\chi^2_v > 10$, and the orange curve the fractions of stars with $\chi^2_v > 100$.

From this comparison, it is possible to conclude that the general distributions and trends found in Ciardi et al. 2011 are also found in this study. However, whilst Ciardi et al. 2011 claimed that around 80% of all G dwarfs are stable (with $\chi^2_v < 2$), from the selection carried out in this study only around 30% of the starting Catalogue 2 sample seem to be stable.

6.4. Discussion

The selection process carried out in Section 6.2 that led to the creation of the Ariel Catalogue filtered the variable stars from the starting sample thanks to criteria on the LSP, the reduced chi-squared, and removed the stars hosting exoplanets. So, the stars that are left do not show a clear, periodic variability in their fluxes, do not have an excess dispersion of more than 1.5 times the measurement uncertainties, and are not affected by known or candidate exoplanets. It must be understood, however, that these stars are not defined as the final Ariel calibrators but are still candidates. As a good first approximation, these stars are not very variable so they represent the most promising Ariel calibrator candidates. They are mainly dominated by noise and are seen as stable by TESS, but there is no guarantee at this stage that these stars will also be seen as stable by Ariel.

This is due to two reasons:

- Selecting stars with power amplitude peaks lower than $100ppm$ in the LSP is not enough to claim their stability inside the Ariel requirement. This is because stability is also dependent on the total noise in the LC and because there could still be some structured components of lower amplitudes that sum up together, possibly increasing the variability beyond the requirement.
- The reduced chi-squared statistic is reliable in recognising if a star is seen as stable by TESS with respect to the instantaneous noise of its measurements. However, the total noise in the TESS LCs might be too high to infer anything about the stability within the Ariel requirement.

So, an additional step should be performed to further assess the noise level in the LCs and more closely relate the selection to the Ariel flux stability requirement. This step should assess the behaviour of the average flux value over the desired integration time (6 hours for the current assumed requirement) within the LC. So, the standard error previously defined in Section 5.5 could be used, as it gives such information. However, it must be made sure that the assumption of no underlying structured signal holds. As previously mentioned, after the selection, it can be assumed that the remaining stars are mainly dominated by noise. However, some of them could still possess underlying structured signals of amplitudes up to $100ppm$, or a sum of different lower-amplitude components, as well as systematic or instrumental effects that were not filtered by the pipeline. So, although the SE might be a good approximation of the behaviour of the 6-hour-long median distribution for some stars, it is not possible to assume that this holds for all of them. As shown in Fig. 5.10, for oscillations with a frequency around 0.1-1d^{-1}, which are common for stars, the SE underestimates the standard deviation of the sample median when an underlying structured signal is present in the data. The extent of the underestimation depends on the noise level in the LC and on the amplitude and frequency of the underlying signal. The standard deviation of the sample median can reach up to double the value that is estimated with the SE. This represents a problem, as stellar flux variation can be highly underestimated by the SE statistic. So, for this additional assessment, a more robust statistic is needed that would give the same information as the SE but without underestimation. This can be performed by directly computing the running median through the LC of the Ariel Catalogue stars and studying its behaviour, as explained below. This additional test is called "running median assessment".

Running median assessment

The standard deviation of the running median distribution is lower than the 1σ *flux stability requirement of Ariel over 6 hours (*$\frac{100}{3}ppm$*).*

$$\sigma_\mu(Mdn_i(6h)) < \frac{100}{3}ppm \tag{6.1}$$

where $Mdn_i(6h)$ is the running median distribution over integration times of 6 hours through the LC. This is a time series generated by calculating the median of the LC within a moving window containing 180 consecutive points. The median of the first 180 points of the LC is recorded, then the window is moved 180 points ahead, so that there is no overlap in the windows, and the median is calculated again and recorded. This is done until the window reaches the

end of the LC. The recorded median values create a time series on which some statistical quantities can be computed to analyse the behaviour of the running median distribution. For instance, its standard deviation is of great interest. If the standard deviation is small, then the average stellar flux is expected to be rather stable. If the standard deviation is large, instead, there can be a large deviation in the average flux when it is computed at different moments in time. To enable the comparison of the stellar flux variations with respect to the Ariel flux stability requirement, the highest allowed standard deviation of the running median is Ariel's 1σ requirement. This method overcomes the underestimation problem of the SE and acts as a sort of "binning" of the LC, so it partly reduces the noise.

This additional assessment is concerned with the total noise level in the LCs. It assesses if the TESS LC noise level is small enough to be able to infer something about the stability of these stars within the Ariel requirement. So, the stars that pass this additional step, will probably be seen as stable with Ariel within the desired flux variation limit. These are only 70 stars ($\sim$ 3.6% of the starting sample). However, it is not possible to claim the contrary, i.e. that the stars not passing this assessment are probably not going to be seen as stable with Ariel. In fact, not passing this assessment can simply be due to the presence of a too-large noise in the TESS LCs. The computation of the running median acts as a binning for the LC so the noise is reduced, but only partially. So, the observed variations in the median can either have a physical nature or be due to fluctuations caused by a too-high noise. In the first case, the star would not be stable within the Ariel requirement, whilst in the second case, the data is too noisy to infer anything about the star's stability within the Ariel requirement. So, only observations with a lower noise level can further assess if the Ariel Catalogue stars are stable within the Ariel flux stability requirement.

Figure 6.16 shows the positions in the sky of the Ariel Catalogue stars also passing the running median assessment.

