



**UNIVERSITÀ
DI TORINO**

Università degli Studi di Torino
Master Degree in Astrophysics and Theoretical Physics

**Digging in the dirt:
Measuring planetary fundamental
parameters in the presence of
stellar activity.
The case of TOI-396.**

Irene Amateis
Registration number: 858028

Supervisor
Prof. Davide Gandolfi
Co-Supervisor
Dr. Andrea Bonfanti

Examiner
Prof. Stefano Camera

October 2023, Turin

Contents

Abstract	5
1 Introduction	7
1.1 Context	7
1.1.1 Other exoplanets detection methods	8
1.2 Aims of this thesis	10
2 Astrophysical Background	11
2.1 Parameters of a Keplerian Orbit	11
2.2 Transit Method	14
2.2.1 Limb Darkening	17
2.2.2 The Rossiter-McLaughlin Effect	18
2.2.3 Transit Timing and Transit Duration Variations	19
2.3 Radial Velocity Method	21
2.3.1 Calculating the Doppler Shift: the Cross-correlation Function	24
2.3.2 Analysis of RV Time Series: the Periodogram	26
2.4 Stellar Activity	29
2.4.1 Stellar Oscillations	29
2.4.2 Stellar Granulation and Supergranulation	29
2.4.3 Short-term Stellar Activity	30
2.4.4 Effects of Short-term Stellar Activity on radial velocities	31
2.4.5 Long-term Stellar Activity	32
3 Methods	35
3.1 Cross-correlation Function Analysis	35
3.1.1 Normal Fit	35
3.1.2 Skew Normal Fit	37
3.1.3 Other techniques for analysing the CCF	38
3.2 RV detrending: the breakpoint method	40
3.3 Markov chain Monte Carlo	41
3.3.1 The MCMCI code	42
3.3.2 Merit function	44
3.3.3 Steps to be performed	46

3.3.4	Input files	46
4	The Planetary System TOI-396	49
4.1	System Overview	49
4.2	Data analysis	50
4.2.1	Radial velocity setup	50
4.2.2	Photometry setup	51
5	Results	55
5.1	Stellar parameters	55
5.2	Detrending baseline	55
5.3	Setup of the <code>mcmc.dat</code>	56
5.4	MCMC Analysis assuming linear ephemerides	58
5.5	Discussion on the mass non-detection of TOI-396 c	59
5.5.1	Hypothesis 1: TOI-396 c is a false-positive	61
5.6	MCMC Analysis assuming TTV	63
5.6.1	Radial velocity	66
5.6.2	Photometry	71
5.6.3	TOI-396 in the exoplanetary context	72
5.7	Hypothesis 2: Stellar activity covers the RV signal of TOI-396 c	72
5.7.1	Test1: adding artificial RV signals with $P = P_c$	83
5.7.2	Test2: adding artificial RV signals with $P = P_{\text{rot}}$	84
6	Conclusions	87
	Bibliography	88

Abstract

The passage of a planet in front of its host star provides us with one of the two fundamental ingredients needed to investigate its nature: the planetary radius. The fraction of light occulted by a planet during a transit is indeed proportional to its size, being about 1% for a Jupiter-sized object and 0.01% for an Earth-sized planet transiting a Sun-like star. The second fundamental quantity, the planetary mass, can be obtained by observing the wobble of the star around the centre of mass, i.e., by measuring the periodic change in the stellar velocity component along the line-of-sight (the so-called "radial velocity") while the star revolves about the centre of mass. This requires the observations of the tiny Doppler shifts of the stellar spectral lines induced by the presence of an orbiting companion.

The knowledge of masses and radii is of paramount importance for exoplanetology, as it allows us to distinguish between gas giants ($\rho \simeq 1 \text{ g cm}^{-3}$), ice giants ($\rho \simeq 1\text{--}3 \text{ g cm}^{-3}$), and terrestrial worlds with or without atmospheric envelopes ($\rho > 3 \text{ g cm}^{-3}$), enabling the study of the internal structure and composition of planets and therefore providing important hints as to planet formation and evolution.

However, stellar magnetic activity (photospheric spots, plages, and flares) coupled with stellar rotation, can also induce quasi-periodic radial velocity variations that can have magnitudes greater than the reflex motion induced by an orbiting planet, hampering its mass determination, or even preventing its detection.

In the course of my MSc thesis work, I focused on the determination of the planetary fundamental parameters in the presence of stellar magnetic activity. Specifically, I characterised the planetary system orbiting the naked-eye star TOI-396 (V=6), a system known to host 3 transiting small planets discovered by NASA's TESS space mission (Vanderburg et al., 2019). I measured the radii, masses, bulk densities, and orbital parameters of the three planets, performing a joint analysis of the TESS transit light curves and the radial velocity measurements collected with the high-precision HARPS spectrograph. I disentangled the activity-induced radial velocity signals from the genuine planetary motions using a novel technique (Simola et al., 2019), which implements a skew-normal fit to a proxy of the average stellar line profile, the so-called cross correlation function. I also detrended the radial velocity time-series using the "break-point" method applied to ancillary activity indicators (Simola et al., 2022). I finally

utilized the code developed by Bonfanti & Gillon (2020) to derive the posterior distributions of the model parameters, using Markov chain Monte Carlo methods.

Nota bene

- The work presented in my MSc thesis work was carried out as part of a five-month traineeship at the Institut für Weltraumforschung (IWF) in Graz, Austria, under the supervision of Dr. Andrea Bonfanti. I was awarded a scholarship by the "Erasmus+ for traineeship" project at the University of Turin.
- The characterization of the TOI-396 planetary system presented in this work will be published in a refereed paper, which will include Irene Amateis as co-author.

Chapter 1

Introduction

1.1 Context

The observations of planets in the Solar System is a very ancient discipline that goes back to the dawn of human history. We know that our planetary system hosts four small rocky planets in its innermost region, between ~ 0.4 and 1.5 au^1 from the Sun (Mercury, Venus, Earth, and Mars), two gas giant planets (Jupiter and Saturn) at 5.2 and 9.6 au from the Sun, and two ice giant planet, namely Neptune and Uranus, with semi-major axis of 19.2 and 30.2 au , respectively.

For long time, the Solar System has been thought to be the standard planetary system template in our Galaxy. The recent discoveries of planets in orbit around other stars – also known as *extrasolar planets* or *exoplanets* – have uncovered several types of planetary objects, such as hot Jupiters (Dawson & Johnson, 2018) and super-Earths (Haghighipour, 2013), which are not represented in our own Solar System. The picture that is emerging is of a realm of exoplanets that challenge our understanding of planetary system formation and evolution, unveiling Nature’s infinite creativity to form worlds with different sizes, masses, and orbital parameters, spectacularly exceeding the variety of the properties of Solar System’s planets.

The two fundamental quantities needed to investigate the nature of these fascinating “new worlds” are the planetary radius and mass. The former can be measured using the transit method, while the latter can be obtained via the radial velocity technique (Seager & Mallén-Ornelas, 2003, Winn, 2010, Hatzes, 2019). Once the planetary radius and mass are known, we can determine the planetary bulk density, which gives fundamental information on the internal structure and composition of planets, as well as important hints as to their formation and evolution (Winn & Fabrycky, 2015, Hatzes, 2016).

The transit method (see Sec. 2.2 for a comprehensive and thorough discussion) is based on the idea that when a planet is orbiting around the centre of

¹The astronomical unit (au) is a unit of length and is equal to the mean distance between the Earth and the Sun ($\sim 1.5 \times 10^8 \text{ km}$).

mass of the star-planet system, it periodically passes between the observer and the star, if the plane of the orbit is seen edge-on. When this happens, a fraction of the light coming from the star is blocked, resulting in a small decrease in the star's brightness. The magnitude of this flux decrease is proportional to the relative size of the planet with respect to its host star.

The principle behind the radial velocity (to which we will simply refer to as RV; see Sec. 2.3 for more details) method is the Doppler effect. The light emitted (or reflected) from an object with a non-null relative radial velocity with respect to the observer² undergoes the Doppler effect. If the object is approaching the observer, the frequencies of the light waves are increased and the corresponding wavelengths become shorter (blue-shift). If the object is receding, frequencies are reduced and wavelengths become longer (red-shift). High resolution spectra of the light emitted (or reflected) from an object will show Doppler blue- or red-shifted spectral lines (i.e., lines shifted towards the blue or the red wavelengths) if the source is moving towards the observer or away from it, respectively. By comparing the positions of the observed spectral lines with the positions where we would expect to see these lines in the laboratory-rest frame we can measure the RV of the object. An orbiting planet induces periodic RV variation of the star due to the Doppler reflex motion of the latter around the star-planet centre of mass. This is the key to reveal an orbiting planet via the Doppler method and measure its minimum mass³.

There are, however, some difficulties in detecting planets via the RV method. The periodic Doppler reflex motion induced by an orbiting planet on the host star is small. The RV semi-amplitude variation is $\sim 95 \text{ m s}^{-1}$ for a Jupiter-mass planet on a 10-d, edge-on, circular orbit, and only $\sim 0.3 \text{ m s}^{-1}$ for an Earth-mass planet on the same orbit. It is therefore mandatory to acquire RV measurements with very high precision, using highly stable and state-of-the-art spectrographs. Moreover, stellar activity (e.g., pulsations, spot-induced variability) can induce apparent RV shifts whose magnitude is greater than the Doppler reflex motion induced by the orbiting planets, hampering the planet detection via the RV method. It is therefore crucial to develop techniques to disentangle the genuine Doppler motion induced by the planet from the unwanted signals associated with stellar activity so that we can detect the planet, and measure its minimum mass by modelling the detrended RV data.

1.1.1 Other exoplanets detection methods

The transit method and RV methods are the more successful detection techniques in terms of number of discovered planets⁴. However, other methods have

²The radial velocity is the velocity component along the line of sight to the object.

³The minimum mass is given by $M_p \times \sin i$, where M_p is the planet's true mass and i is the inclination of the orbital plane with respect to the plane of the sky, i.e., the plane perpendicular to the line of sight from Earth to the object. For a transiting planet, the orbit is seen almost edge-on ($i \approx 90^\circ$) and $M_p \times \sin i \approx M_p$.

⁴See the following link from the NASA Exoplanet Archive: https://exoplanetarchive.ipac.caltech.edu/exoplanetplots/exo_dischist_cumulative.png.

been developed and will be briefly described in the following paragraphs (see Perryman, 2018, for a comprehensive description).

- *Astrometry.* Generally speaking, astrometry is the measurement of positions and motions of celestial bodies in the sky. It is used to measure parallaxes⁵ and proper motions⁶. The change in the proper motion can be used to detect planets by observing the wobble of the star’s photocentre about the centre of mass. This method consists in measuring the transverse component of the displacement of the host star due the gravitational perturbation of an orbiting planets. In this regard astrometry is complementary to the RV technique, which observes instead the motion of an object along the line of sight. ESA’s Gaia space mission (Gaia Collaboration et al., 2022) is currently searching for exoplanets using the astrometric method, among other goals.
- *Gravitational microlensing.* It is based on the effect predicted by Albert Einstein in his theory of general relativity. From general relativity we know that a massive object is able to bend light. If the right alignment occurs, light travelling towards an observer from a distant star can be bent around an intervening star, which acts as a “gravitational lens”. If the intervening star hosts an orbiting planet, both star and planet can act as lenses, with the latter causing what we call a “microlensing” effect. Due to the relative motion of the background star, planetary system, and observer, the magnification of the brightness of the background source changes with time as the alignment varies between the source, lens, and observer. A star and its planet acting as lenses will produce a double-peaked magnification of the background star’s flux, unveiling the presence of the orbiting planet. This method allows us to discovery very distant planetary systems, but has the great disadvantage that a microlensing does not repeat itself again, due to the low geometric probability of reobserving an alignment. This implies that planets detected using the microlensing method cannot be reobserved again.
- *Direct imaging.* It refers to the detection of the point source image of an exoplanet. Directed imaging typically uses infrared wavelengths to minimize the star-to-planet flux ratio and increase the probability of detecting

⁵Generally speaking, the parallax is the apparent shift of an object’s position relative to more distant background objects caused by a change in the observer’s position. As the Earth orbits the Sun, the position of nearby stars will appear to slightly shift against more distant background objects. These shifts are angles in an isosceles triangle, with 2 au making the base leg of the triangle and the distance to the star being the height of the triangle. The parallax angle π is the angle subtended by the Earth’s orbital radius as seen from the star. The parallax angle provides the distance to an object via the equation $d = 1/\pi$, where π is measured in arcseconds, and d in parsecs.

⁶The proper motion is the angular velocity of an object due its motion with respect to the centre of mass of the Solar System. It is a measure of the observed changes in the apparent position of an object in the sky, as seen from the centre of mass of the Solar System, compared to more distant celestial bodies whose proper motion is negligible.

an exoplanet. As a matter of fact, at optical wavelengths the light reflected/emitted by a planet is usually drowned out by the massive amounts of radiation coming from its host star. Unlike the indirect methods listed above, direct imaging enables the direct detection of an exoplanet. This technique works best for young planetary systems, whose planets are hot and bright, as they are still contracting and radiating away the energy from their formation. By studying real images and spectra of exoplanets, we can find out what the planetary atmospheres are made of. With the current observation technology, direct imaging is possible on very rare occasions, as it works only for nearby systems hosting very long-period planets, due to the large angular separation between the planet and its host star.

1.2 Aims of this thesis

My master of science (MSc) thesis work focuses on the characterization of the three transiting small planets orbiting around TOI-396. The host star is a relatively young F-type star, with a projected equatorial velocity⁷ of $v_s \sin i_s \approx 8.3 \text{ km s}^{-1}$, showing strong magnetic activity. The primary objective of my thesis work is to detrend the RV measurements acquired with the high-precision HARPS spectrograph, as part of a Doppler follow-up project of transiting small planets carried out by the Physics Department of the University of Turin. I aim at detecting the Doppler reflex motion induced by the three transiting planets and measuring their masses.

The TOI-396 planetary system was discovered and announced by Vanderburg et al. (2019), who detected three different sets of transit signals in the NASA's *TESS* light curve of TOI-396. The three planets are about twice the size of the Earth and have orbital periods of ~ 3.6 , 6.0 , and 11.2 days. Combining *TESS* transit photometry with *HARPS* RV measurements I made a step forward and measured the planetary mass, radius, and bulk density of the three planets, enabling the first characterization of their internal structure and composition.

⁷The projected equatorial velocity is the modulus of the equatorial velocity of the star projected along the line of sight. It is given by $v_s \sin i_s$, where v_s is the magnitude of the equatorial velocity, while i_s is the inclination of the stellar spin with respect to the line of sight to the star.

Chapter 2

Astrophysical Background

2.1 Parameters of a Keplerian Orbit

In this section I will briefly present the parameters that are needed to identify a Keplerian orbit, specifically referring to elliptical orbits¹. Let's start by describing the *relative motion* of a planet about its host star, which is located at one of the two foci of the ellipse. To define the orbital elements of a Keplerian orbit we need a reference frame, which is given by a reference plane and a reference direction. With reference to Fig. 2.1, we identify:

- *Reference plane*: the plane tangent to the celestial sphere at the position of the star. This is the so-called plane of the sky.
- *Reference direction*: the direction pointing towards the Vernal Equinox point Υ , i.e., the point on the ecliptic where the Sun crosses from the southern celestial hemisphere to the northern. The point Υ lies in the intersection between the equatorial plane of the Earth and the ecliptic plane (i.e., the plane that contains the orbit of the Earth around the Sun).
- *Orbital plane*: the plane of the planetary and stellar orbits, which forms the angle i with respect to the reference plane.
- *Periastron* (resp. *Apastron*): the point at which the planet is closest (resp. farthest) to (resp. from) its stellar host.
- *Line of nodes*: the line defined by the intersection of the planet's orbital plane with the reference plane. The planet orbit intersects the plane of the sky in two points, namely, the ascending and descending nodes, as described below.
- *Ascending node*: point where the planet crosses the reference plane while moving away from the observer.