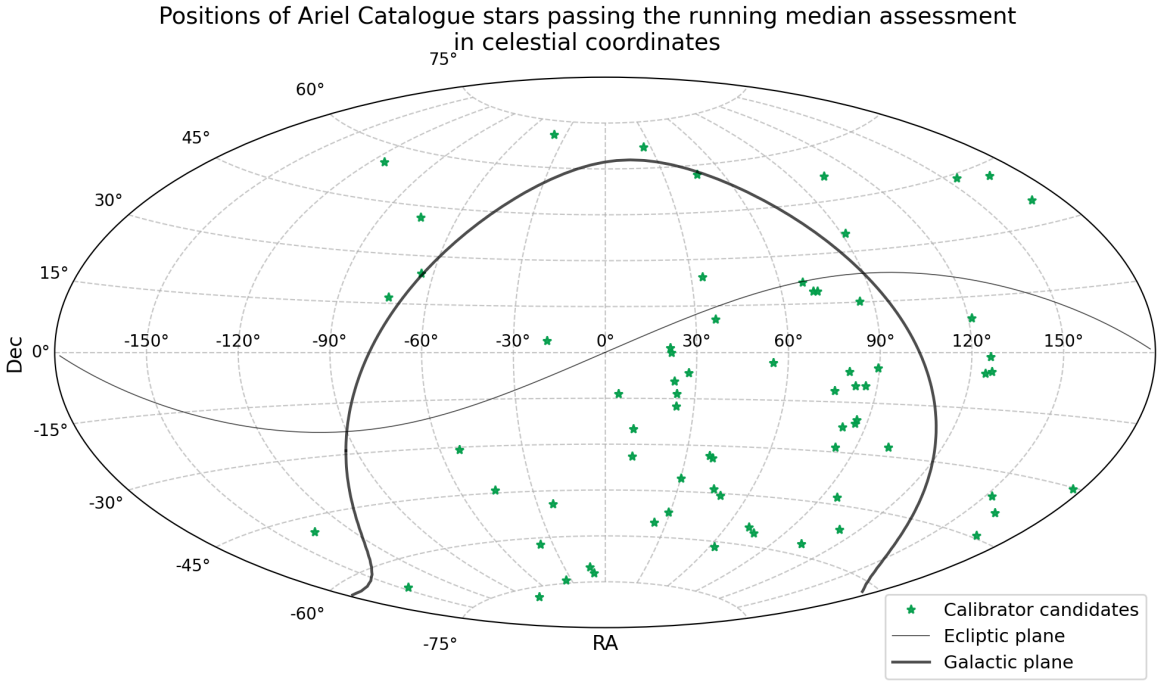


Figure 6.16: Positions in the sky of the Ariel Catalogue stars passing the running median assessment, in ICRS and epoch J2000.

The stars passing the running median assessment are mainly found at negative declination angles with respect to the ecliptic plane and seem to be clustered in restricted regions of the sky. This appears strange, as we would expect a homogeneous distribution of stable stars in the sky. This could be explained by physical reasons, like sampling specific stellar populations with a high flux stability that are clustered in precise areas. However, it could also be due to differences in the flux contamination and background noise in different sky regions. This would affect the LC noise levels, possibly making some TESS Sectors more precise than others. So, only observations from specific Sectors might possess a noise level that is small enough to identify stars within the Ariel flux stability requirement, whilst for the other Sectors all LCs possess a too-high noise and fail the running median assessment.

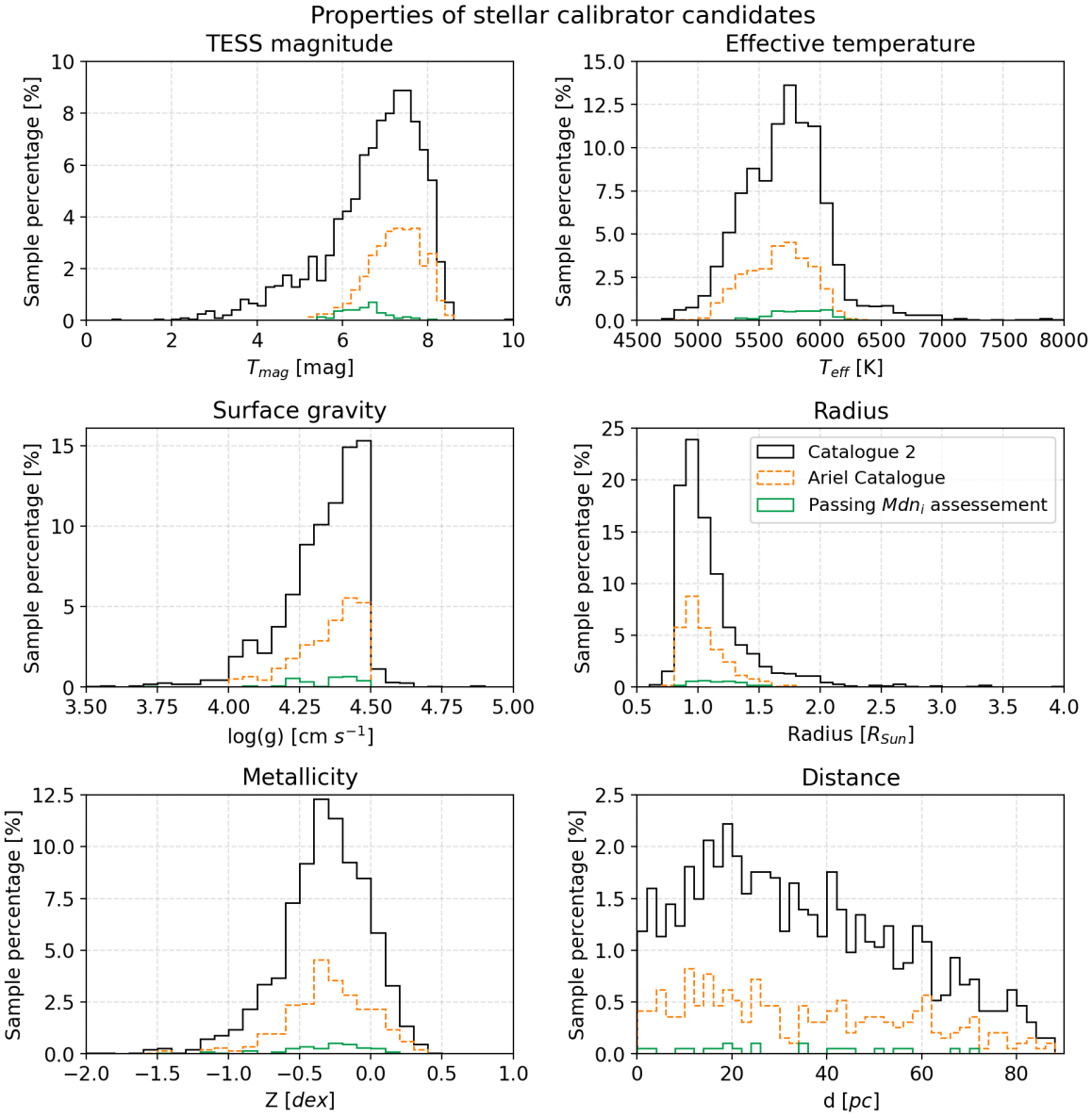


Figure 6.17: Properties of the original candidate sample, of the Ariel Catalogue stars, and of the Ariel Catalogue stars passing the running median assessment, including TESS magnitude, effective temperature, surface gravity, radius, metallicity, and distance distributions.

The noise level also depends on some stellar properties like the magnitude. To study this, Fig. 6.17 shows the distributions of some stellar properties of interest for the Ariel Catalogue stars also passing the running median assessment, with respect to the whole sample and all the Ariel Catalogue stars.

The Ariel Catalogue magnitude distribution showed a peak around $T_{mag} \simeq 7.5 mag$. The stars passing the running median assessment however seem to have smaller magnitudes (with a peak around $T_{mag} \simeq 6.5 mag$). This makes sense, as the noise level is larger for dimmer stars, so these are filtered much more by the running median assessment. This was also noticed in Fig. 5.2 depicting the TESS photometric precision with respect to the target's magnitude. With regards to the effective temperature, the Ariel Catalogue stars had a range between around $5,000$-$6,300 K$, but the stars passing the running median assessment are only found at the higher end of this range (around $5,500$-$6,300 K$). In fact, the slightly hotter stars also possess lower magnitudes. For the surface gravity instead, the best candidates are split into two small groups centered around values of 4.4 and $4.25 log(cm\ s^{-1})$, which represent the solar value and slightly smaller stars. Radius, metallicity, and distance show distributions in accordance with the ones from the Ariel Catalogue stars.

Figure 6.18 shows a sensitivity analysis on the number of stars passing the running median assessment for a varying Ariel flux stability requirement.

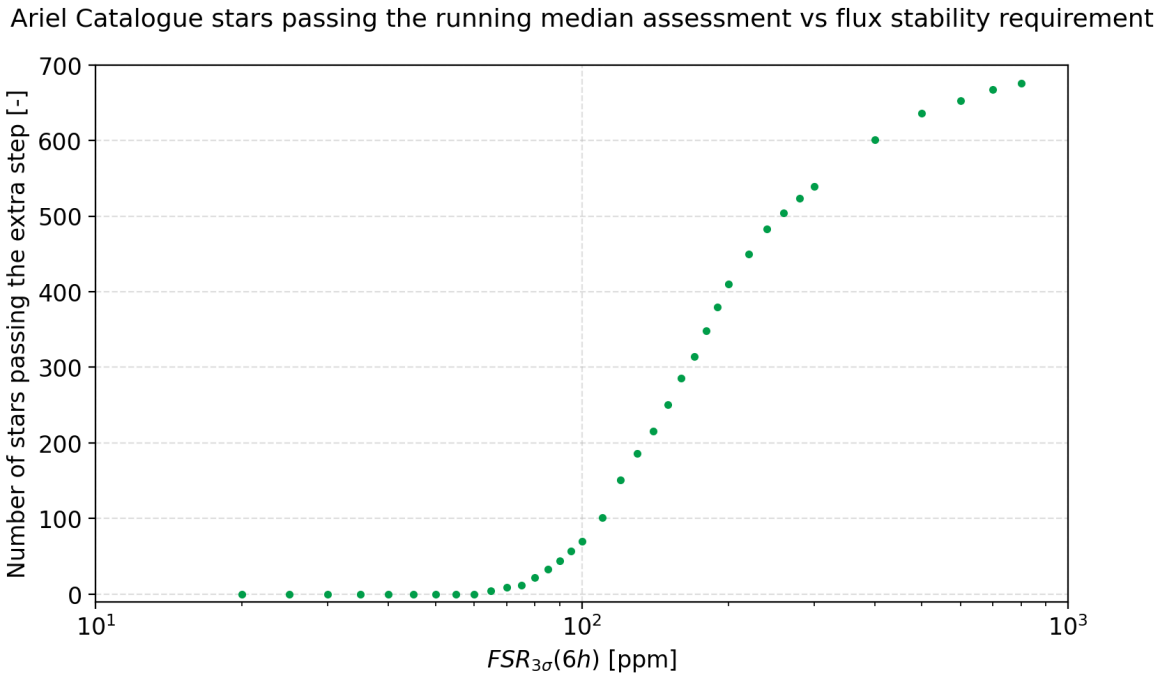


Figure 6.18: Number of stellar calibrator candidates passing the selection procedure and the additional running median assessment for different values of the flux stability requirement.

With a requirement of $100 ppm$, not many, but some suitable calibrators are still found. However, a strange selection effect was shown from their sky positions whose origin is not clear. From this sensitivity analysis, it is shown that, by relaxing the requirement, for example to $200 ppm$, many more stars pass the running median assessment. The number of stars passing the assessment rapidly increases as the requirement is relaxed, much more than it was shown

in Fig. 6.12 for the Ariel Catalogue stars. At around $FSR_{3\sigma}(6h) = 800ppm$, all the Ariel Catalogue stars pass the running median assessment. No good stellar calibrator candidates are found up to $FS_{req}(6h) = 65ppm$, which was expected as the TESS LC noise floor is around $60ppm$ (Vanderspek et al. 2018).

Figure 6.19 shows the positions in the sky of the Ariel Catalogue stars also passing the running median assessment if the Ariel flux stability requirement is relaxed to $200ppm$ over 6 hours. They are more homogeneously spread in the sky than in Fig. 6.16. This is an indication that the strange selection effect restraining stable stars in specific regions of the sky does not have a physical origin but is more probably caused by a too-high noise level in the LCs.