¹The mechanic energy of an elliptical orbit is negative.

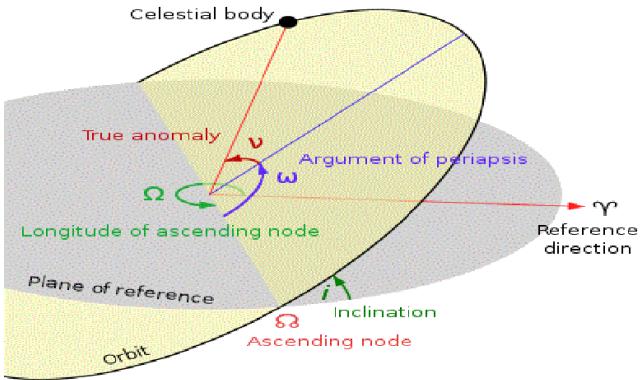


Figure 2.1: Keplerian orbital elements (Zamaro, 2011).

- *Descending node*: point where the planet crosses the reference plane while moving towards the observer.

Let us now define the so called *Keplerian elements*, which are the quantities needed to uniquely parameterize a Keplerian orbit:

- a : semi-major axis. It is the semi-distance between the periastron and apastron. The semi-major axis a defines the size of the orbit.
- e : eccentricity. It is defined as $e = c/a$, where c is the distance between the centre of the ellipse and one of its foci, and a is the semi-major axis. For an elliptical orbit $0 < e < 1$. For a circular orbit $e = 0$. The eccentricity e describes the shape of the orbit and how much it is elongated compared to a circumference.
- Ω : longitude of the ascending node. It is the angle, measured on the plane of the sky and whose vertex is at the star, between the direction of the Vernal equinox Υ , and the direction of the ascending node. It defines the orientation of the orbit in the plane of the sky and can assume any values in the range $0 \leq \Omega < 360^\circ$.
- ω : argument of periastron. Measured on the orbital plane and with its vertex at the star, the argument of periastron is the angle between the direction towards the ascending node and the semi-major axis containing the periastron. It defines the orientation of the orbit on the orbital plane and can assume any value in the range $0 \leq \omega < 360^\circ$. For a circular orbit the argument of periastron of a star is conventionally set to 90° .
- i : orbital inclination. It is the angle between the orbital plane and the plane of the sky. If $i = 0^\circ$ the orbit is seen face-on, i.e., tangential to the celestial sphere. If $i = 90^\circ$ the orbit is seen edge-on.

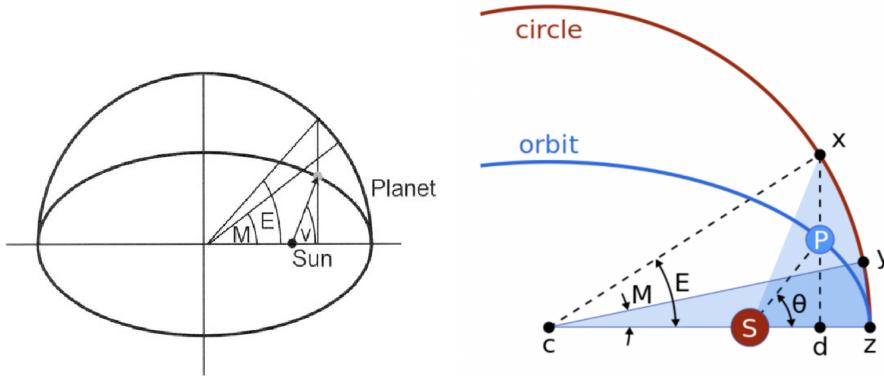


Figure 2.2: True (ν on the left and θ on the right panel), eccentric (E) and mean anomaly (M). Credit: left panel from Varvoglis (2006) and right panel from Wikipedia.

- $\nu(t)$: true anomaly. The angle whose vertex is at the star, measured on the orbital plane, between the direction of the periastron and the position of the planet at the time (epoch) t .

To parameterize the motion of a planet along its orbit, we further need

- P : orbital period, that is the time taken by the planet to complete one orbit.
- T_0 : the time (epoch) of periastron passage.

Since it is not possible to directly express the true anomaly $\nu(t)$ as a function of time t , the temporal dependence of $\nu(t)$ is usually parameterized using two additional angles, namely, the eccentric anomaly $E(t)$ and the mean anomaly $M(t)$. With reference to Fig. 2.2, let us consider a planet, graphically represented by \textcircled{P} on an elliptical orbit around its host star \textcircled{S} (e.g. the Sun), which is in the focus of the ellipse. The periastron is z , the ellipse centre is c , its eccentricity is e , while its semi-major and semi-minor axes are a and b , respectively. We start drawing a circle whose centre coincides with c and whose radius is equal to the semi-major axis of the ellipse. The perpendicular to the semi-major axis passing through \textcircled{P} intersects the circle in x . The eccentric anomaly of \textcircled{P} is then defined by the angle $E = \widehat{zcx}$. The corresponding mean anomaly $M = \widehat{zcy}$ is defined such that the area A_c of the circular sector zcy is $A_c = \frac{a}{b}A_e$, where A_e is the area of the elliptical sector $z\textcircled{S}x$. In other words, y is the position that the planet would have if it orbited its host at a uniform rate with the same orbital period P of \textcircled{P} on the eccentric orbit. As such, the mean anomaly $M(t)$ is defined as

$$M(t) = \frac{2\pi}{P}(t - T_0). \quad (2.1)$$

$M(t)$ is linked to $E(t)$ through Kepler equation, that is

$$M(t) = E(t) - e \sin E(t), \quad (2.2)$$

while $\nu(t)$ depends on $E(t)$ through

$$\tan \frac{\nu(t)}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E(t)}{2} \quad (2.3)$$

2.2 Transit Method

The transit method has played a fundamental role in exoplanet discoveries and, more generally speaking, in exoplanetary science. This technique has under its name the largest number of exoplanets discovered so far. This has been possible mainly thanks to NASA's Kepler mission launched in 2009 and retired in 2018².

Let us consider a planet orbiting around its host star, where the two bodies are expected to have very unequal sizes. We call transit the passage of the smaller body (e.g., the planet) in front of the larger body (e.g., the star) and occultation the passage of the smaller body behind the larger body. Both these events can be generically defined as eclipses. Sometimes we have grazing eclipses, in which the bodies' disks do not overlap completely.

A very fortunate situation occurs when the planet has a nearly edge-on orbit with respect to the line of sight (i.e., $i \simeq 90^\circ$). In this case, we observe the host star and the planet periodically eclipsing each other, causing a periodic decrease in the total flux. The plot showing the stellar flux as a function of time (time series) as the planet transits in front of the stellar disk is the so-called *transit light curve*. When the planet transits its host star, it produces a U-shaped curve (Fig. 2.3).

Let us consider a star of radius R_s and mass M_s , and a planet of radius R_p and mass M_p . Following Winn (2010), it is useful to introduce the quantity $k = \frac{R_p}{R_s}$, i.e., the planet-to-star radius ratio. The impact parameter b is the sky-projected distance between the star and the planet at conjunction (i.e., when they reach the minimum sky-projected distance). It is a dimensionless quantity normalized to R_s . For a circular orbit, it is given by

$$b = \frac{a}{R_s} \cos i. \quad (2.4)$$

For an eccentric orbit, we should include in Eq. (2.4) a corrective factor that accounts for the non-null eccentricity e . The transit happens if $|b| < 1 + k$.

As in Murray & Correia (2010), we choose a coordinate system centred on the star, with the sky in the X–Y plane and the +Z axis pointing at the observer (see Fig. 2.3). Since the orientation of the line of nodes relative to reference direction (i.e., the direction pointing towards the Vernal Equinox point) is usually unknown and of limited interest, we align the X axis with the line of nodes. We

²https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html.

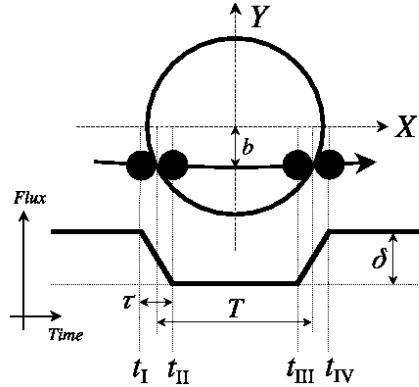


Figure 2.3: Transit of a planet (Winn, 2010).

also place the descending node of the planet's orbit along the +X axis, giving $\Omega_p = 180^\circ$. Using Newton's law of universal gravitation, we can derive the following equation for the distance between the star and planet

$$r(t) = \frac{a(1 - e^2)}{1 + e \cos \nu(t)}, \quad (2.5)$$

where a is the semi-major axis of the relative orbit, e is the orbit eccentricity, and ν is the true anomaly. Eq. (2.5) is the equation of an ellipse in polar coordinates.

For non grazing transits, there are four points where stellar and planetary disks are tangent. These points correspond to four different contact times (see Fig. 2.3), which define:

- the total duration of the transit: $T_{\text{tot}} = t_{\text{IV}} - t_{\text{I}}$
- the ingress duration: $\tau_{\text{ing}} = t_{\text{II}} - t_{\text{I}}$
- the egress duration: $\tau_{\text{egr}} = t_{\text{IV}} - t_{\text{III}}$

It is also useful to define, in reference to Fig. 2.3, $T = T_{\text{tot}} - \tau$ (where τ represents either the ingress or egress duration). Computing T is not trivial, but if we assume $e \rightarrow 0$, $R_p \ll R_s \ll a$ and $b \ll 1 - k$, then T can be expressed as (Murray & Correia, 2010)

$$T \approx \bar{T} \sqrt{1 - b^2}, \quad (2.6)$$

where \bar{T} is the characteristic time scale defined as:

$$\bar{T} \equiv \frac{R_s P}{\pi a} \quad (2.7)$$

This time scale refers to the transit duration for a central event ($b = 0$). Eq. (2.6) and (2.7) suggest that accurately and precisely measuring the T -related time

intervals is important for having a good estimate of the impact parameter b . Moreover, by inverting Eq. (2.4) we can obtain the scaled semi-major axis $\frac{a}{R_s}$, that is indeed a transit observable. Knowing $\frac{a}{R_s}$ is useful because we can derive the mean stellar density ρ_s via Kepler's third law of planetary motion, which can be written as:

$$\rho_s + k^3 \rho_p = \frac{3\pi}{GP^2} \left(\frac{a}{R_s} \right)^3 \quad (2.8)$$

Since k^3 (i.e., the planet-to-star radius ratio to the third power) is usually very small, the second term on the left-hand side containing the planetary mean density ρ_p can be neglected and ρ_s can be computed purely from transit photometry.

An additional source of information for characterizing transiting planets is the flux drop. During the non-eclipsing phase, the total flux $F(t)$ coming from the system is the superposition of the stellar flux $F_s(t)$ and the planetary flux $F_p(t)$. Given their temperature, planets emit very little light of their own (they rather reflects the stellar light), and can be treated as blackbodies whose emission essentially peaks in the infrared. The resulting total flux $F(t)$ decreases both during the transit phase as the planet is partially covering the star, and during the occultation when the planet is behind the star. Conceptually, we can describe $F(t)$ as

$$F(t) = F_s(t) + F_p(t) - \begin{cases} k^2 \alpha_{\text{tra}}(t) F_s(t) & \text{transit} \\ 0 & \text{outside eclipses} \\ \alpha_{\text{occ}}(t) F_p(t) & \text{occultation} \end{cases} \quad (2.9)$$

where the α coefficients are dimensionless functions of order unity, which quantify the fraction of overlapping area between the planetary and stellar disks. Both F_s and F_p may vary in time for different reasons. For example, F_s may change due to flares or spots, while F_p changes because of the so-called phase curve, i.e., the change in the amount of light reflect by the planet along its orbit due to the variation of the planetary disk's fraction illuminated by the star (Parmentier & Crossfield, 2018).

With reference to Fig. 2.3, we can introduce the transit depth, which is the relative flux variation, i.e., the difference between the out-of-transit and in-transit fluxes, normalized to the out-of-transit flux

$$\delta_{\text{tra}} = \frac{F_{\text{out-of-transit}} - F_{\text{in-transit}}}{F_{\text{out-of-transit}}} = \frac{\Delta F}{F_{\text{out-of-transit}}} \quad (2.10)$$

During the full phase of the transit $\alpha_{\text{tra}} = 1$. Entering Eq. (2.9) into Eq. (2.10), we can derive the maximum loss of light

$$\delta_{\text{tra}} \approx k^2 \left[1 - \frac{F_p(t_{\text{tra}})}{F_s} \right], \quad (2.11)$$

which accounts for the $F_p \ll F_s$ condition. If we neglect F_p , we can write

$$\delta_{\text{tra}} \approx k^2 \equiv \left(\frac{R_p}{R_s} \right)^2, \quad (2.12)$$

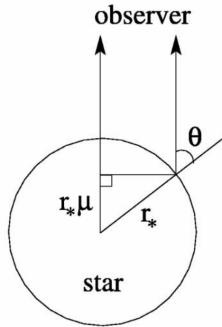


Figure 2.4: Geometry of limb darkening. Star is seen edge-on, with the observer off the top of the page. The star has radius r_* and θ is defined as the angle between the observer and the normal to the stellar surface, while $\mu = \cos \theta$ (Mandel & Agol, 2002).

which shows that the transit depth allows us to derive the planet-to-star radius ratio. If the stellar radius is known (e.g., from stellar evolutionary models, or broad-band photometry combined with parallax and spectroscopic parameters) we can determine the planetary radius.

2.2.1 Limb Darkening

Limb darkening (LD) is an optical effect seen on stars, where the central part of the disk appears brighter than the edge, (also referred to as the limb). When modelling transit light curves, it is important to account for this effect as it affects the shape and the depth of the event (Csizmadia et al., 2013). Limb darkening is usually modelled as a function of both wavelength λ and $\mu = \cos \theta$, where θ is the angle between the normal to the stellar surface and the line-of-sight to the star (see Fig. 2.4).

Different functions have been proposed to model the specific intensity $I_X(\mu)$ for a given bandpass X as a function of μ . Each choice presents advantages and disadvantages as discussed in Maxted (2018). Here we recall the most common laws, whose coefficients, which depend upon the stellar spectral type, define the LD coefficients:

- Quadratic limb-darkening: $I_X(\mu) = 1 - c_1(1 - \mu) - c_2(1 - \mu)^2$
- Non linear limb-darkening: $I_X(\mu) = 1 - \sum_{j=1}^4 a_j(1 - \mu^{j/2})$
- Power-2 limb-darkening: $I_X(\mu) = 1 - c(1 - \mu^\alpha)$.

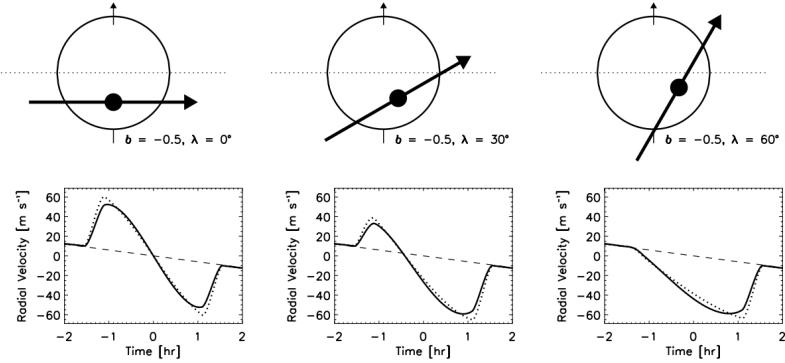


Figure 2.5: Variations of the anomalous RV for different values of the sky-projected spin-orbit angle λ (Gaudi & Winn, 2007)

2.2.2 The Rossiter-McLaughlin Effect

The following description of the Rossiter-McLaughlin effect is based on the work of Winn (2010) and Triaud (2018). The Rossiter-McLaughlin (RM) effect appears as an anomalous radial velocity variation of a rotating star during a transit event. If a star rotates, half of its visible disk approaches the observer, while the other half recedes. Thus the light coming from the first half is blue-shifted, while the light from the second half is red-shifted and this broadens the spectral lines. When a planet is transiting, it progressively covers different parts of the stellar disk. If it covers part of the blue-shifted side, the overall light from the star is slightly red-shifted and vice versa. This causes a time-dependent distortion in the star spectrum, which translates into an anomalous measurement of the stellar radial velocity.

The shape of the deformed spectral lines gives important information about the star and the planet. The maximum amplitude of the anomalous radial velocity is approximately:

$$\Delta RV_{RM} \approx k^2 \sqrt{1 - b^2} (v_s \sin i_s) \quad (2.13)$$

where $v_s \sin i_s$ is the line-of-sight component of the stellar equatorial rotational velocity, being i_s the angle between the stellar spin axis and the line-of-sight. As expected, ΔRV_{RM} depends on the relative dimension between the two bodies in the system through k . But more interestingly it gives an independent measure of the projected rotational velocity $v_s \sin i_s$.

The RM effect can also be employed to measure the sky-projected angle between the planet's orbital plane and the star's equatorial plane (that is the angle between the stellar spin axis and the angular momentum vector), often referred to as λ . Fig. 2.5 shows three trajectories of a transiting planet that have the same impact parameter (hence they produce the same light curves), but different values for λ , which produce different RV time series. If $\lambda = 0^\circ$, the

transiting planet generates a symmetric RV signal (left panel), while a strongly misaligned planet that blocks, for example, only the receding part of the star produces a blue-shifted signal (right panel).

2.2.3 Transit Timing and Transit Duration Variations

In the presence of a multi-planetary system, perturbations of the planetary Keplerian orbits can occur as a consequence of the mutual gravitational pull induced by the planets on one another. If the planets transit their host star, such mutual perturbations can cause the time and the duration of transits to vary. Following Agol et al. (2005) and Agol & Fabrycky (2018), I will herewith briefly discuss two phenomena that arise from planet-planet interaction: the transit timing variation (TTV) and the transit duration variation (TDV), which are of particular interest as they have been used for detecting, confirming, and measuring the masses of transiting planets. The first planet discovered using gravitation means has been Neptune in our Solar System. Adams (1846) and Le Verrier (1877) were able to predict the presence of Neptune by observing the perturbations (in terms of accelerations and decelerations) of Uranus.

Perturbations on the Keplerian orbits specifically generate changes to the instantaneous orbital semi-major axis a , eccentricity e , and longitude of periastron ω . If we consider a constant-period model, where the orbit is not perturbed, the time of a transit is given by

$$C = T_0 + nP \quad (2.14)$$

where n is an integer number, called epoch, T_0 is the reference transit time at $n = 0$ and P is the orbital period. C stands for "calculated" transit time. In case TTVs are significant, the transit time as computed in Eq. (2.14) would differ from the observed transit time, to which we will refer to as O . The difference between the observed transit time and the calculated one, $O - C$, is the so-called TTV.

Similarly to TTVs, TDVs are caused by changes in a , e , and ω , but also by orbital plane re-orientations, as the inclination of the planetary orbit affects the length of the transit cord. As such, if i_p changes, the amount of time the planet spends in front of its host stars also changes and, consequently, the transit duration too.

Let us take a deeper look at the physics behind TTVs. The gravitational interaction between planets and the host star occur on orbital timescales. As a consequence, the magnitude of TTVs must be proportional to the orbital period of each planet. We can also easily assume that the magnitude will also be a function of the mass of the bodies that generate the gravitational interaction on a given planet. Let us consider a two-planet system. The planets are here referred to as 1 e 2 according to the proximity to their host star, with planet 1 being the closest one. To the lowest order in mass ratio, $O - C$ is given by

$$\delta t_1 = P_1 \frac{M_2}{M_s} f_{12}(\alpha_{12}, \theta_{12}) \quad \delta t_2 = P_2 \frac{M_1}{M_s} f_{21}(\alpha_{12}, \theta_{21}) \quad (2.15)$$

where M_s is the stellar mass, whereas M_1 and M_2 are the masses of planet 1 and 2, and P_1 and P_2 their orbital periods. f_{ij} is a function that describes the perturbations of planet j on planet i and it depends on the semi-major axis ratio $\alpha_{ij} = \min(a_i/a_j, a_j/a_i)$ and on the angular orbital elements of the interacting planets, therefore $\theta_{ij} = (\lambda_i, e_i, \omega_i, I_i, \Omega_i, \lambda_j, e_j, \omega_j, I_j, \Omega_j)$. Different papers report different estimates for these functions, e.g. Nesvorný & Morbidelli (2008), Nesvorný (2009), Nesvorný & Beaugé (2010), Agol & Deck (2016), Deck & Agol (2016).

In a system of multiple perturbing planets, we can approximate the TTVs as the linear combinations of the perturbations due to every other planet:

$$\delta t_i = P_i \sum_{j \neq i} \frac{M_j}{M_s} f_{ij}(\alpha_{ij}, \theta_{ij}) \quad (2.16)$$

Note that this approximation is valid if the ratios of the planetary masses to the stellar mass are sufficiently small, and if none of the pairs of planets are in mean-motion resonance³.

For planets in mean-motion resonance, the TTV amplitudes are generally greater. We can compute such amplitudes by imposing that the two planets are trading energy with each others, while the total energy must be conserved. Following the formalism presented in Agol & Fabrycky (2018), assuming that the changes in the orbital periods are small ($\delta P_i \ll P_i$), we can conclude that

$$\delta P_2 = -\delta P_1 \left(\frac{M_1}{M_2} \right) \left(\frac{P_2}{P_1} \right)^{5/3} \quad (2.17)$$

or equivalently:

$$\delta t_2 = -\delta t_1 \left(\frac{M_1}{M_2} \right) \left(\frac{P_2}{P_1} \right)^{2/3} \quad (2.18)$$

Consequently, the TTV curves of two planets in mean-motion resonance are anti-correlated. If the two planets have similar masses, the outer planet is subject to a greater TTV as the orbit dimension needs to change more so that its Keplerian orbital energy can balance the change in the Keplerian orbital energy of the inner planet.

One final remark on the importance of TTVs is the fact that they provide a mean of measuring the mass and the density of planets. As described in Sec. 2.2, the transit method provides the planetary radius once the stellar radius is known. If two or more transiting planets are subject to mutual gravitational interaction, we can estimate the radius ratio of the planets to the star. Using both information we can infer the density of the planets.

³A mean-motion orbital resonance occurs when two bodies have periods of revolution that are a simple integer ratio of each other.

2.3 Radial Velocity Method

The radial velocity (RV) method is playing a fundamental role in exoplanet studies as it enables us to determine the planetary mass once the mass of the host star and orbit inclination are known. It is the second most prolific method for discovering exoplanets, after the transit method. As of September 2023, about 20% of all confirmed exoplanets have been discovered through the RV technique and $\sim 76\%$ exploiting the transit method⁴. The transit method is usually combined with the RV technique because together they provide the planetary radius, mass, and bulk density, as well as the orbital parameters, such as, the inclination, period, semi-major axis, and eccentricity.

The RV method is based on the Doppler effect, i.e., the apparent change in frequency (or wavelength) of a wave in relation to an observer moving relative to the wave source. Over the past century the Doppler effect has been used in several fields of astrophysics, e.g., for measuring the kinematics of stellar clusters, young associations, and, more generally, our Galaxy, determining the orbital parameters of multiple systems, identifying radial stellar pulsations, measuring the redshift of galaxies and the Hubble's constant. In 1952, the forward-thinking Otto Struve wrote a short paper pointing out that in the early '50s the precision of Doppler measurements had become precise enough ($< 1 \text{ km s}^{-1}$) to detect massive planets ($\sim 10 M_{\text{Jup}}$), but only if they existed in short-period orbits ($P < 3 \text{ d}$). Setting aside the question of how such a planet might have formed, he realized there is no law of physics that forbids such planets from existing (Struve, 1952). His paper might have ignited a whole new area of astronomy, but in fact, it languished in obscurity. It took nearly 43 years before the first exoplanet orbiting a Sun-like star *51 Peg b* was discovered by Mayor & Queloz (1995) using the RV methods. Since then this method has played a major role in the discovery and characterization of exoplanetary systems, far beyond what Otto Struve could have imagined.

Let us now discuss the principle behind the radial velocity method. We will follow the formalism presented in Lovis & Fischer (2010) and Hatzes (2016). According to special relativity (i.e., flat space-time), if a light source emits a photon of a given wavelength λ_0 in the rest frame of the source, an observer that is moving with respect to the source with a relative radial velocity RV will detect the photon at a different wavelength λ , where the latter is given by

$$\lambda = \lambda_0 \gamma_L (1 + RV/c) = \lambda_0 \frac{1 + RV/c}{\sqrt{1 - v^2/c^2}}, \quad (2.19)$$

where RV is the radial component of v , γ_L is the relativistic Lorentz factor, and c is the speed of light in vacuum. In the context of general relativity, where the space-time curvature also gives rise to a gravitational redshift, further terms of order c^{-4} and higher should be included in Eq. (2.19). However, we are usually only interested in low-amplitude radial velocity variations, such as those

⁴See https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html.

produced by exoplanets. This simplifies the description of the Doppler effect, as this implies that the relativistic term can be dropped ($\gamma_L \approx 1$) from Eq. (2.19).

The Doppler effect allows us to measure the relative radial velocity between the light source and the observer. If we acquire measurements using ground-based facilities, we have to correct for the motion of the observer along the line of sight, such as those due to the Earth's rotation and revolution around the Sun-Earth and Moon-Earth barycenter. To account for the motion of the observer, we can transform the measured wavelength to the wavelength that a reference frame at rest would detect. To this end, following the recommendation from the International Astronomical Union (IAU), RV measurements are always referred to the International Celestial Reference System (ICRS), which is centred on the Solar System barycenter.

Neglecting the relativistic effects, we can express the radial velocity RV observed in the ICRS as

$$RV = c \frac{\lambda_B - \lambda_0}{\lambda_0}, \quad (2.20)$$

where λ_B is the wavelength that is measured in the ICRS.

As long as the orbit is not seen face-on ($i \neq 0^\circ$), a planet induces variations of the radial velocity of its host star due to the mutual gravitational pull between the two objects. We can derive an expression to describe how the RV of a star varies due to the presence of an orbiting planet studying the motion of the star around the centre of mass. As described in Sec. 2.2, the relative elliptical orbit of a planet about its star in polar coordinates is given by Eq. (2.5). We can write a similar equation for the absolute orbit of the star about the centre of mass

$$r_s(t) = \frac{a_s(1 - e^2)}{1 + e \cos \nu_s(t)}, \quad (2.21)$$

where $r_s(t)$ is the distance between the star and centre of mass, a_s is the semi-major axis of the absolute stellar orbit, e is the eccentricity⁵, and $\nu_s(t)$ is the true anomaly of the star. Describing the orbital motion of a star around the centre of mass using the RV method translates into studying the stellar motion along the line of sight of the system, to which we will refer to as z -axis.

With reference to Fig. 2.6, the z_s coordinate of a star⁶ orbiting around the centre of mass is given by

$$z_s(t) = r_s(t) \sin i \sin[\omega_s + \nu_s(t)] \quad (2.22)$$

where ω_s is the argument of pericentre of the star and i is the inclination of the orbit with respect to the plane of the sky. The radial velocity of the star $RV_s(t)$ is equal to the sum of the barycenter radial velocity v_b and the orbital velocity

⁵The stellar and planetary orbits about the centre of mass, and the relative orbit of the planet about the star share the same eccentricity.

⁶Here z_s refers to the distance between the star and the plane of the sky (see Fig. 2.6).

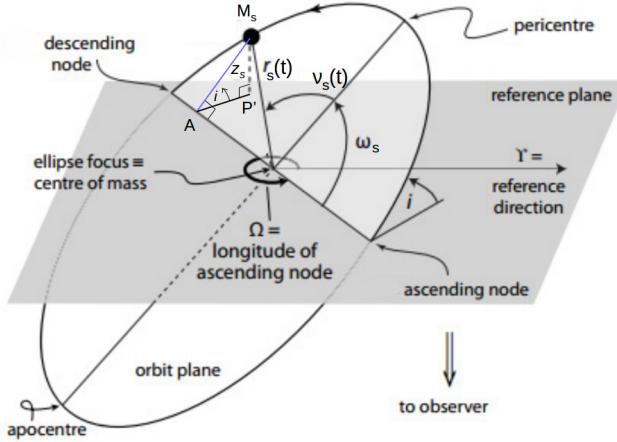


Figure 2.6: Elliptical orbit of the star about the star-planet centre of mass.

component along the line of sight $\dot{z}_s(t)$, i.e., the time derivative of Eq. (2.22):

$$RV_s(t) = v_b + \dot{z}_s(t) = v_b + \sin i \left[\sin[\omega_s + \nu_s(t)] \dot{r}_s(t) + r_s(t) \cos[\omega_s + \nu(t)] \dot{\nu}(t) \right] \quad (2.23)$$

Combining Eq. (2.23) with Eq. (2.21) and its first derivative, and using the conservation of angular momentum, we can derive an expression for the radial velocity curve of the star as a function of its true anomaly

$$RV_s(t) = v_b + K_s \left[\cos[\omega_s + \nu_s(t)] + e \cos \omega_s \right] \quad (2.24)$$

where K_s is the *radial velocity semi-amplitude* variation and is given by

$$K_s = \frac{2\pi}{P} \frac{a_s \sin i}{(1 - e^2)^{1/2}} \quad (2.25)$$

We can relate the radial velocity semi-amplitude variation of the star K_s to the planetary (M_p) and stellar (M_s) mass using Kepler's third law of planetary motion

$$P^2 = \frac{4\pi^2 a^3}{G(M_s + M_p)} \approx \frac{4\pi^2 a^3}{GM_s} \quad (2.26)$$

where the approximation holds because $M_p \ll M_s$. We note that a in Eq. (2.26) refers to the semi-major axis of the relative orbit of the planet around the star and is given by

$$a = a_p + a_s \quad (2.27)$$

where a_p and a_s are the semi-major axes of the planetary and stellar orbits about the centre of mass, respectively. From the definition of barycenter it follows

$$M_s a_s = M_p a_p \quad (2.28)$$

Combining Eqs. (2.25), (2.27), and (2.28) with Kepler's third law given in Eq. (2.26), we obtain:

$$K_s = \left(\frac{2\pi G}{P} \right)^{1/3} \frac{M_p \sin i}{M_s^{2/3}} \frac{1}{(1 - e^2)^{1/2}} \quad (2.29)$$

We can rewrite the previous equation as

$$K_s = 28.4 \frac{M_p \sin i}{M_{\text{Jup}}} \left(\frac{M_s}{M_\odot} \right)^{-2/3} \left(\frac{P}{1 \text{ yr}} \right)^{-1/3} \frac{1}{\sqrt{1 - e^2}} \text{ m s}^{-1}, \quad (2.30)$$

where K_s is given in m s^{-1} , the stellar mass M_s is in solar masses⁷ (M_\odot), the planetary mass M_p in Jupiter masses (M_{Jup}), and the orbital period P in years.