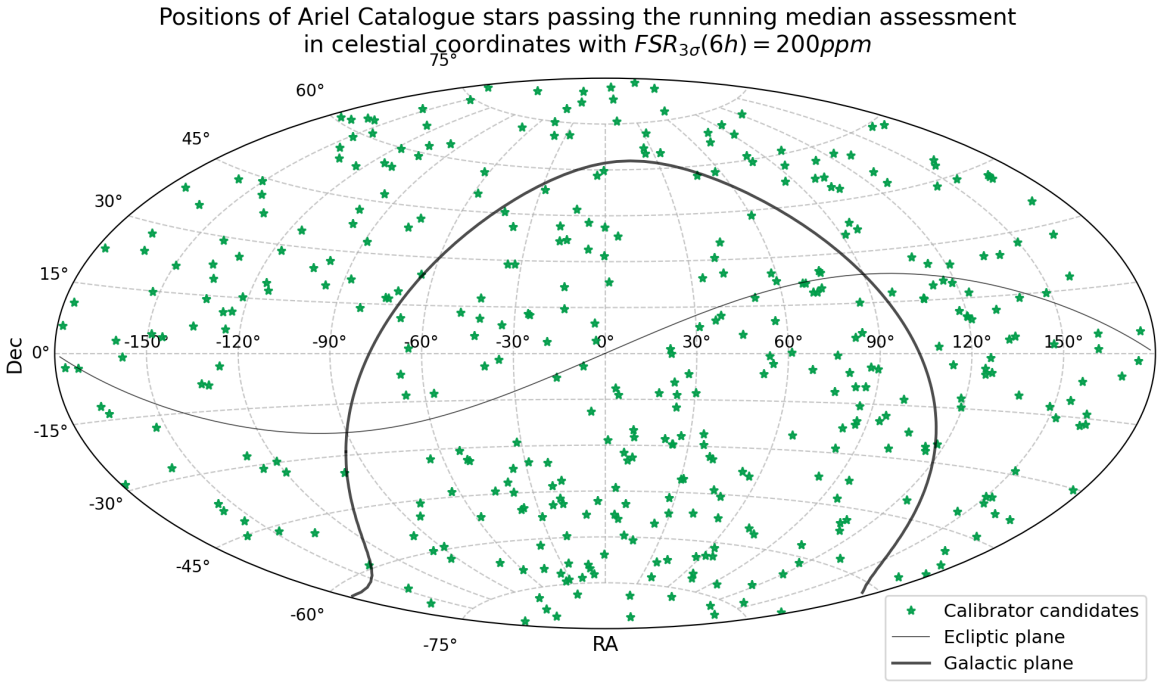


Figure 6.19: Positions in the sky of the Ariel Catalogue stars passing the running median assessment, in ICRS and epoch J2000, for a relaxed Ariel flux stability requirement of $200ppm$ over 6 hours.

Following the findings of the running median assessment, the most stable stars within the Ariel Catalogue seem to be distributed slightly towards the brighter end and the higher temperatures end of the stable ranges. However, this might also be a bias caused by the noise in the data, rather than having a physical origin. The strange selection effect that restrains stars in specific areas of the sky and the fact that this effect disappears as soon as the requirement is relaxed, gives a strong indication of an underlying non-suitability of the TESS data for this specific assessment. It is possible to claim that the Ariel Catalogue stars are "stable" for TESS standards, however, there are some difficulties in extending this claim to Ariel. Using data from a different telescope has some limitations, and noise is a very important one for this application. TESS and Ariel are very different telescopes. The tacit, underlying assumption that Ariel will have the same noise level as TESS is actually not true. The noise in TESS is generally higher. So, although it is true that a final selection criterion on the running median would be needed to complement the three-step selection process described in Section 5.6, currently this is too strict. The running median assessment is probably theoretically solid, but the underlying data is unsuitable. This is because the noise level in TESS LCs is generally too high to effectively

analyse the stellar stability within the current assumed Ariel requirement of 100*ppm* over 6 hours.

6.5. Recommendations for future work

TESS LCs and the mathematical tools used in this study, like the LSP and other statistical quantities, are effective in identifying variable stars and discarding bad calibrator candidates. However, determining if stars are stable within a certain flux stability requirement is a harder task that is mainly limited by the noise level in the data. So, it is recommended to keep under consideration all the stars identified by the selection procedure of this study that make up the so-called Ariel Catalogue. This catalogue represents the most advanced flux stability study of calibrators candidates for the Ariel mission. An additional criterion on the running median of the light curves over the desired integration time is generally needed to assess the flux stability within a specific requirement. However, it has been shown that this procedure is too strict as applied to the currently available data from TESS. To greenlight the stars in the Ariel Catalogue as final Ariel calibrators, follow-up observations with a lower noise level are required. This can allow us to further assess the stability level of the stars within the currently assumed Ariel 3σ flux stability requirement of 100*ppm* over 6 hours. Such follow-up observations could be proposed for current space-based observatories like JWST, as well as for future missions like PLAnetary Transits and Oscillations of stars (PLATO, Heras et al. 2020). This would enable us to investigate if the selected stars look stable as observed by these higher-precision telescopes. Moreover, since Ariel will only be launched in 2029, it is beneficial to monitor the stars that appear to be stable with current and past TESS observations to verify if they appear to be stable also in future Sectors, or if their behaviour has changed. Finally, if not with new observations, a comparison between past TESS and Kepler LCs can be made to further assess the behavior of the stars. 6 stars in the Catalogue 2 sample are in the Kepler field. Kepler observed those stars for a long period of time, and it is more similar to Ariel than TESS, as a telescope. Comparing LCs of the same star from TESS and Kepler could give new insights into the limitations or, on the contrary, capabilities of the TESS data.

7

Summary

This study was primarily aimed at **selecting stars to build the stellar calibrator catalogue of ESA's Ariel mission**. Such stars should be seen as stable, so as to be used as an external flux reference to calibrate Ariel's instruments. The starting sample was defined, comprising 1937 stars, and then filtered with a three-step procedure. Such steps took care of removing:

1. Stars with semi-periodic variations thanks to the Lomb-Scargle Periodogram.
2. Stars with a large excess dispersion with respect to the flux measurement uncertainties, through the use of the reduced chi-squared statistic.
3. Stars whose flux might be affected by sporadic flux variation events or phase curve effects because of any confirmed or candidate exoplanets.

Assuming an Ariel 3σ flux stability requirement of $100ppm$ over an integration time of 6 hours, the stars passing the selection are 581 ($\sim$ 30% of the starting sample). These make up the so-called Ariel Catalogue which represents the most refined stellar calibrator candidate list for the Ariel mission. The properties of these stars include being fairly homogeneously distributed over the sky, possessing a TESS magnitude in the range 5-8.5mag, an effective temperature in the range 5,000-6,300K, a surface gravity between 4-4.5$log(cm\ s^{-1})$, a radius between 0.7-1.8R_{Sun}, and metallicity and distance ranges that correspond to the ones of the starting sample. The distributions of the analysed stellar properties, when inside these ranges, appear to be extremely similar to the ones of the original sample. So, it can be concluded that the probability of finding a stable star inside those ranges is constant and around 30%. There is no specific value for any property that appears to be related to a higher probability of flux stability apart from the magnitude. This is due to the increased noise level for dimmer stars which makes it more difficult to detect flux variability. So, dimmer stars are found to be more stable, although this is probably because of a limitation in the instrument rather than having a physical explanation. The definition of these ranges and the probability of finding stable stars is an important result not only for the selection of calibrators for Ariel but also for future missions requiring stellar calibrators that could use these considerations as a starting point for their selection.