From radial velocity observations spanning the entire orbit, we can constrain P , e , ω_s , and K_s . If M_s is known through, e.g., isochrones or evolutionary tracks, we can use Eq. (2.30) to compute $M_p \sin i$, which is the so-called “minimum mass”, as $\sin i \leq 1$. We stress that the minimum planetary mass can still be considered a good estimate of M_p . In fact, as the orbital planes are randomly oriented in space, it can be proved that inclination angles close to 90° (edge-on systems) are much more frequent (see, e.g., Lovis & Fischer, 2010).

The Doppler shift of a star induced by planetary companions is small regardless of the typical planetary mass, stellar mass, and semi-major axis. We therefore need instruments with very high-precision and stability to measure the RV variation induced by a planet. For instance, assuming the orbit is seen edge-on and following Eq. (2.30), to detect a planet on a circular orbit at 0.1 au from a solar-mass star, the required RV precision is $\sim 90 \text{ m s}^{-1}$ for a Jupiter-mass planet, $\sim 5 \text{ m s}^{-1}$ for a Neptune-mass planet, $\sim 1.5 \text{ m s}^{-1}$ for a super-Earths, and $\sim 0.3 \text{ m s}^{-1}$ for an Earth-mass planet. For an hypothetical Earth-mass planet orbiting at 1 au, the required precision would be even higher, being $\sim 0.1 \text{ m s}^{-1}$. The second obstacle in detecting low-mass planets through the RV method is the Doppler signal induced by stellar activity, which can drown out the planetary signal. Please see Sec. 2.4 for a thorough description of stellar activity and how it impacts our capability to detect planets with the RV method. In the following subsections I will present the tools that are commonly used to extract the RV measurements, analyse the time series, and identify potential signals arising from stellar activity.

2.3.1 Calculating the Doppler Shift: the Cross-correlation Function

The instantaneous Doppler shift could in principle be computed from the several lines that are present in a stellar spectrum. However, single spectral lines can

⁷The mass of the Sun is $\sim 1.989 \times 10^{30} \text{ kg}$.

be noisy, if the spectrum has a low signal-to-noise (S/N) ratio, and/or highly distorted by stellar activity, granulation, and convection. Therefore, it is preferred to use a tool which averages out the Doppler information carried out from the many spectral lines produced in the stellar photosphere. To this end, the common approach is to compute the so-called cross-correlation function (CCF).

Let us consider a star and its spectrum $s(x)$, which is generally retrieved as a function of the pixel location x . The CCF between $s(x)$ and a template spectrum $t(x)$ is defined as:

$$\text{CCF}(\Delta x) = s(x) \otimes t(x) = \int_{-\infty}^{+\infty} s(x)t(x + \Delta x)dx, \quad (2.31)$$

where Δx is the so-called “lag” of the CCF. Eq. (2.31) gives the theoretical definition of the CCF. However, since we have to deal with discretely sampled spectra, we have to compute the CCF numerically by expressing the integral through a discrete sum, that is:

$$\text{CCF}(\Delta x) = \sum_{x=1}^N s(x)t(x + \Delta x) \quad (2.32)$$

Both Eq. (2.31) and Eq. (2.32) describe an operation in which the template spectrum is shifted over the real spectrum, while the CCF reaches its maximum for a Δx that makes the template spectrum overlap the real spectrum. Unfortunately, the Doppler formula in Eq. (2.20) is non-linear which means at different wavelengths the Doppler shift in pixels will be different. This can be remedied by re-binning the linear wavelength scale onto a logarithmic one transforming the Doppler formula to

$$\Delta \ln \lambda = \ln \lambda - \ln \lambda_0 = \ln \left(1 + \frac{RV}{c} \right) \quad (2.33)$$

The lag of the CCF is then

$$\Delta x = \Delta \ln \lambda = \ln \left(1 + \frac{RV}{c} \right) \quad (2.34)$$

which is a constant for a given radial velocity RV . To conclude, we cross-correlate the observed spectrum with a template, search for the lag Δx at which the CCF reaches its maximum, and use Eq. (2.34) to determine the corresponding radial velocity RV .

We now have a powerful tool to extract RV measurements, but how do we choose the template $t(x)$? To compute absolute RVs, the adopted template should be the spectrum of our star as if it were in a reference frame at rest. The key to having a CCF that produces a good velocity measure is to have a template that matches as closely as possible our stellar spectra and that has a high signal-to-noise ratio. In the searching for exoplanets one is primarily interested in changes in the star’s velocity, or relative RVs. As such, we could

use one observation of the target star as a template and cross-correlate the other spectra of the star with this one. This would provide an excellent match between the template and the star spectra. However, the drawback of using an actual observation is that the template would contain noise, as it comes from real observations, introducing noise into the CCF. To partially mitigate this effect, one could co-add several observations of the target star to produce one master, high S/N ratio template. Alternatively, one could use a synthetic spectrum.

A very common approach is to use a digital stellar mask that is noise-free. The mask is a binary function that contains “holes” at the rest wavelength of stellar lines, i.e., a binary transmission function with 1’s at the position of stellar lines and 0’s elsewhere. The mask should be created *ad hoc*, based on the spectral type of the target star. However, differences between close spectral types are not significant from the point of view of signal extraction and experience has shown that a single mask for each main type (G, K, M) is usually sufficient. The pipeline of the High Accuracy Radial velocity Planet Searcher (*HARPS*; Mayor et al., 2003) spectrograph uses G2, K5, M2 numerical masks to compute the CCF.

The presence of photospheric inhomogeneities, such as spots and plages, combined with stellar rotation can affect the shape of spectral lines, mimicking a wavelength shift of their centroids. The amplitude of the apparent Doppler shift can be of the same order of magnitude of the Doppler reflex motion induced by an orbiting planet, leading to false positive detections (see, e.g., Queloz et al., 2001). The CCF is a powerful tool to disentangle *bona fide* planetary signals from stellar activity. While planets induce a rigid shift of the CCF, any change of its shape and symmetry is linked to stellar activity. We could probe the effect of stellar activity on our data by quantifying the CCF asymmetry. By identifying and properly removing the contribution of stellar activity from the RV signal, we can, e.g, assess whether a planet detected in transit photometry is also present in RV time series, providing a spectroscopic confirmation of transiting planets (see, e.g., Simola et al., 2019, and references therein).

2.3.2 Analysis of RV Time Series: the Periodogram

In order to detect periodic signals, which may hint at the presence of planetary companions, we could start by plotting the RV time-series as in Fig. 2.7. Sometimes we can find a periodic signal in the data simply by looking at our time series. This only works if the amplitude of our signal is large and the sampling is good. However, if the amplitude of the signal is comparable to the measurement error, the signal is not so easy to detect by eye. For this reason, we usually prefer to visualize the periodogram of our time-series. A periodogram is essentially a power spectrum of the time-series (see, e.g., Fig. 2.8).

The most used tools are largely based on the Discrete Fourier Transform (DFT). For a time series of measurements $X(t_j)$ where t_j is the epoch of the

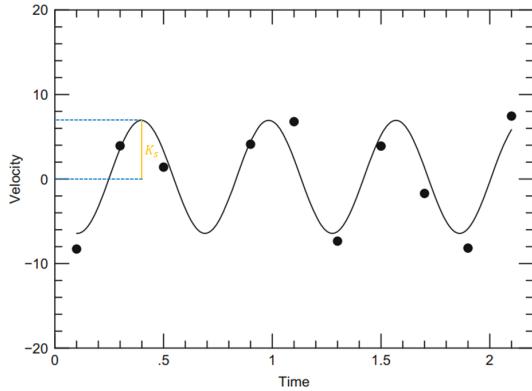


Figure 2.7: Simulated RV time series and best-fitting sinusoidal fit. The RV semi-amplitude variation K_s is marked with a yellow vertical line (Hatzes, 2016).

j -th measurement, the DFT is defined as

$$\text{DFT}_X(\omega) = \sum_{j=1}^N X(t_j) e^{-i\omega t_j}, \quad (2.35)$$

where $e^{-i\omega t}$ is the complex trigonometric function $\cos(\omega t) + i \sin(\omega t)$, $\omega = \frac{2\pi}{P}$ the pulsation (also known as angular frequency⁸), and N the number of data points. The power is defined by

$$\mathcal{P}_X(\omega) = \frac{1}{N} |\text{DFT}_X(\omega)|^2 \quad (2.36)$$

and it is often called the classic periodogram. To visualize our data we can then plot the power spectrum $\mathcal{P}(\omega)$ or, alternatively, the amplitude spectrum $A(\omega)$, where $A = \sqrt{\mathcal{P}_X}$ as a function of the frequency $f = 1/P = \omega/2\pi$. The utility of the DFT is that, if a periodic signal is in our data, it appears (nearly) as a δ -function in Fourier space at the frequency and with the amplitude of our sine wave.

Ideally, one would like to have data that are taken in equally spaced time intervals, but this is rarely possible with astronomical observations. A different mathematical tool that accounts for unevenly-spaced measurements is the so-called Lomb–Scargle (LS) periodogram (Lomb, 1976, Scargle, 1982). This tool computes the power spectrum in a more sophisticated way (see, e.g., Hatzes, 2016) compared to Eq. (2.36). An improvement of the LS periodogram is the Generalized Lomb–Scargle (GLS, Zechmeister & Kürster, 2009) periodogram. Unlike the LS, the GLS periodogram is able to deal with eccentric orbits, and

⁸Frequency is often measured as angular frequency, which is related to the period by $\omega = 2\pi/P$. When I refer to a frequency it is merely the inverse of the period.

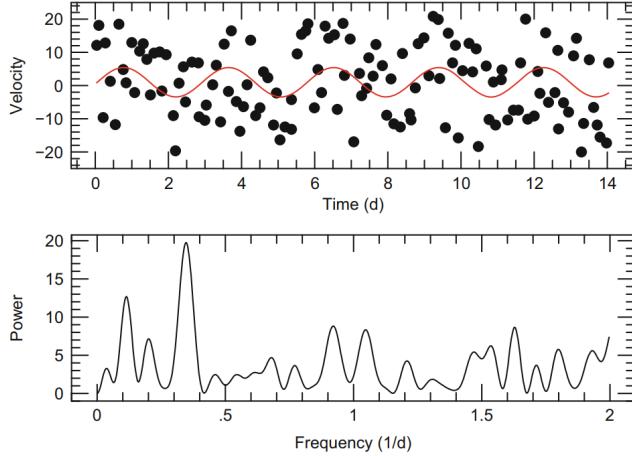


Figure 2.8: *Top:* Time series (dots) of a sampled sine function (curve) with a period of 2.58 days whose amplitude is the same level as the noise. The sine variations are barely visible in the time series. *Bottom:* The DFT power spectrum of the above time series. A dominant peak at the frequency of $1/2.58 \text{ days}^{-1}$ is clearly seen (Hatzes, 2016).

accounts for the measurement errors, which enter the computation of the power spectrum in the form of weights.

When interested in finding a periodic signal we usually look for the highest peak in the periodogram of a given time-series. It is important to check the statistic significance of the peak. To this end, we can define the so-called false-alarm probability (FAP), i.e., the probability that noise could produce a peak with a power higher than the one we found in your data, over a wide frequency range. Following Scargle (1982) the FAP is defined as

$$\text{FAP} = 1 - (1 - e^{-z})^N, \quad (2.37)$$

where z is the power of the peak in the LS periodogram and N is the number of independent frequencies.

It is also possible to define the FAP at a given frequency. Let us suppose we know that there must be a periodic signal at a given frequency in our data and we want to determine the probability that noise can produce a peak higher than what is observed exactly at that frequency in our power spectrum. In this case $N = 1$ as there is only one independent frequency. The above expression reduces to

$$\text{FAP} = e^{-z} \quad (2.38)$$

This case is of particular interest for a transiting planet, where we know the period of the orbiting body – it has been measured via transit photometry! – and we are interested only in the significance of the power at that orbital frequency in the periodogram of our RV measurements.

2.4 Stellar Activity

Stellar activity is one of the main limitations for precisely measuring the K_s semi-amplitude induced by an orbiting planet. It can strongly affect our capability to detect new planets. In this context, stellar activity is a sort of “umbrella term”, which includes all the Doppler signals induced by the star, and that are detectable in our data.

Following Lovis & Fischer (2010) and Dumusque (2016), we can classify the different sources of stellar noise according to their average timescales. In a typical solar-like star we may have:

- Stellar oscillations on timescales of a few minutes.
- Stellar granulation and supergranulation on timescales of a few minutes, up to 48 hours.
- Short-term stellar activity on the timescale of the stellar rotation (mainly due to spots and plages).
- Long-term stellar activity on timescales of several years (mainly related to stellar magnetic cycles).

2.4.1 Stellar Oscillations

Stars having an outer convective envelope can stochastically generate pressure waves (also called p-mode oscillations) at their surface through turbulent convection. P-waves induce expansions and contractions of the external envelope over timescales of a few minutes. These stellar oscillations are seen in the RV signal of the star as superposition of a large number of modes, which may cause RV variations typically between 10 and 400 cm/s, depending on the stellar type and evolutionary stage.

According to theoretical predictions, the oscillation frequencies increase with the square root of the mean stellar density, while the induced RV amplitudes scale with the luminosity-to-mass ratio (Christensen-Dalsgaard, 2004). Low-mass, non-evolved stars have lower $\frac{L}{M}$ and are therefore easier targets for RV planet detection as the noise due to p-modes is lower. The most precise spectrographs today, such as ESPRESSO (Pepe et al., 2014) and EXPRESS (Fischer et al., 2016), can reach sub- $m\text{s}^{-1}$ RV precision and can therefore resolve stellar p-mode oscillations in solar-like stars.

2.4.2 Stellar Granulation and Supergranulation

Unlike stellar oscillations, granulation is the result of large-scale convective motions, which occur in the outer layers of stars having a convective envelope. The granulation pattern is made of a large number of cells showing bright up-flows and darker down-flows, which are due to hot matter coming from deeper layers and cooled down matter at the surface, respectively. On the Sun, convective motions of the single cells have typical velocities of 1-2 kms^{-1} in the vertical

direction. However, the large number of granules on the visible stellar surface efficiently averages out the velocity field. As a consequence, there is just a remaining noise affecting RV measurements of the order of the m s^{-1} level.

We can also differentiate between granulation and supergranulation. We specifically refer to granulation when timescales are of the order of a few tens of minutes. In case of the Sun, this is the typical timescale over which a given granule evolves significantly. On longer timescales, from a few hours to about 48 hours, other similar phenomena called meso- and supergranulation occur. Their origin is not very clear, although they are suspected to be due to larger convective structures in the stellar photosphere.

2.4.3 Short-term Stellar Activity

Short-term stellar activity is induced by stellar rotation in the presence of active regions and surface magnetic inhomogeneities, such as dark spots and bright plages and faculae.

Starspots

Starspots are regions on the stellar surface that appear as temporary spots that are darker and cooler than the surrounding photosphere. In case of the Sun, spots are ~ 700 K cooler than the Sun's effective temperature⁹ ($T_{\text{eff}} = 5777$ K; Meunier et al., 2010). They are common in F, G, K, and M-type stars, i.e., stars with an outer convective zone and an inner radiative region.

Spots are magnetically active regions: for instance, the magnetic field intensity of the solar quiet photosphere is about 1 Gauss, whereas the typical field strength in sunspots is a few thousands Gauss. The strong magnetic field inhibits convection, preventing hot plasma from deeper stellar layers to emerge at the photosphere. Spots appear usually in pairs of opposite magnetic polarity. The typical spot structure consists of a darker inner region called *umbra* surrounded by a filamentous rim called *penumbra*. Their number, size, and lifetime are strongly linked to the strength of the magnetic flux that emerges from the star photosphere, and vary according to the magnetic cycle of the star. As a reference, the Sun has a magnetic cycle of about 11 years.

Plages and Faculae

Plages and Faculae, similarly to spots, are also linked to the strength of the magnetic flux in active regions of the stellar surface. However, unlike spots, plages and faculae appear as brighter regions on the surface of the star (Chatzistergos et al., 2022). While plages appear in the chromosphere of the star, faculae are typical phenomena of the photosphere and are often detected in the vicinity of spots.

⁹The effective temperature of a star, usually denoted as T_{eff} , is the temperature of a blackbody that would emit the same amount of integrated radiation as the star.

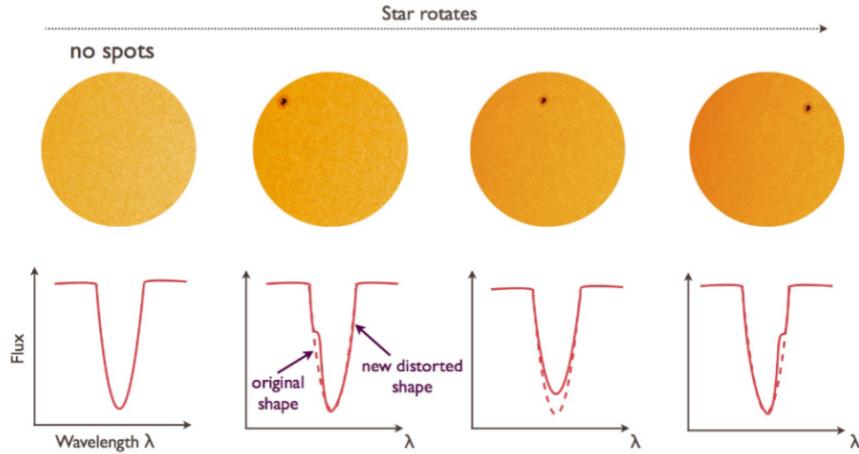


Figure 2.9: Diagram illustrating how flux blocked by starspots on the rotating stellar disk induces asymmetries in the spectral lines, leading to variations in the RV measurement (from Haywood, 2015).

2.4.4 Effects of Short-term Stellar Activity on radial velocities

Let us now discuss the effect of spots, plages, and faculae on radial velocity measurements. Due to stellar rotation, these short-term stellar activity phenomena induce RV variations according to two different physical processes.

1. As the star rotates, half side of the stellar disk is blue-shifted, while the other one is red-shifted. When a spot crosses the visible disk, we observe a flux deficit, which moves from the blue-shifted to the red-shifted part of the stellar disk as the star rotates (Fig. 2.9) (depending on the inclination of the spin axis). This changes the shape of spectral lines and introduces anomalous RV variations on timescales comparable to the stellar rotation period. In the case of plages and faculae, instead, we observe a flux increase. This flux excess is, however, smaller compared to the deficit caused by spots, so the plages- and faculae-related effect on the RV time series is overall weaker.
2. The magnetic field strength is also related to stellar rotation. As mentioned in Sec. (2.4.3), the strong magnetic field inside active regions inhibits convection. This means that we no longer observe the hot plasma coming from the interior toward the surface, hence its corresponding blue-shift effect is suppressed. The net balance is that active regions appear red-shifted, and this introduces RV variations on timescale of the rotation period when active regions are visible.

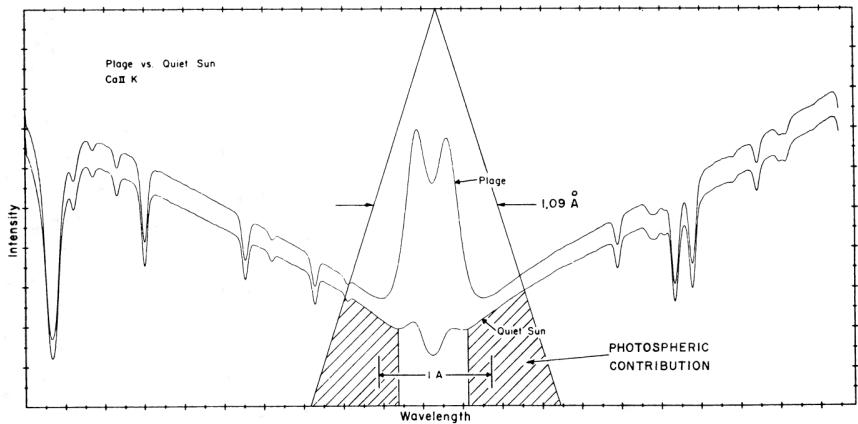


Figure 2.10: Spectrum of the quiet Sun (the lowest, indicated by the arrow saying "Quite Sun", as taken from White & Livingston, 1981) together with the spectrum of a more active star affected by plages. The triangular bandpass of the HKP-2 Spectrometer is superimposed to mark the photospheric contribution to the S-index from the quiet Sun. The chromospheric emission due to the plages falls within the HKP-2 bandpass (from Hartmann et al., 1984).

2.4.5 Long-term Stellar Activity

Long-term stellar activity is induced by stellar magnetic cycles, which affect RV measurements on timescales of several years. During the phases of high stellar activity, the stronger magnetic field reduces convection. As a consequence, the star appears redder and this adds a positive RV contribution to the RV measurement. A positive correlation between RVs and the level of stellar activity is thus expected and different indicators are indeed used by astronomers to trace stellar activity.

- *S-index.* Magnetic heating in chromospheric active regions of F, G, K, and M-type stars leads to emission in the cores of the CaII H & K absorption lines. Named by Joseph von Fraunhofer, the H and K lines are prominent absorption lines at about 3934 Å and 3968 Å due to singly ionized calcium (CaII). The S-index is defined as the flux ratio of two bandpasses centred around the cores of the CaII H & K lines and two continuum regions on either side of the lines (Fig. 2.10). In detail, the S-index definition is linked to the HKP-2 spectrometer, which was designed to measure CaII H & K line fluxes (Vaughan et al., 1978). It was put in operation in 1977 with the aim of detecting four channels of spectral data: two channels are the 20-Å-wide continuum bands (called "R" and "V"), and the other two are triangular bandpasses with a 1.09-Å FWHM centred around the cores of

the Ca II H & K lines (called "H" and "K"). The S-index is given by

$$S = \alpha \frac{H + K}{R + V}, \quad (2.39)$$

where H , K , R , and V are the fluxes in the corresponding bandpasses, and the multiplicative constant is $\alpha = 2.4$, as determined from observations made by Wilson (1978). For further details, see, e.g., Duncan et al. (1991), Wright et al. (2004). The chromospheric activity S-index is sensitive to the intensity of the emission component in the cores of the Ca II H & K lines (Wilson, 1978), which in turn depends on the magnetic activity of the star. The S-index shows variations on timescales of the stellar rotation period due to active regions moving in and out of sight as the star rotates around its axis.

- R'_{HK} . It is the ratio between the chromospheric component of the cores of the Ca II H & K emission lines and the total bolometric emission of the star. R'_{HK} is computed from the S -index using the relations in Middelkoop (1982). The R'_{HK} index contains only the chromospheric contribution, while the S -index contains both the photospheric and chromospheric contributions.
- *X-Ray luminosity*: The luminosity of the star in the X-Ray spectral domain (from 10 pm to 10 nm, see, e.g., Robrade et al., 2012).
- *H_α line*. The H_α line is the first spectral line of the Balmer series, which originates when the electron in an hydrogen atom transits from its third to its second lowest energy level. The line falls at a wavelength of $\lambda_\alpha = 6562.81\text{\AA}$. The H_α -index is defined as $H_\alpha = 2F_H/(F_{c1} + F_{c2})$ (e.g., Sissa et al., 2016), where F_H is the instrumental flux measured at the H_α line location, while F_{c1} and F_{c2} are the fluxes measured in two continuum intervals symmetrically located with respect to the H_α . This indicator is sensitive to stellar activity (e.g., Robinson et al., 1990).
- *Na I D lines*: Absorption spectral lines of the atomic sodium. Historically called D after Joseph von Fraunhofer's study, this spectral feature actually consists of two closely spaced lines, D1 and D2, which form the sodium doublet. Their corresponding wavelengths are $\lambda_{D1} = 5895.92\text{\AA}$ and $\lambda_{D2} = 5889.95\text{\AA}$. The presence of these lines has also been proven to be an useful index for long-period activity cycle (Robertson et al., 2015).
- *CCF profile diagnostics*. As mentioned in Sec. 2.3.1, changes in the shape and profile of the CCF gives important hints as to stellar activity on the average line profile. After properly fitting the CCF, some of the fit-extracted parameters can be used to trace the effects of stellar activity. For more details, see Sec. 3.1.
- *crx: chromatic RV index*. For a given epoch spectrum, we can extract $RV(\lambda)$ measurements from different chunks of the spectrum (i.e., from

spectral ranges centred around different wavelengths λ). We can then use a linear correlation in the form $RV\lambda = \alpha + \beta\lambda$, where the slope β defines the *crx* chromatic RV index (Zechmeister et al., 2018). While a *bona fide* planet produces an RV signal that does not depend on wavelength, spot-induced variability is expected to show “a chromatic effect”, owing to the changing contrast between spots and quiet photosphere with wavelength.

- *dLW*: *Differential line width*. This indicator evaluates the width changes of the spectral absorption lines at different wavelengths (for further details, see Zechmeister et al., 2018).

Chapter 3

Methods

3.1 Cross-correlation Function Analysis

In Sec. 2.3.1 we introduced the cross-correlation function, whose shape and asymmetry correlate with stellar activity as they trace the profile variation of spectral lines. We also explained (Sec. 2.4) that stellar activity is indeed the main limitation to detecting low-mass exoplanets. Therefore, it is important to develop a technique to fully characterise the CCF shape along with its temporal variations, so to remove the activity contribution from our data. We are interested in estimating three CCF parameters: the barycenter (linked to the stellar radial velocity), the width and the asymmetry (which are stellar activity tracers, see e.g. Dumusque, 2016). One very common technique for analysing the CCF consists in fitting a Gaussian function onto the CCF (see Sec. 3.1.1). Other techniques aiming at evaluating the line profile variations have been developed (Sec. 3.1.3) and, in particular, a new method based on a Skew Normal fit has been recently proposed (Simola et al., 2019, Sec. 3.1.2).

3.1.1 Normal Fit

Computing a Normal fit onto the CCF is the standard technique for CCF analysis. The mean of the fitted Gaussian provides a measure of the CCF barycenter, while the FWHM quantifies the CCF width. Instead, the CCF asymmetry cannot be directly obtained from the Normal fit since a Gaussian function is symmetric by definition. One way to estimate the CCF asymmetry is through the bisector velocity span (BIS SPAN, Queloz et al., 2001). In detail, for each CCF flux level, the mid point between the red and blue wing of the CCF is computed. Then the CCF is divided in two sections: the top section is located between 60% and 90% of the CCF full depth, while the bottom section is located between 10% and 40% of the CCF full depth. The BIS-SPAN is finally computed by the difference between the average mid-point of the top section (top velocity \bar{V}_t) and the average mid-point of the bottom section (bottom velocity \bar{V}_b); see Fig. 3.1 for a schematic representation.

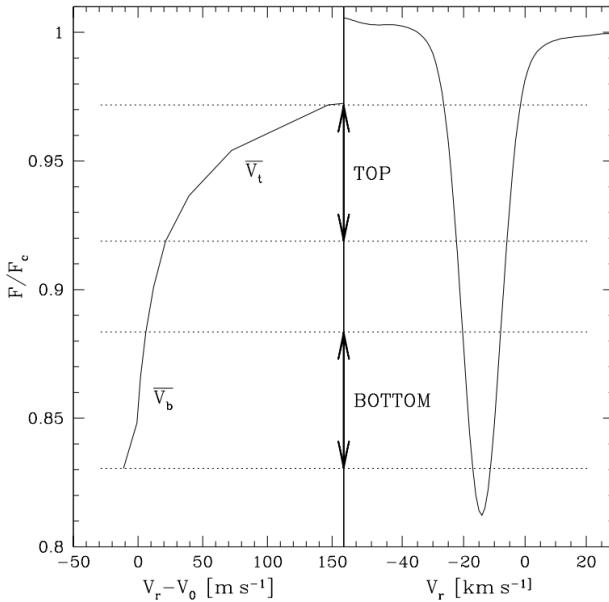


Figure 3.1: Computation of the BIS-SPAN $\equiv \bar{V}_t - \bar{V}_b$. *Right panel:* the CCF of a star. *Left panel:* the CCF bisector, where the velocities computed from the top (\bar{V}_t) and bottom (\bar{V}_b) section of the CCF are shown; V_0 is an arbitrary offset. The change in the bisector slope emphasizes the CCF asymmetry (Queloz et al., 2001).

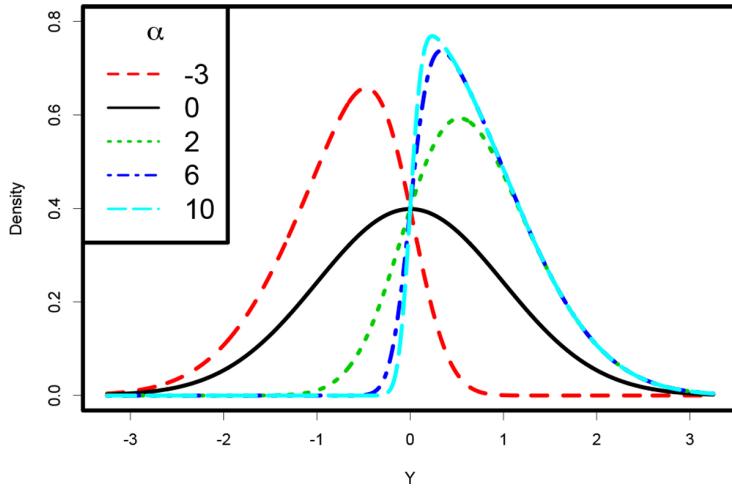


Figure 3.2: Representation of a Skew Normal distribution with location parameter $\xi = 0$, scale parameter $\omega = 1$ and different values of the Skewness parameter α indicated by different colours (Simola et al., 2019).