Moreover, a secondary objective was **performing a variability population study of the starting stellar calibrator candidate sample**. With regard to different variability types, the majority of the sample ($\sim$ 40%) shows a high-amplitude periodic variability that can be attributed to modulation of the light curves due to rotation. These are mainly clustered around values of

the frequency at maximum power amplitude corresponding to rotation periods between 2 and 14 days. Roughly another $\sim 45\%$ shows a lower-amplitude periodic variability or no variability. As the LSP maximum power peak decreases, the peak is found at a wider range of frequencies as the sum of random noise components. Pulsators, which seem to make up less than 2% of the sample possess the highest temperatures in the sample (7,000-8,000K) and some are also found at 5,500-6,000K. As expected, binaries and stars with exoplanet transit events ($\sim 11\%$ of the sample) do not show any correlation with the maximum power peak or its correspondent frequency. Moreover, following the analysis of some statistical quantities of interest including the dispersion about the median of the light curve and the reduced chi-squared statistic, a stability analysis of the sample was performed and compared with the findings by Ciardi et al. 2011. It was possible to confirm the general distributions and trends found by Ciardi et al. 2011, although some differences in the variability fractions were found. These are probably due to the limited size of the stellar sample used in this study with respect to the one from Ciardi et al. 2011. The main results include a bimodal distribution of the flux dispersion, with most of the stars belonging to either a low or high-dispersion group (100-1,000ppm and 1,000-10,000ppm, respectively). Furthermore, the variability fractions with respect to the stellar magnitude roughly follow the same trend of Ciardi et al. 2011 with some fluctuations. Most importantly, stars brighter than $T_{mag} = 5mag$ are always variable, and the number of stable stars found increases as the stars get dimmer. This is again probably due to the increased noise level in the LCs for faint stars.

It finally must be noted that the defined Ariel Catalogue still comprises calibrator candidates and not the final Ariel calibrators. In fact, although these stars are seen as stable by TESS, in order to infer their stability within the Ariel flux stability requirement, an additional assessment should be carried out. This should study the running median behaviour over 6-hour integration times. It was shown that by carrying out this assessment on the TESS light curves, only stars at negative declination angles with respect to the ecliptic were selected, which is a strange and unexpected result. If the Ariel requirement is relaxed to a higher flux value, this strange selection effect disappears, and many more stars, distributed homogeneously in the sky, pass the assessment. This gives a strong indication that the noise level in TESS light curves is generally too high to effectively analyse the stellar stability within the current assumed Ariel requirement of 100ppm over 6 hours. So, at this stage, the candidates cannot be properly assessed with this method. Therefore, follow-up observations with a lower noise level are needed to greenlight the Ariel Catalogue stars as final Ariel calibrators. It is recommended to use observations from current and future missions like JWST and especially PLATO to further investigate their stability.

Bibliography

Andrae, R., Schulze-Hartung, T., and Melchior, P. (2010). "Dos and don'ts of reduced chi-squared". In: *arXiv preprint*. DOI: 10.48550/arXiv.1012.3754.

Anthony, L. (2003). "The Cambridge dictionary of statistics". In: *Ref. Rev.* 17, pp. 29–30. DOI: 10.1108/09504120310455993.

Auvergne, M., Bodin, P., Boisnard, L., et al. (2009). "The CoRoT satellite in flight: description and performance". In: *A&A* 506, pp. 411–424. DOI: 10.1051/0004-6361/200810860.

Balona, L. (2017). "Starspots on A stars". In: *Monthly Notices of the Royal Astronomical Society* 467, pp. 1830–1837. DOI: 10.1093/mnras/stx265.

Bohlin, R. C., Gordon, K. D., and Tremblay, P.-E. (2014). "Techniques and review of absolute flux calibration from the ultraviolet to the mid-infrared". In: *PASP* 126, p. 711. DOI: 10.1086/677655.

Bohlin, R., Gordon, K., Rieke, G., et al. (2011). "Absolute flux calibration of the IRAC instrument on the Spitzer Space Telescope using Hubble Space Telescope flux standards". In: *AJ* 141, p. 173. DOI: 10.1088/0004-6256/141/5/173.

Breger, M. (2000). *Delta Scuti stars (Review)*. ASP Conf. Series, 210.

Cardoso, J. V. d. M., Hedges, C., Gully-Santiago, M., et al. (2018). *Lightkurve: Kepler and TESS time series analysis in Python*. Astrophysics Source Code Library, p.1812.

Ciardi, D. R., Von Braun, K., Bryden, G., et al. (2011). "Characterizing the variability of stars with early-release Kepler Data". In: *AJ* 141, p. 108. DOI: 10.1088/0004-6256/141/4/108.

Claytor, Z. R., Saders, J. L. van, Llama, J., et al. (2022). "Recovery of TESS Stellar Rotation Periods Using Deep Learning". In: *AJ* 927, p. 219. DOI: 10.3847/1538-4357/ac498f.

Deustua, S., Kent, S., and Smith, J. A. (2012). *Absolute Calibration of Astronomical Flux Standards*. Vol. 375. Springer. DOI: 10.1007/978-94-007-5618-2_8.

Edwards, B. and Tinetti, G. (2022). "The Ariel Target List: The Impact of TESS and the Potential for Characterizing Multiple Planets within a System". In: *AJ* 164, p. 15. DOI: 10.3847/1538-3881/ac6bf9.

Fetherolf, T., Pepper, J., Simpson, E., et al. (2022). "Variability Catalog of Stars Observed During the TESS Prime Mission". In: *ASS*. DOI: 10.48550/arXiv.2208.11721.

Gardner, J. P., Mather, J. C., Clampin, M., et al. (2006). "The James Webb Space Telescope". In: *Space Sci. Rev.* 123, pp. 485–606. DOI: 10.48550/arXiv.astro-ph/0606175.

Gautschy, A. and Saio, H. (1995). "Stellar pulsations across the HR diagram: Part 1". In: *ARAA* 33, pp. 75–113. DOI: 10.1146/annurev.aa.33.090195.000451.

Gilliland, R. L. (2008). "Photometric oscillations of low-luminosity red giant stars". In: *AJ* 136, p. 566. DOI: 10.1088/0004-6256/136/2/566.

Gordon, K. D., Bohlin, R., Sloan, G., et al. (2022). "The James Webb Space Telescope Absolute Flux Calibration. I. Program Design and Calibrator Stars". In: *AJ* 163, p. 267. DOI: 10.3847/1538-3881/ac66dc.

Grigahcène, A., Antoci, V., Balona, L., et al. (2010). "Hybrif γ Doradus–δ Scuti pulsators: New insights into the physics of the oscillations from Kepler observations". In: *ApJL* 713, p. L192. DOI: 10.1088/2041-8205/713/2/L192.