3.1.2 Skew Normal Fit

Almost all the methods that are based on a Normal fit of the CCF require a separate approach to quantify the CCF asymmetry, as described in Sec. 3.1.1. For this reason, Simola et al. (2019) suggest fitting the CCF with a Skew-Normal (SN) function. The free parameters of a SN function are a location and a scale parameter (i.e. the counterpart of the mean and the standard deviation of a Normal function). The SN function has a further free parameter, the skewness, which makes the distribution asymmetric (Azzalini, 1985). This means that by applying a skew-normal fit to the CCF, we can extract the RV, FWHM and asymmetry of the CCF at the same time.

Going into the details, let us consider a random variable $Y \in \mathbb{R}$ which follows a SN distribution. We introduce a location parameter $\xi \in \mathbb{R}$, a scale parameter $\omega \in \mathbb{R}^+$, and a skewness parameter $\alpha \in \mathbb{R}$, which are called direct parameters (DPs). Through this parametrization, the density function of the Y variable at some value $y \in Y$ can be written as:

$$SN(y; \xi, \omega, \alpha) = \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left(\frac{\alpha(y - \xi)}{\omega}\right) \quad (3.1)$$

where ϕ and Φ are the density function and distribution function of a standard Gaussian distribution, respectively. Note that the Normal distribution is the special case of the SN distribution with $\alpha = 0$, which means absence of asymmetry. Some SN functions with different α values are shown in Fig. 3.2 as examples.

For reasons related to the interpretation of the DPs, it is convenient to introduce another set of parameters describing the SN functions (the so called

centred parameters: CPs), which are (μ, σ, γ) . DPs may be converted into CPs via the following relations (Arellano-Valle & Azzalini, 2008):

$$\mu = \xi + \omega\beta, \quad \sigma^2 = \omega^2(1 - \beta^2), \quad \gamma = \frac{1}{2}(4 - \pi)\beta^3(1 - \beta^2)^{-3/2} \quad (3.2)$$

where $\beta = \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{\sqrt{1+\alpha^2}} \right)$. The new parameters μ and σ are the actual mean and variance of the SN distribution, while γ quantify the skewness.

Now, how shall we use these parameters to characterise the CCF after the SN fit? We will take γ as an estimate of the CCF asymmetry, also noting that standard errors associated to γ measurements are on average smaller by 15% than the standard errors associated to the BIS SPAN (see Fig. 3.3, left panel). Regarding the CCF width – in analogy to the relation holding for the FWHM of a Gaussian – Simola et al. (2019) suggest using the $\text{FWHM}_{\text{SN}} \equiv 2\sqrt{2\ln 2}\sigma$, where σ is the standard deviation of the SN function. Simola et al. (2019) further prove that both γ and FWHM_{SN} are more sensitive to stellar activity if compared to their counterparts obtained within the Normal fitting framework.

Finally, we want to compute the radial velocity from the barycenter of the CCF; to this end we could use either the mean or the median of the SN function. Simola et al. (2019) show that the median is a better choice as its associated uncertainty is on average smaller than the one associated to the SN mean. Moreover, they show that the SN-related RV inferred from the median has a lower uncertainty if compared with the RV extracted from the Normal fit, hence confirming the better performance of SN functions in fitting the CCFs (see Fig. 3.3, right panel).

3.1.3 Other techniques for analysing the CCF

Although less common, other techniques have also been developed to quantify the CCF asymmetry. Boisse et al. (2011) propose to retrieve two different RV measurements by fitting Normal functions onto the upper and lower part of the CCF separately. After that, they compute the V_{span} as the difference between the two velocities, giving a hint on the asymmetry.

A different way of quantifying the CCF asymmetry, firstly proposed by Nardetto et al. (2006), uses a bi-Gaussian, which is a function obtained by merging two half Gaussians having different standard deviations. By fitting both a Gaussian and a bi-Gaussian onto the CCF, two different RV measurements are extracted and the difference between the two represents a proxy for the CCF asymmetry.

Finally, Figueira et al. (2013) evaluate the RV information content that can be extracted from spectral lines following Bouchy et al. (2001) and then they compute the V_{asy} indicator as the RV difference between the average information content of the blue and red wings of spectral lines.

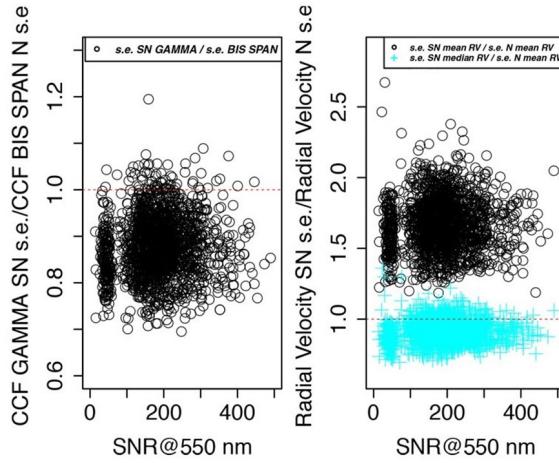


Figure 3.3: Relative precision of the SN-based γ and RV parameters with respect to their N-based counterparts, as quantified by Simola et al. (2019) by bootstrapping ~ 2000 CCFs of the following stars: HD 215152, HD 192310, and CoRoT-7. *Left panel:* ratio between the standard errors associated to γ and the standard errors associated to BIS SPAN. Standard errors of the SN-based asymmetry parameter, γ , are on average smaller by 15%. *Right panel:* standard errors of the SN-based mean (black circle) and median (cyan pluses) RV normalised with respect to the standard errors of the RV as extracted from the N-fit. The standard errors of the SN mean (resp. SN median) RV are on average greater by 60% (resp. lower by 10%) than the standard errors associated to the N-fit RV.

3.2 RV detrending: the breakpoint method

After obtaining fundamental information from the CCF as explained in Sec. 3.1, it is important to properly detrend the RV data. To this end we will use the breakpoint (bp) method, here it is presented following Simola et al. (2022) and Bai & Perron (2003).

As we already mentioned, stellar activity represents the biggest obstacle in planet detection when using the RV method. A very common strategy for removing the stellar activity is setting up some kind of detrending along the entire time series, which we can refer to as overall correction (oc) approach. However, active regions on the photosphere constantly evolve over time, as can be seen by the temporal change of the activity indicators obtained from the CCF (e.g. its asymmetry). In other words, stellar activity is not stationary, but rather *piecewise stationary*, which means that it does not change significantly within certain properly selected temporal intervals. As a consequence, the correlations between the RV observations and the activity indicators are expected to change over time. If we are able to find the temporal locations where the correlation changes are statistically significant, we could split the time series in different piecewise stationary segments and detrend the RV time series on a chunk-wise base, rather than performing a global detrending along the whole time series. This is the general idea behind the change point detection (CPD) methods, among which the bp method represents a possible implementation.

Let us go into the details. CPD methods can be classified as either online or offline (van den Burg & Williams, 2022). Online methods are used when wanting to know the change points within the data in real time. Offline methods, on the other hand, operate “*a posteriori*”: they estimate the change point locations when the collecting data process is over. In addition to this classification, if only a single-parameter change is monitored, we talk about univariate CPD methods, otherwise we deal with multivariate methods.

Let us now consider a univariate time series $y = y_{1,\dots,T} = \{y_i\}_{i=1}^T$, where T is the number of data points. Let us also assume that the time series is piecewise stationary, so that there are D generally unknown locations where the activity-RV correlation changes significantly. The goal of the CPD methods is to compute the set of indices l corresponding to the D locations, $l = l_1, \dots, l_D$ (by convention $l_0 \equiv 1$ and $l_{D+1} \equiv T$). This means finding the best segmentation of the time series and these indices are used to delimit the different regions in which y is stationary.

To retrieve the best possible segmentation, CPD methods rely on the maximization of the penalized log-likelihood function:

$$\mathcal{L}_P(D; y) = \mathcal{L}(D; y) - \lambda P(T) = \sum_{k=1}^{D+1} \log L(y_{[l_{k-1}:l_k-1]}) - \lambda P(T), \quad (3.3)$$

where k is an index used to identify the $D+1$ segments of the time series bounded by l_0, \dots, l_{D+1} , $\log L(y_{[l_{k-1}:l_k-1]})$ is the log-likelihood function evaluated at the k -th segment, λ is a positive constant, and $P(T)$ is a penalty function that

is added if the total number of change points is not assigned a priori (Bai & Perron, 2003).

Let us now illustrate in more detail the break-point method. With specific reference to the SN-fit onto the CCF (Sec. 3.1.2), Simola et al. (2022) choose to model the stellar activity via:

$$RV_{\text{activity}} = \beta_0 + \beta_1 A + \beta_2 \gamma + \beta_3 \text{FWHM}_{\text{SN}} + \epsilon, \quad (3.4)$$

where the covariates are A (i.e. the contrast parameter of the CCF), γ (i.e. the CCF asymmetry parameter), and FWHM_{SN} (i.e. the CCF width parameter). Then, β_i represents the model coefficients to be optimised, while ϵ is the random error having a multivariate normal distribution, whose vector of means is 0 and its variance-covariance matrix is equal to $\sigma^2 I$ (being I the identity matrix). We can also arrange the covariates in a matrix X , the parameters in a vector β and write our time series as: $y = X\beta + \epsilon$. If we assume piecewise stationarity for y , we can divide it in $D+1$ segments over which the coefficients are constant, thus there is a set of model coefficients per segment, that is β_k with $k = 1, \dots, D+1$. We can then write our vector of observations as:

$$y_{[l_{k-1}:l_k-1]} = X_{[l_{k-1}:l_k-1]} \beta_k + \epsilon_k \quad (3.5)$$

Following Bai & Perron (2003), the break-point indices are estimated by minimising the residual sum of squares (RSS) of Eq. (3.5), which is equivalent to maximizing the log-likelihood in Eq. (3.3). This computation is done through the dynamic programming algorithm (Fisher, 1958, Bellman & Roth, 1969, Guthery, 1974) following the implementation outlined in Zeileis et al. (2003).

3.3 Markov chain Monte Carlo

In order to fully characterise an exoplanetary system, in this work both the light curves (LCs) from the transit method (Sec. 2.2) and the RV time series acquired via the radial velocity method (Sec. 2.3) are used. The analysis of these data is achieved using a Markov chain Monte Carlo (MCMC) algorithm developed by Bonfanti & Gillon (2020). Other commonly employed codes are *EXOFAST* (Eastman et al., 2013) and its new version *EXOFASTv2* (Eastman et al., 2019), *TLCM* (Smith et al., 2017), *PlanetPack3* (Baluev, 2018), *Juliet* (Espinoza et al., 2019), *Pyaneti* (Barragán et al., 2019), *Exoplanet*¹, *Allesfitter* (Günther & Daylan, 2021). In this section I will give an overview on MCMC algorithms and then I will focus on the specific working scheme of the here employed code.

A MCMC process is a tool that is used to sample the probability distribution functions of parameters of interest assuming a given model. This is achieved by generating a sequence (*chain*) of data points (called *states*), starting from an initial state that is then perturbed. There are different implementations of a MCMC scheme such as the Metropolis-Hastings algorithm (Hastings, 1970),

¹<https://github.com/exoplanet-dev/exoplanet>

the Gibbs sampler (Casella & George, 1992), the Differential Evolution (Ter Braak, 2006), the Affine Invariant (Foreman-Mackey et al., 2013) or the Hamiltonian (Betancourt, 2017) approach. These types of algorithms generate different chains with a certain number of steps given by the user. One important thing to keep in mind is that the first few generated states of a chain probably do not follow the limiting distribution just yet, therefore it is a good practice to discard them: this process is called *burn-in*. The burn-in length (that is the number of initial steps to be discarded) depends on the initial state and the speed of convergence to the limiting distribution.

3.3.1 The MCMCI code

The code proposed by Bonfanti & Gillon (2020) is called MCMCI and allows to combine a MCMC approach for the LC and RV analysis of exoplanets with the isochrone placement algorithm (Bonfanti et al., 2015, 2016) for the stellar host characterisation. In fact, in order to obtain an estimate of planetary parameters, we also need to know some stellar parameters, namely the radius and mass. The great advantage of the MCMCI approach is that, instead of inferring the stellar parameters separately from evolutionary models, this code allows one to analyse photometric and/or radial velocity time series with a proper interpolation within stellar evolutionary isochrones directly. In this work, however, given the large number of LCs to be analysed, isochrone interpolation was not used to save computational time. Therefore, the following discussion will be focused on the MCMC aspects of the MCMCI code.

The MCMCI code is implemented in Fortran and uses the Metropolis-Hastings algorithm, possibly combined with the Gibbs sampler. The user shall give in input ASCII files containing the LCs and/or RV time series, besides an input text file that contains all the necessary information to set up the run (see Sec. 3.3.4 for more details).

The MCMCI code implements the photometric model by Mandel & Agol (2002) to reproduce the eclipse, as well as a classical Keplerian model for analysing the RV signal, and the Rossiter-McLaughlin effect model by Giménez (2006) if RVs are obtained during the transit. As default time standard, the code uses the Barycentric Dynamic Time (TDB), which avoids the drift due to leap seconds and it refers to the barycentre of the Solar System, thus it corrects for relativistic effects due to the gravitational potential of Earth. A built-in function which converts Coordinated Universal Time (UTC) to TDB is also available, if needed. For deepening the topic about time standards, the interested reader may refer to Eastman et al. (2010).

Using polynomials, the code is able to detrend the LCs against several effects, such as the inhomogeneous intra-pixel sensitivity of the detector, the ramp effect (i.e. an increasing gain of individual pixels over time, depending on the illumination history of the pixels), and signal variations due to rotating spots, the sky background, the airmass, the centroid location on the CCD, and the width of the point spread function during observation, once the appropriate detrending vectors are given in input. The eclipse model is then multiplied

by the polynomial trend model with the goal of reproducing all the systematic effects that can arise from astrophysical or instrumental effects. Therefore, as a preliminary step, one needs to search for the best detrending baseline. This is achieved by varying the polynomial orders of the different terms and picking up the set of orders with the smallest Bayesian information criterion (BIC, Schwarz, 1978, see below for further details).

Similarly, it is also possible to build a detrending baseline for the RV time series to model any RV trend against time, CCF width, CCF contrast, CCF skewness, and a stellar activity indicator (e.g. $\log R'_{\text{HK}}$).

There are numerous parameters that can be sampled at each step, the so called *jump parameters* or *step parameters*. Here is a list of the jump parameters. The stellar jump parameters are:

- Stellar mass M_s .
- Stellar radius R_s .
- Stellar effective temperature T_{eff} .
- Stellar metallicity [Fe/H].

The planetary jump parameters (if more than one planet has to be characterised, all the relevant parameters have to be listed in the input file) are:

- Transit depth dF .
- Occultation depth dF_{occ} .
- Impact parameter in the case of circular orbit $b' = \frac{a \cos i_p}{R_s}$.
- Eclipse duration W (previously called T_{tot} in Sec. 2.2).
- Eclipse timing T_0 (see Sec. 2.2).
- Orbital period P .
- $\sqrt{e \sin \omega}$ and $\sqrt{e \cos \omega}$ where e is the orbital eccentricity and ω is the argument of the periastron.
- The parameter $K_2 = K \sqrt{1 - e^2} P^{1/3}$ where K is the RV semi-amplitude.
- $\sqrt{v \sin i_s} \cos \beta$ and $\sqrt{v \sin i_s} \sin \beta$ where $v \sin i_s$ is the stellar rotational velocity along the line of sight and β is the projected angle between the stellar spin axis and orbital axis.
- Limb-darkening coefficients (see Sec. 2.2.1).

It is useful to discuss the choices of some jump parameters. Firstly, the code uses b' as jump parameter rather than the actual impact parameter $b = b' \frac{1-e^2}{1+e \sin \omega}$ since this choice minimises the correlation between the other jump parameters as shown in Gillon et al. (2009). Similarly, the parameter K_2 is used as a jump parameter instead of K to minimise such correlations. In fact,

K_2 results to be (Gillon et al., 2009): $K_2 = (2\pi G)^{1/3} \frac{M_p \sin i_p}{M_s^{2/3}} \left(1 + \frac{M_p}{M_s}\right)^{-2/3}$. It is independent from e and P and, for a given stellar mass, assuming a uniform prior on K_2 implies assuming a uniform prior on $M_p \sin i_p$.

Instead of directly sampling e and ω , the code uses $\sqrt{e \sin \omega}$ and $\sqrt{e \cos \omega}$ as jump parameters, which ensures both orthogonality and uniform priors on e and ω (Anderson et al., 2011). If RV time series are available, a good approach would be to try to characterise the system setting $e = 0$ and then leaving e as a free parameter (or setting an upper limit) to then check which approach is preferable using the BIC.

Let us note that so far only the β of one planet can be handled by the code also when dealing with multi-planetary systems. However, this is not too problematic: this parameter can be inferred from the Rossiter-McLaughlin effect which is not commonly observed for multiple planets belonging to the same system.

On the stellar side, two Limb-darkening laws are implemented within the code: the non linear model (*nl*, four LD coefficients) or the quadratic model (*qd*, two LD coefficients: u_1 and u_2). In case the *qd* model is employed, the actual jump parameters come from the following linear combination $c_1 = 2u_1 + u_2$ and $c_2 = u_1 - 2u_2$ to minimise their mutual correlation as shown by Holman et al. (2006). For each available filter, LD coefficients are inferred through interpolation in the tables of Claret & Bloemen (2011) further complemented by ATLAS9-based² LD grids computed using the code by Espinoza & Jordán (2015). Since these tables of *qd* and *nl* coefficients (stored within the `quadratic.dat` and `nonlinear.dat` text files, respectively) are affected by systematics, rather than fixing the LD coefficients at the interpolated values, the default approach of the code is to set a Gaussian prior to each LD coefficient.

Finally, imposing a Gaussian prior to both M_s and R_s (which gives an external constraint to ρ_s) may be useful in those cases where the LC data are not able to strongly constrain the impact parameter. In fact, in those cases both $\frac{a}{R_s}$ and ρ_s cannot be precisely determined from the observations alone.

3.3.2 Merit function

The decision of either accepting or rejecting a MCMC state is handled by a merit function, which is generally built upon a χ^2 -base in the case of a MCMC scheme. At each step the code computes a χ^2 fot the photometric data:

$$\chi_{\text{ph}}^2 = \sum_{i=1}^{n_{\text{LC}}} \sum_{j=1}^{np_i} \left(\frac{f_{ij} - \bar{f}_{ij}}{\sigma_{f_{ij}}} \right)^2, \quad (3.6)$$

where n_{LC} is the number of available LCs, np_i are the numbers of points of the i -th LC, f_i and σ_{f_i} are the vectors of flux measurements and respective errors referring to the i -th LC, while \bar{f}_i are the fluxes that have been computed

²<http://kurucz.harvard.edu/grids.html>

according to the adopted eclipse model and depend upon the jump parameters that are sampled at that state. Similarly, the code evaluates also

$$\chi_{\text{rv}}^2 = \sum_{i=1}^{n_{\text{RV}}} \sum_{j=1}^{n_{r_i}} \left(\frac{v_{ij} - \bar{v}_{ij}}{\sigma_{v_{ij}}} \right)^2, \quad (3.7)$$

which is written in analogy to Eq. (3.6), but referring to the RV time series. The sum of χ_{ph}^2 and χ_{rv}^2 gives the basic merit function.

In case some jump parameters are subject to Gaussian priors, a Bayesian penalty (BP) is added to the merit function:

$$\text{BP} = \left(\frac{x - \bar{x}}{\sigma_{\bar{x}}} \right)^2, \quad (3.8)$$

where \bar{x} and $\sigma_{\bar{x}}$ are the mean and standard deviation of the prior, while x is the value drawn at that given step of the MCMC.

Finally, the complete merit function, which is computed at each step, is:

$$\chi^2 = \chi_{\text{ph}}^2 + \chi_{\text{rv}}^2 + \sum_{p=1}^{n_{\text{pp}}} \text{BP}_p, \quad (3.9)$$

where n_{pp} is the number of jump parameters subject to a Gaussian prior.

As mentioned above, at each step of a MCMC, a new state is proposed; whether this state is accepted or not depends on the merit function. If the χ^2 at a given step is lower than the one computed at the previous step, this state is accepted. If that is not the case, the state is accepted with a probability given by $\exp(-\frac{1}{2}\Delta\chi^2)$ with $\Delta\chi^2$ being the χ^2 -difference between the last two steps. When the state is not accepted, the new state becomes a copy of the previous set of parameters.

In the MCMCI code the new state is generated by sampling a Gaussian distribution centred on the previous state. Its variance, that is the step size, depends on the acceptance rate which is the fraction of accepted states over a given number. During the burn-in phase only one parameter changes at each step, and Ford (2005) proved that the optimal acceptance rate in this case is $\sim 40\%$. After the burn-in phase, all the jump parameters are modified at a time, the code allows to select a desired values which typically is smaller the during the burn-in-phase.

Before I mentioned the BIC as a useful indicator when we need to compare different models and select the best one. We are now able to mathematically define it:

$$\text{BIC} = \hat{\chi}^2 + k \log N, \quad (3.10)$$

where $\hat{\chi}^2$ is the smallest χ^2 found at the end of the MCMC run, k is the number of free model parameters and N is the number of data points. We recall that the preferable model is the one with the smallest BIC.

The BIC is a sort of penalised χ^2 . On the one hand, by increasing the number of free model parameters we expect to lower down the χ^2 as it is easier

for the model to replicate the data set; on the other hand, one would risk to face overfitting issues. The BIC counterbalances these two aspects by quantifying if it is actually worth to add free parameters to the model.

One final aspect is checking whether the MCMC analysis is able to reach convergence. This is checked using the Gelman-Rubin statistics (Gelman & Rubin, 1992), \hat{R} , which compares the variance of a given parameter from each individual chain to the variance of the means of that parameter inferred from the different chains. Good convergence occurs when \hat{R} is close to 1, with Gelman & Rubin (1992) recommending $\hat{R} < 1.1$.

3.3.3 Steps to be performed

The analysis of LC and RV time series using the MCMCI code typically consist of the following steps:

1. Search for the best polynomial detrending baselines to be applied to each of the available LC and RV time series.
2. Preliminary run adopting the best detrending baseline to check whether photometric errors need to be inflated.
3. Final run that provides the posterior distributions of the relevant stellar and planetary parameters.

More specifically, the preliminary run has the goal of estimating the amplitude of both white and red noise of each light curve following Pont et al. (2006). This extra source of noise is quantified by correction factors (CFs) by which the flux errors should be multiplied to get reliable error bars. Once the final MCMCI run is performed accounting for the CFs, for each LC the average flux error is equal to the rms of the photometric residuals (that is data minus best-fit transit model). A similar approach enables the MCMC scheme to estimate the RV jitter, which takes care of any underestimation in the RV error budget and shall be summed in quadrature to the input RV errors.

3.3.4 Input files

In order to perform the joint LC+RV analysis, the MCMCI code requires an input file detailing the setup (`mcmc.dat`); one or more text files containing the RV time series; one or more text files containing the light curves.

The `mcmc.dat` reports some crucial information such as the number of RV time series and LC to be analysed, the orders of the detrending polynomials, the number of MCMC steps, the kind of prior to impose on the jump parameters along with their initial guess values.

The files containing the RV time series (.txt extension) are called `rv` followed by four digits which represent the progressive number of the RV time series (starting from 1); if necessary the left-most digits are padded with zeros. As an example, the file containing the first RV time series is called `rv0001.txt`.

Each of these files contains eight columns. The first three columns are essential, as they refer to the date of observation, the measured radial velocity and its error. The epoch of observation is expressed in Julian date (JD) shifted by $-2\,450\,000$, while both the RV measurements and the errors shall be provided in $\text{km}\,\text{s}^{-1}$. The other columns are: the CCF width, the CCF contrast, the CCF asymmetry, a further activity indicator (e.g. $\log R'_{\text{HK}}$) and the exposure time (in seconds). If some of these quantities are not available, a column of zeros is provided.

The text files containing the LCs are built similarly, but their filename starts with `phot` (e.g. `phot0001.txt`). They contain 11 columns where the first three are always needed. These three columns contain the observation epoch (in JD $-2\,450\,000$), the measured flux and its error. Columns 4—9 contain vectors of ancillary parameters relevant for the detrending, while the last two columns report the airmass and the exposure time. If a parameter is not available, that column is set to zero.

It is also possible to perform a TTV analysis to check the significance of transit timing variations caused by the mutual gravitational pull of the planets in the system. In this case, for each planet, an additional input file must be provided. These input files are text files with the extension `.dat`, whose name starts with `ttv` followed by two digits which represent the progressive number of the planet that file is referring to (e.g. `ttv01.dat`). The `ttv` input files contain three columns, while the number of rows is equal to the number of times that planet transits in all the available input LCs. For each row (i.e. for each transit event) the first column contains an integer number N_{te} computed as

$$N_{\text{te}} = \text{round} \left(\frac{t_{\text{te}} - T_0}{P} \right), \quad (3.11)$$

where t_{te} is the transit timing of that given transit event, P is the planet orbital period, while T_0 is the guess transit timing as reported in the `mcmc.dat`. As we are dealing with guess values, it is likely that $(t_{\text{te}} - T_0)/P$ is not exactly an integer, hence the need for the `round` function. Finally, the second and third columns contain initial guess values (both in days) for the TTVs and their errors, respectively.

Chapter 4

The Planetary System TOI-396

4.1 System Overview

This work will focus on the characterization of the exoplanetary system orbiting around the star TOI-396, also known as HR-858 or TIC-178155732. These names refer to different catalogues where this star is present, with TOI standing for “*TESS* object of interest”, HR for “Harvard Revised photometry” and TIC for “*TESS* input catalogue”. A first analysis of this system was done by Vanderburg et al. (2019), therefore I will firstly introduce TOI-396 following their work.

TOI-396 is a F6 V (Gray et al., 2006) bright naked-eye star with an apparent visual magnitude of $V = 6.4$ mag (Høg et al., 2000). It is located 31.7 pc away, having a parallax $\tilde{\omega} = 31.542$ mas (Gaia Collaboration et al., 2022), and can be visible in the constellation of Fornax in the southern hemisphere. It is member of a visual binary system and its companion is a faint M-dwarf ($G \sim 16$ mag), about $8'',4$ away. The analysis carried by Vanderburg et al. (2019) suggests that TOI-396 has a mass of $\sim 1.1 M_{\odot}$, a radius of $\sim 1.3 R_{\odot}$ and is rotating rapidly, having a measured radial velocity of $v \sin i = 8.3 \text{ km s}^{-1}$. Using *TESS* (*Transiting Exoplanet Survey Satellite*; Ricker et al., 2014) light curves, Vanderburg et al. (2019) found three planets orbiting around this star, namely HR-858 b, c and d. These planets are all about twice the size of Earth (they are therefore referred to as super Earths) and have an estimated orbital period of ~ 3.6 , ~ 6.0 and ~ 11.2 days. HR-858 b and c may be in mean motion resonance as they orbit within 0.03% of the 3:5 period ratio.

"SIMPLE"	"T"
"BITPIX"	"-32"
"NAXIS"	"1"
"NAXIS1"	"91"
"EXTEND"	"T"
"CDELT1"	"0.01"
"CTYPE1"	"%"
"CRVAL1"	"0.05"
"CTYPE1"	"km/s"
"ORIGIN"	"TEST"
"DATE"	"2019-01-31T00:36:08.576"
"TELESCOP"	"ESO-3P6"
"INSTRUME"	"HARPS"
"OBJECT"	"TOI-396"
"RA"	"42.984668"
"DEC"	"-30.81187"
"EQUINOX"	"2000."
"RADECSYS"	"FK5"
"EXPTIME"	"382.9981"
"MJD-OBS"	"58514.02064717"
"DATE-OBS"	"2019-01-31T00:29:43.915"
"UTC"	"1779.000"
"LST"	"15989.839"
"PI-COI"	"Armstrong"
"ORIGFILE"	"HARPS_ech_obs_all_0001.fits"
"DATASUM"	"0"

Figure 4.1: A portion of the header, as extracted from a *HARPS* .fits file

4.2 Data analysis

4.2.1 Radial velocity setup

Differently from Vanderburg et al. (2019), in this work the complete analysis of the system is achieved using both the transit method and the radial velocity technique. Specifically, the RV analysis is based on 78 *HARPS* spectra taken in 2019 between 31st January and 27th July, as kindly provided by the *HARPS* consortium. *HARPS* is a radial velocity spectrograph (Mayor et al., 2003) dedicated to the discovery of exoplanets. It is attached to the ESO 3.6-metre telescope in La Silla and it was installed in 2003. *HARPS* is able to reach a very high long term radial velocity accuracy on the order of 1 m/s¹.

Data obtained from *HARPS* spectra are available as .fits files (Wells et al., 1981, Harten et al., 1988). I wrote an R (R Core Team, 2019) script to read these files and extract the desired information. Each of these files contains a header (see an example in Fig. 4.1) and a matrix made of 241 rows and 73 columns storing the CCF values. In particular, each row represents the pixel number on the spectrograph detector, while each of the first 72 columns represents a wavelength. The last column contains the CCF integrated over the entire wavelengths range, that is the sum of all the previous columns. The *HARPS* pipeline also provides us with an auxiliary text file that contains the observation time (expressed in BJD_{TDB}) and different stellar activity indicators, namely log R'_{HK} , S-index, H_{α} , and NaI D.

From the header of each .fits file I first extracted two parameters called CRVAL1 and CDELT1, which are needed to convert the pixel number in the

¹<https://www.eso.org/sci/facilities/lasilla/instruments/harps/overview.html>.

physical RV measurement according to

$$\text{RV}_n = \text{CRDELT1} + n\text{CRVAL1}, \quad (4.1)$$

where n is the n -th pixel starting from zero. After extracting the last matrix column from each `fits` file and applying Eq. (4.1), I ended up with 78 CCFs, where the cross-correlation between the stellar spectrum and the template mask is expressed as a function of the stellar RV.

I then wrote an `R` routine to fit each CCF with a skew-normal function (Sec. 3.1.2). From the fit I extracted the median, the asymmetry γ , the FWHM_{SN} , and the contrast of the CCF. With these data in hands, I could built the input `rv0001.txt` file as required by the MCMCI code, following the format described in Sec. 3.3.4.

The following key aspect of the RV analysis I performed is the break-point method (see discussion in Sec. 3.2). I tested this method on my RV time series by allowing up to five break-points; the BIC favours the presence of a single break point at observation 48, that is where the bp method has identified a significant change in the stellar activity.