Hayes, D. S. and Latham, D. (1975). *A rediscussion of the atmospheric extinction and the absolute spectral-energy distribution of Vega*. ApJ, 197, p.593.

Heras, A. M., Rauer, H., Aerts, C., et al. (2020). "The PLATO mission: Studying the diversity of exoplanets orbiting up to the habitable zone of Sun-like stars". In: *EPSC*, p. 396. DOI: 10.5194/epsc2020-396.

Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. (2010). "Initial characteristics of Kepler long cadence data for detecting transiting planets". In: *The Astrophysical Journal Letters* 713, p. L120. DOI: 10.1088/2041-8205/713/2/L120.

Jenkins, J. M., Twicken, J. D., McCauliff, S., et al. (2016). "The TESS science processing operations center". In: *Software and Cyberinfrastructure for Astronomy IV*. Vol. 9913. SPIE, pp. 1232–1251. DOI: 10.1117/12.2233418.

Karttunen, H., Kröger, P., Oja, H., et al. (2007). *Fundamental astronomy*. Springer.

Kaye, A. B., Handler, G., Krisciunas, K., et al. (1999). "γ Doradus stars: defining a new class of pulsating variables". In: *PASP* 111, p. 840. DOI: 10.1086/316399.

Kirk, B., Conroy, K., Pra, A., et al. (2016). "Kepler eclipsing binary stars. VII. The catalog of eclipsing binaries found in the entire Kepler data set". In: *AJ* 151, p. 68. DOI: 10.3847/0004-6256/151/3/68.

Kjeldsen, H. and Bedding, T. R. (1994). "Amplitudes of stellar oscillations: the implications for asteroseismology". In: DOI: 10.48550/arXiv.astro-ph/9403015.

Kjeldsen, H. and Frandsen, S. (1992). "High-precision time-resolved CCD photometry". In: *PASP* 104, p. 413. DOI: 10.1086/133014.

Koch, D. G., Borucki, W. J., Basri, G., et al. (2010). "Kepler mission design, realized photometric performance, and early science". In: *ApJL* 713, p. L79. DOI: 10.1088/2041-8205/713/2/L79.

Lomb, N. R. (1976). "Least-squares frequency analysis of unequally spaced data". In: *Ap&SS* 39, pp. 447–462. DOI: 10.1007/BF00648343.

Ma, S., Ali, E., Lv, C., et al. (2022). "KIC 5768203: A δ Sct Pulsator Modulated by Rotation and Spots". In: *AJ* 164, p. 22. DOI: 10.3847/1538-3881/ac6fde.

McQuillan, A., Mazeh, T., and Aigrain, S. (2014). "Rotation periods of 34,030 Kepler main-sequence stars: the full autocorrelation sample". In: *ApJS* 211, p. 24. DOI: 10.1088/0067-0049/211/2/24.

Mugnai, L. V., Pascale, E., Edwards, B., et al. (2020). "ArielRad: the Ariel radiometric model". In: *Exp.Astron.* 50, pp. 303–328. DOI: 10.1007/s10686-020-09676-7.

Mullally, S. E., Sloan, G., Hermes, J., et al. (2022). "Searching for TESS Photometric Variability of Possible JWST Spectrophotometric Standard Stars". In: *AJ* 163, p. 136. DOI: 10.3847/1538-3881/ac4bce.

NExScI (2020). *Planetary Systems Table*. DOI: 10.26133/NEA12.

Oke, J. and Schild, R. (1970). *The absolute spectral energy distribution of Alpha Lyrae*. ApJ, 161, p.1015.

Pascale, E., Eccleston, P., and Tinetti, G. (2018). "The ARIEL space mission". In: *5th IEEE Int. Workshop on MetroAeroSpace*. IEEE, pp. 31–34. DOI: 10.1109/MetroAeroSpace.2018.8453588.

Pearson, C., Malaguti, G., Sarkar, S., et al. (2022). "The Ariel ground segment and instrument operations science data centre: Organization, operation, calibration, products and pipeline". In: *Exp.Astron.* 53, pp. 773–806. DOI: 10.1007/s10686-020-09691-8.

Petralia, A., Micela, G., et al. (2017). *Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL), Assessment Phase Payload Study, SKY CALIBRATORS FOR ARIEL*. ARIEL-INAF-SCI-TN-0003 Issue 0.1.

Price-Whelan, A. M., Lim, P. L., Earl, N., et al. (2022). "The Astropy Project: sustaining and growing a community-oriented open-source project and the latest major release (v5. 0) of the core package". In: *AJ* 935, p. 167. DOI: 10.3847/1538-4357/ac7c74.

Pra, A., Kochoska, A., Conroy, K. E., et al. (2022). "TESS Eclipsing Binary Stars. I. Short-cadence Observations of 4584 Eclipsing Binaries in Sectors 1–26". In: *ApJS* 258, p. 16. DOI: 10.3847/1538-4365/ac324a.

Prusti, T., De Bruijne, J., Brown, A. G., et al. (2016). "The Gaia mission". In: *A&A* 595, A1. DOI: 10.1051/0004-6361/201629272.

Reach, W. T., Megeath, S., Cohen, M., et al. (2005). "Absolute Calibration of the Infrared Array Camera on the Spitzer Space Telescope". In: *PASP* 117, p. 978. DOI: 10.1086/432670.

Ricker, G. R., Latham, D., Vanderspek, R., et al. (2010). "Transiting Exoplanet Survey Satellite (TESS)". In: *AAS* 215, p. 450. DOI: 10.1117/1.JATIS.1.1.014003.

Savitzky, A. and Golay, M. J. (1964). "Smoothing and differentiation of data by simplified least squares procedures." In: *AnaCh* 36, pp. 1627–1639. DOI: 10.1021/ac60214a047.

Scargle, J. D. (1982). *Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data*. AJ, 263, pp.835–853.

Schuster, A. (1898). "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena". In: *Terrestrial Magnetism* 3, pp. 13–41. DOI: 10.1029/TM003i001p00013.

Skrutskie, M., Cutri, R., Stiening, R., et al. (2006). "The Two Micron All Sky Survey (2MASS)". In: *AJ* 131, p. 1163. DOI: 10.1086/498708.

Stassun, K. G., Oelkers, R. J., Paegert, M., et al. (2019). "The revised TESS input catalog and candidate target list". In: *AJ* 158, p. 138. DOI: 10.3847/1538-3881/ab3467.

Stone, R. P. (1996). "Spectrophotometry of flux calibration stars for Hubble Space Telescope". In: *AJS* 107, p. 423. DOI: 10.1086/192369.

Tenenbaum, P. and Jenkins, J. M. (2018). *TESS Science Data Products Description Document: EXP-TESS-ARC-ICD-0014 Rev D*. Tech. rep.

Tinetti, G., Beaulieu, J.-P., Henning, T., et al. (2012). "EChO: Exoplanet Characterisation Observatory". In: *Exp.Astron.* 34, pp. 311–353. DOI: 10.1007/s10686-012-9303-4.

Tinetti, G., Eccleston, P., Haswell, C., et al. (2021). "Ariel: Enabling planetary science across light-years". In: *arXiv preprint*. DOI: 10.48550/arXiv.2104.04824.

Uytterhoeven, K., Moya, A., Grigahcène, A., et al. (2011). "The Kepler characterization of the variability among A-and F-type stars-I. General overview". In: *A&A* 534, A125. DOI: 10.1051/0004-6361/201117368.

Vallenari, A., Brown, A., and Prusti, T. (2022). "Gaia Data Release 3. Summary of the content and survey properties". In: *A&A*. DOI: 10.1051/0004-6361/202243940.

VanderPlas, J. T. (2018). "Understanding the Lomb–Scargle Periodogram". In: *ApJS* 236, p. 16. DOI: 10.3847/1538-4365/aab766.

Vanderspek, R., Doty, J., Fausnaugh, M., et al. (2018). *TESS Instrument Handbook*. Version 0.1.

Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2020). "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Waldmann, I., Pascale, E., Swinyard, B., et al. (2013). "Photometric stability analysis of the Exoplanet Characterisation Observatory". In: DOI: 10.48550/arXiv.1302.6425.

Wamsteker, W. (1993). *The International Ultraviolet Explorer Project*. UNPSA, 4, p.173.

Wenger, M., Ochsenbein, F., Egret, D., et al. (2000). "The SIMBAD astronomical database-The CDS reference database for astronomical objects". In: *ApJS* 143, pp. 9–22. DOI: 10.1051/aas:2000332.

Werner, M. W., Roellig, T., Low, F., et al. (2004). "The Spitzer Space Telescope Mission". In: *ApJS* 154, p. 1. DOI: 10.1086/422992.

Winn, J. N. (2010). “Transits and occultations”. In: *arXiv preprint*. DOI: 10.48550/arXiv.1001.2010.

A

Stellar Calibrator Catalogues

All these catalogues and all LC plots with their associated LSP are available for all the stars from the starting sample Catalogue 2 at this link.

A.1. EChO list

First EChO stellar calibrator candidates, *by J.P. Beaulieu and A. Moneti, 2013* (private communication).

Column descriptions
0. Calibration star: Arbitrary ID of the EChO list.
1. R.A.: Right ascension in ICRS at the reference epoch J2000.0. Unit: $[h : min : s]$
2. Dec: Declination in ICRS at the reference epoch J2000.0. Unit: $[° : ' : '']$
3. mag (J): J band magnitude. Unit: $[mag]$
4. mag (H): H band magnitude. Unit: $[mag]$
5. mag (K_s): K_s band magnitude. Unit: $[mag]$

All values come from 2MASS.

A.2. Catalogue 1

Updated stellar calibrator candidates list with Gaia DR3, *by J.P. Beaulieu and L. Bernard, IAP, 2022* (private communication).

Column descriptions
0. source_id: Gaia DR3 ID of the source.
1. ra: Right ascension in ICRS at the Gaia DR3 reference epoch (J2016.0). Unit: [°]
2. dec: Declination in ICRS at the reference epoch. Unit: [°]
3. parallax: Source's parallax in ICRS at the reference epoch. Unit: $[mas]$
4. pmra: Proper motion in the right ascension direction in ICRS at the reference epoch. Unit: $[mas\ yr^{-1}]$
5. pmdec: Proper motion in the declination direction in ICRS at the reference epoch. Unit: $[mas\ yr^{-1}]$
6. phot_g_mean_mag: mean magnitude in the G band ($\sim$330-1050nm). Unit: $[mag]$
7. bp_rp: Colour in the G-band. Unit: $[mag]$

8. bp_g: Colour in Gaia's blue band (∼330-700nm). Unit: [mag]
9. g_rp: Colour in Gaia's red band (∼600-1050nm). Unit: [mag]
10. phot_variable_flag: Flag indicating if photometric variability was identified. The available flags are 'NOT_AVAILABLE' - source not processed or exported to the catalogue, 'CONSTANT' - source not identified as variable, and 'VARIABLE' - course identified as variable. Note that in this data release, many known variable sources were not processed so they are flagged as 'NOT_AVAILABLE'.
11. non_single_star: Flag indicating if the source is in a binary system. It assumes the value of 0 - single star or not a known binary, 1 - astrometric binary, 2 - spectroscopic binary, or 3 - eclipsing binary.
12. teff_gspphot: Effective temperature. Unit: [K]
13. logg_gspphot: Surface gravity. Unit: [$log(cm\ s^{-1})$]
14. mh_gspphot: Iron abundance. Unit: [dex]
15. distance_gspphot: Distance of the observer to the source. Unit: [pc]

The parameters 6, 7, 8, 9, and 10 are inferred by GSP-Phot Aeneas best library using BP/RP spectra.

A.3. Catalogue 2

Unified catalogue including EChO list and Catalogue 1 stars with at least one TESS light curve observation, *by E. Tonucci (TU Delft, SRON), 2023.*

Column descriptions
0. C2_index: Position of the star within Catalogue 2.
1. EChO_Candidate_ID: Arbitrary ID from the EChO list, if available.
2. GAIA_ID: ID from Gaia DR3, if available.
3. TIC: TESS Input Catalog ID.
4. RA: Right ascension in ICRS at the epoch J2000.0. Unit: [°]
5. Dec: Declination in ICRS at the epoch J2000.0. Unit: [°]
6. pmRA: Proper motion in the right ascension direction in ICRS at the epoch J2000.0. Unit: [$mas\ yr^{-1}$]
7. pmDec: Proper motion in the declination direction in ICRS at the epoch J2000.0. Unit: [$mas\ yr^{-1}$]
8. plx: Parallax in ICRS at the epoch J2000.0. Unit: [mas]
9. Tmag: Magnitude in the TESS band. Unit: [mag]
10. Gmag: Magnitude in the Gaia band. Unit: [mag]
11. Vmag: Magnitude in the V band. Unit: [mag]
12. Jmag: Magnitude in the J band. Unit: [mag]
13. Hmag: Magnitude in the H band. Unit: [mag]
14. Kmag: Magnitude in the K band. Unit: [mag]
15. d: Distance of the observer to the source. Unit: [pc]
16. Teff: Effective temperature of the source. Unit: [K]
17. mass: Mass of the source. Unit: [M_{Sun}]
18. radius: Radius of the source. Unit: [R_{Sun}]
19. logg: Surface gravity of the source. Unit: [$log(cm\ s^{-1})$]
20. metallicity: Iron abundance of the source. Unit: [dex]
21. non_single_star: Flag indicating if the source is in a binary system. It assumes the value of 0 - single star or not a known binary, 1 - astrometric binary, 2 - spectroscopic binary, or 3 - eclipsing binary.

22. N_obs: Number of available TESS observations of the source.
23. sector: TESS Sector number of the observation used in this study.
24. mean: Mean value of the light curve time series observation. Unit: $[e^- \ s^{-1}]$
25. Mdn: Median value of the light curve time series observation. Unit: $[e^- \ s^{-1}]$
26. sigma: Standard deviation of the light curve time series observation. Unit: $[e^- \ s^{-1}]$
27. sigma_ppm: Standard deviation of the light curve time series observation in a normalised unit. Unit: $[ppm]$
28. sigma_Mdn: Dispersion around the median of the light curve time series observation. Unit: $[e^- \ s^{-1}]$
29. sigma_Mdn_ppm: Dispersion around the median of the light curve time series observation in a normalised unit. Unit: $[ppm]$
30. chi2v: Reduced chi-squared computed around the median of the light curve time series observation.
31. Pmax: Maximum value of the LSP normalised power amplitude spectrum. Unit: $[ppm]$
32. f_Pmax: Frequency value corresponding to the maximum value of the LSP normalised power amplitude spectrum. Unit: $[d^{-1}]$
33. C1_flag: Flag indicating if the source passes the first criterion of the Ariel calibrator selection process. 0 - if the star does not comply with the criterion, 1 - if the star complies with the criterion.
34. C2_flag: Flag indicating if the source passes the second criterion of the Ariel calibrator selection process. 0 - if the star does not comply with the criterion, 1 - if the star complies with the criterion.
35. C3_flag: Flag indicating if the source passes the third criterion of the Ariel calibrator selection process. 0 - if the star does not comply with the criterion, 1 - if the star complies with the criterion.
36. Ariel_flag: Flag indicating if the source passes the full Ariel calibrator selection process. 0 - if the star does not pass the selection, 1 - if the star passes the selection, which means it is included in the Ariel Catalogue.

The stellar parameters from column 4 to column 18 in this catalogue are stored as defined for TESS, and their specific source can be found in the TIC v7 Data Release Notes[1]. Columns 19 to 21 come from Gaia DR3. The others are defined or computed in this study.

A.4. Ariel Catalogue

Catalogue containing the stellar calibrator candidates from Catalogue 2 passing the three-step selection process defined in this study *by E. Tonucci (TU Delft, SRON), 2023.*

Column descriptions
0. Ariel_Candidate_ID: Arbitrary ID from the Ariel Catalogue.
1. C2_index: Position of the star within Catalogue 2.
2. EChO_Candidate_ID: Arbitrary ID from the EChO list, if available.
3. GAIA_ID: ID from Gaia DR3, if available.
4. TIC: TESS Input Catalog ID.
5. RA: Right ascension in ICRS at the epoch J2000.0. Unit: [°]
6. Dec: Declination in ICRS at the epoch J2000.0. Unit: [°]
7. pmRA: Proper motion in the right ascension direction in ICRS at the epoch J2000.0. Unit: $[mas \ yr^{-1}]$

[1] https://outerspace.stsci.edu/display/TESS/TIC+v7+and+CTL+v7.xx+Data+Release+Notes

8. pmDec: Proper motion in the declination direction in ICRS at the epoch J2000.0. Unit: $[mas\ yr^{-1}]$
9. plx: Parallax in ICRS at the epoch J2000.0. Unit: $[mas]$
10. Tmag: Magnitude in the TESS band. Unit: $[mag]$
11. Gmag: Magnitude in the Gaia band. Unit: $[mag]$
12. Vmag: Magnitude in the V band. Unit: $[mag]$
13. Jmag: Magnitude in the J band. Unit: $[mag]$
14. Hmag: Magnitude in the H band. Unit: $[mag]$
15. Kmag: Magnitude in the K band. Unit: $[mag]$
16. d: Distance of the observer to the source. Unit: $[pc]$
17. Teff: Effective temperature of the source. Unit: $[K]$
18. mass: Mass of the source. Unit: $[M_{Sun}]$
19. radius: Radius of the source. Unit: $[R_{Sun}]$
20. logg: Surface gravity of the source. Unit: $[log(cm\ s^{-1})]$
21. metallicity: Iron abundance of the source. Unit: $[dex]$
22. non_single_star: Flag indicating if the source is in a binary system. It assumes the value of 0 - single star or not a known binary, 1 - astrometric binary, 2 - spectroscopic binary, or 3 - eclipsing binary.
23. N_obs: Number of available TESS observations of the source.
24. sector: TESS Sector number of the observation used in this study.
25. mean: Mean value of the light curve time series observation. Unit: $[e^-\ s^{-1}]$
26. Mdn: Median value of the light curve time series observation. Unit: $[e^-\ s^{-1}]$
27. sigma: Standard deviation of the light curve time series observation. Unit: $[e^-\ s^{-1}]$
28. sigma_ppm: Standard deviation of the light curve time series observation in a normalised unit. Unit: $[ppm]$
29. sigma_Mdn: Dispersion around the median of the light curve time series observation. Unit: $[e^-\ s^{-1}]$
30. sigma_Mdn_ppm: Dispersion around the median of the light curve time series observation in a normalised unit. Unit: $[ppm]$
31. chi2v: Reduced chi-squared computed around the median of the light curve time series observation.
32. Pmax: Maximum value of the LSP normalised power amplitude spectrum. Unit: $[ppm]$
33. f_Pmax: Frequency value corresponding to the maximum value of the LSP normalised power amplitude spectrum. Unit: $[d^{-1}]$
34. Mdn_i_flag: Flag indicating if the source passes the additional running median assessment. 0 - if the star does not pass the assessment, 1 - if the star passes the assessment.

This is the most refined Ariel stellar calibrator candidate list.