4.2.2 Photometry setup

TESS is an MIT-led NASA mission launched in April 2018 that searches for transiting exoplanets orbiting bright close-by stars. *TESS* aims at observing the whole sky by focusing on individual sectors of $24^\circ \times 96^\circ$ at a time. Each sector is observed nearly uninterruptedly for about 27 days². The southern ecliptic hemisphere was observed during year 1 of the mission (from July 2018 to July 2019; sectors 1–13). Instead, the northern hemisphere was observed during year 2 (from July 2019 to July 2020; sectors 14–27). Afterwards, observations were repeated.

TOI-396 was observed by *TESS* during Sectors 3 and 4 in 2018 and then during Sectors 30 and 31 in 2020. Sector 3 data were collected with a cadence of 30 min, while all the other sectors are characterised by a 2 min cadence (see table 4.1 for further details). In this work we analysed the transit events as coming from all the four available sectors (3, 4, 30 and 31), while Vanderburg et al. (2019) could analyse photometric data from Sectors 3 and 4 only.

TESS observations can be retrieved as `.fits` file format from the MAST portal³ after inputting the TIC identifier of the star. I then wrote an `R` routine which is able to

- read the `.fits` *TESS* observations;
- reject all the data points marked with a bad-quality flag and to further discard photometric outliers according to a 5-MAD-clipping⁴;

²<https://tess.mit.edu/observations/>

³<https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html>

⁴MAD: median absolute deviation

Table 4.1: Summary of the four *TESS* sectors containing the observations of TOI-396. This table lists the starting and ending date of data acquisition, the total number of data points and the number of data points actually used after applying both the quality-flag and the MAD rejection and after keeping only those temporal windows centred around each transit event.

Sector	Start	Stop	Cadence [min]	Points [#]	Used points [#]
3	2018 Sep 21	2018 Oct 17	30	1288	250
4	2018 Oct 19	2018 Nov 14	2	18684	3152
30	2020 Sep 23	2020 Oct 20	2	19687	3464
31	2020 Oct 22	2020 Nov 18	2	18314	3417

- split the entire *TESS* sectors into chunks, which possibly contain one single transit event (unless two or more planets are transiting almost simultaneously).

Each chunk of data extracted from the *TESS* sectors contain one transit event (or more in case of almost overlapping transits) plus 4 h of observations both before and after the transit event for detrending purposes. For each chunk of photometric data, the R routine is also in charge of building one `phot` input file following the format detailed in Sec. 3.3.4. In particular, from the *TESS* `.fits` files I extracted the following columns

- `TIME`: the epoch of observation.
- `PDCSAP_FLUX`: stellar flux, as corrected by the *TESS* pipeline, except for Sector 3 for which only the raw flux was available (`FLUX_RAW`).
- `PDCSAP_FLUX_ERR`: flux error (in case of Sector 3 the variable is `FLUX_RAW_ERR`).
- `MOM_CENTR1` and `MOM_CENTR2`: the column and row position of target's flux weighted centroid, hereafter denoted with `x` and `y`, respectively.
- `Pos_CORR1` and `Pos_CORR2`: the column and row local motion differential velocity aberration, pointing drift, and thermal effects, hereafter denoted with `dx` and `dy`, respectively.

From this vectors, I produced a total of 41 `phot` files, each reporting the epoch of observation, the flux and its error normalised with respect to the median, plus the detrending vectors `x`, `y`, `dx`, and `dy`. A summary of the 41 input LCs is presented in table 4.2.

Table 4.2: Summary of the input file name with respect to the sector of origin and planet (or planets) that it contains (e.g. TOI-396 b is simply referred to as b). Sector 3 had a 30 min cadence while sectors 4, 30, 31 a 2 min cadence.

phot name	Sector	Planet	phot name	Sector	Planet
0001	3	b	0022	30	d
0002	3	c	0023	30	b
0003	3	b, d	0024	30	c
0004	3	b	0025	30	b
0005	3	b	0026	30	b, c
0006	3	b	0027	30	c
0007	3	c	0028	30	b
0008	3	b, d	0029	30	b
0009	3	b	0030	30	c
0010	3	c	0031	30	d
0011	3	b	0032	30	b
0012	4	b	0033	31	b
0013	4	c	0034	31	c, d
0014	4	b	0035	31	b
0015	4	c	0036	31	b
0016	4	b	0037	31	c
0017	4	c	0038	31	b, c, d
0018	4	b	0039	31	b
0019	4	d	0040	31	c
0020	4	c	0041	31	b
0021	4	b			

Chapter 5

Results

5.1 Stellar parameters

The stellar age, mass, and radius were derived homogeneously by A. Bonfanti using the isochrone placement algorithm (Bonfanti et al., 2015, 2016) and its capability of interpolating a flexible set of input parameters within pre-computed grids of PARSEC¹ v1.2S (Marigo et al., 2017) and CLES (Code Liègeois d’Évolution Stellaire, Scuflaire et al., 2008) isochrones and tracks. The input set for interpolating within PARSEC grids is constituted by the spectroscopic parameters reported in Vanderburg et al. (2019) (i.e. T_{eff} , [Fe/H], $\log g$, and $v \sin i$), the G band magnitude, and the Gaia parallax (Gaia Collaboration et al., 2022, offset-corrected following Lindegren et al. 2021). In the case of CLES-based grids, instead, only the spectroscopic parameters were inputted as the CLES code does not provide any magnitude. The respective pairs of outcomes from the two independent analyses were compared using the χ^2 -based test described in Bonfanti et al. (2021) and finally the results were combined together obtaining the parameters synthesized in table 5.1.

5.2 Detrending baseline

As anticipated in Sec. 3.3.1, the MCMCI code performs data detrending using polynomials. For each detrending parameter and for each LC, I varied the polynomial order from 0 up to 4 and checked the corresponding BIC outputted by each MCMCI mini-run made of 10,000 steps. The combination of polynomial orders corresponding to the lowest BIC is reported in table 5.2 and constitutes the reference detrending baseline that has been adopted in all the following analyses.

Similarly, I searched the best detrending baseline for the radial velocity data. As previously mentioned, the detrending parameters contained in the `rv` input

¹<http://stev.oapd.inaf.it/cgi-bin/cmd>

Table 5.1: Synthesis of the stellar parameters.

Parameter	Value	Source
T_{eff} (K)	6201 ± 50	Vanderburg et al. (2019)
[Fe/H] (dex)	-0.14 ± 0.08	Vanderburg et al. (2019)
$\log g$ (cm s^{-2})	4.19 ± 0.1	Vanderburg et al. (2019)
$v \sin i$ (km s^{-1})	8.3 ± 0.5	Vanderburg et al. (2019)
G (mag)	6.26498 ± 0.00021	Gaia Collaboration et al. (2022)
$\tilde{\omega}$ (mas)	31.564 ± 0.035	Gaia Collaboration et al. (2022)
M_s (M_\odot)	$1.096^{+0.054}_{-0.057}$	isochrones
R_s (R_\odot)	$1.310^{+0.014}_{-0.013}$	isochrones
age (Gyr)	4.4 ± 1.1	isochrones
L_s (L_\odot)	2.277 ± 0.087	from T_{eff} & R_s

file are the observation time, FWHM_{SN} , γ , the contrast, and a further activity indicator. Since four different activity indicators were available from the *HARPS* pipeline (i.e. $\log R'_{\text{HK}}$, S-index, H_α , NaI D), I created four different `rv0001.txt` files whose fifth column contained one of these activity indicators at a time and then I studied the detrending baseline separately. However, only the CCF-related activity indicators were actually effective in lowering down the BIC, while none of the other indicators turned out to further improve the detrending. The final adopted polynomial baseline for detrending the RV time series is summarised in table 5.3.

5.3 Setup of the `mcmc.dat`

The relevant parameters for setting up the MCMC process shall be specified in the `mcmc.dat` input file. After performing a first preliminary run made of 300,000 steps to derive the correction factors as anticipated in Sec. 3.3.3, for the final analysis I opted for 3 independent runs with 300,000 steps per run. As mentioned in Sec. 3.3, it is usual practice to discard the first few steps of each run. To this end, I set the burn-in length equal to 20% of the number of steps, following Bonfanti & Gillon (2020).

After that, for each light curve to be analysed (i.e. for each `phot` file), I specified the photometric filter (the *TESS* passband in our case), the sequence of polynomial orders constituting the detrending baseline (as reported in table 5.2), and the correction factor previously computed as described in Sec. 3.3.3 to properly blow up the photometric error bars. In addition, for those light curves extracted from Sector 3 (whose cadence was 30 min), I opted for an over-sampling of 120 s, which is the same cadence characterising the other three sectors. In fact, Kipping (2010) proves that long-cadence photometry can cause the retrieval of erroneous system parameters, with a sensible underestimation e.g. of the mean stellar density inferred from the transit fitting.

Table 5.2: Photometric detrending baseline. The first column reports the `phot` name of each light curve and the listed numbers are the polynomial orders to be associated to the different detrending terms.

phot name	time	dx	dy	x-y
0001	1	0	1	0
0002	0	0	0	4
0003	0	0	0	0
0004	1	0	2	0
0005	0	0	0	0
0006	3	0	0	0
0007	0	0	2	0
0008	0	0	0	0
0009	1	0	0	0
0010	1	0	0	0
0011	1	0	0	0
0012	0	1	0	0
0013	0	0	1	0
0014	0	0	0	0
0015	2	0	1	1
0016	0	1	0	0
0017	1	0	0	0
0018	2	0	0	0
0019	0	0	0	0
0020	4	0	0	1
0021	0	0	1	0
0022	1	0	0	0
0023	0	0	0	0
0024	0	0	0	0
0025	2	0	0	0
0026	0	0	1	0
0027	0	0	0	0
0028	0	0	0	0
0029	0	0	0	0
0030	1	2	0	0
0031	0	2	0	1
0032	0	0	0	0
0033	0	0	0	0
0034	0	1	0	1
0035	1	0	0	0
0036	1	0	0	0
0037	2	0	0	0
0038	0	1	0	0
0039	0	0	0	0
0040	1	0	0	0
0041	0	0	0	0

Table 5.3: RV detrending baseline. First row reports the detrending parameters and second row the polynomial orders adopted for the optimal detrending of the RV time series.

time	FWHM_{SN}	γ	contrast
3	1	4	1

On the RV side, I specified that I have one RV time series (i.e. the `rv0001.txt` file), and I then filled in the `mcmc.dat` with the polynomial orders of the RV detrending baseline (see table 5.3) along with a guess value for the RV jitter as inferred from the preliminary run. I also specified to detrend the RV data via a piece-wise polynomial interpolation by setting one break-point at observation 48, as it turned out from the analysis in Sec. 4.2.1.

The `mcmc.dat` setup also requires to specify the jump parameters within the MCMC and the kind of prior they are subject to. I imposed a Gaussian prior on the stellar mass, radius, metallicity and effective temperature using the values reported in table 5.1 to define the mean and rms of the Gaussian distribution. The reason is twofold: first to impose a prior on ρ_s (via the priors on both M_s and R_s) to drive the convergence of the transit fitting; second to retrieve the best-fit LD coefficients following interpolation within the `quadratic.dat` grid, as I assume a *qd* LD model in all the analyses. In this regard, a Gaussian prior on the LD coefficients was imposed as well, according to the parameterisation outlined in sec. 3.3.1. Instead, I imposed uniform unbounded priors (except for the physical limits) on the jump parameters of three transiting planets, namely the transit depth, the impact parameter, the period, the transit timing, the RV semi-amplitude, the eccentricity, and the argument of periastron. The specific re-parameterisation of the jump parameters are discussed in Sec. 3.3.1.

I carried out two different analyses. During my first analysis I assumed linear ephemerides, that is absence of TTVs (Sec. 5.4). In my second analysis, instead, I assumed that the planets could be subjected to TTVs (Sec. 5.6).

5.4 MCMC Analysis assuming linear ephemerides

Assuming linear ephemerides, I first executed different runs varying the eccentricity setup of each planet according to the following options:

- Null eccentricity, $e = 0$ (fixed).
- Eccentricity is a jump parameter, with e that may vary from 0 up to 1.
- e has an upper limit of 0.15.

For each simulation I compared the BIC and it turned out that the BIC was smaller when assuming null eccentricity for all the three planets. Therefore, I will set circular orbits for all the planets in all the following analyses.

Table 5.4: Main parameters of TOI-396 b, c and d, as derived from the MCMCI run when assuming linear ephemerides and null eccentricity. The RV semi-amplitude, mass and density of TOI-396 c are reported as their 3σ upper limits because of their non-detection.

Parameter	b	c	d
P (days)	$3.585291^{+0.000007}_{-0.000010}$	$5.973858^{+0.000015}_{-0.000017}$	$11.230513^{+0.000034}_{-0.000028}$
T_0 (BJD _{TBD})	$8409.18801^{+0.00060}_{-0.00071}$	$8415.6352^{+0.0014}_{-0.0011}$	$8409.7334^{+0.0015}_{-0.0020}$
δ_{tra} (ppm)	207 ± 7	205 ± 9	220 ± 10
K_p (m s^{-1})	1.31 ± 0.34	< 1.25	$1.79^{+0.41}_{-0.40}$
R_p (R_\oplus)	$2.057^{+0.042}_{-0.041}$	2.048 ± 0.050	2.121 ± 0.056
M_p (M_\oplus)	3.33 ± 0.88	< 3.84	$6.7^{+1.6}_{-1.5}$
ρ_p (g cm^{-3})	$2.11^{+0.58}_{-0.57}$	< 2.54	$3.85^{+0.99}_{-0.90}$
T_{eq} (K)	1574^{+21}_{-21}	1328 ± 17	1076 ± 14

The joint analysis of RV data (Sec. 4.2.1) and photometric data (Sec. 4.2.2) reached convergence according to the Gelman-Rubin test. Specifically, the Gelman-Rubin statistic yields $\hat{R} \lesssim 1.004$ for each jump parameter.

If compared with the outcomes from Vanderburg et al. (2019), I improved the transit depth precision of all the three planets by a factor ~ 1.4 . Moreover, by combining the SN-fit onto *HARPS* CCFs along with the breakpoint method, I was able to provide the first mass estimates of TOI-396 b and TOI-396 d (at the 3.8 and 4.3 σ -level, respectively). Instead, I could not find any significant Keplerian signal of TOI-396 c within the RV time series; the possible reasons for this null detection are discussed in the following sections.

The output of this analysis is summarised in table 5.4. The posterior distributions of the orbital periods of TOI-396 b, c and d are shown in Fig. 5.1, where – as already done by Vanderburg et al. (2019) – I note that the ratio between P_c and P_b is commensurable, being close to 5:3. Finally, the RV semi-amplitude of TOI-396 c is displayed in Fig. 5.2.

5.5 Discussion on the mass non-detection of TOI-396 c

Given the essentially null-detection of the RV semi-amplitude of planet c (see Fig. 5.2), K_c and the consequent M_c and ρ_c are reported as 3σ upper limits in table 5.4. In fact, the median value inferred from the ρ_c -posterior distribution, that is $\rho_c = 0.54^{+0.56}_{-0.38} \text{ g cm}^{-3}$, looks unreliable and extremely low. Such a low density could be justified e.g. assuming an extended atmospheric envelope, which is, however, at odd with the high equilibrium temperature retrieved from the MCMCI run, that is $T_{\text{eq},c} = 1328 \pm 17$ K.

A confirm that any RV periodic signal with $P = P_c \sim 6$ d cannot be detected comes from the GLS (see Sec. 2.3.2) periodogram analysis. In fact, I took the

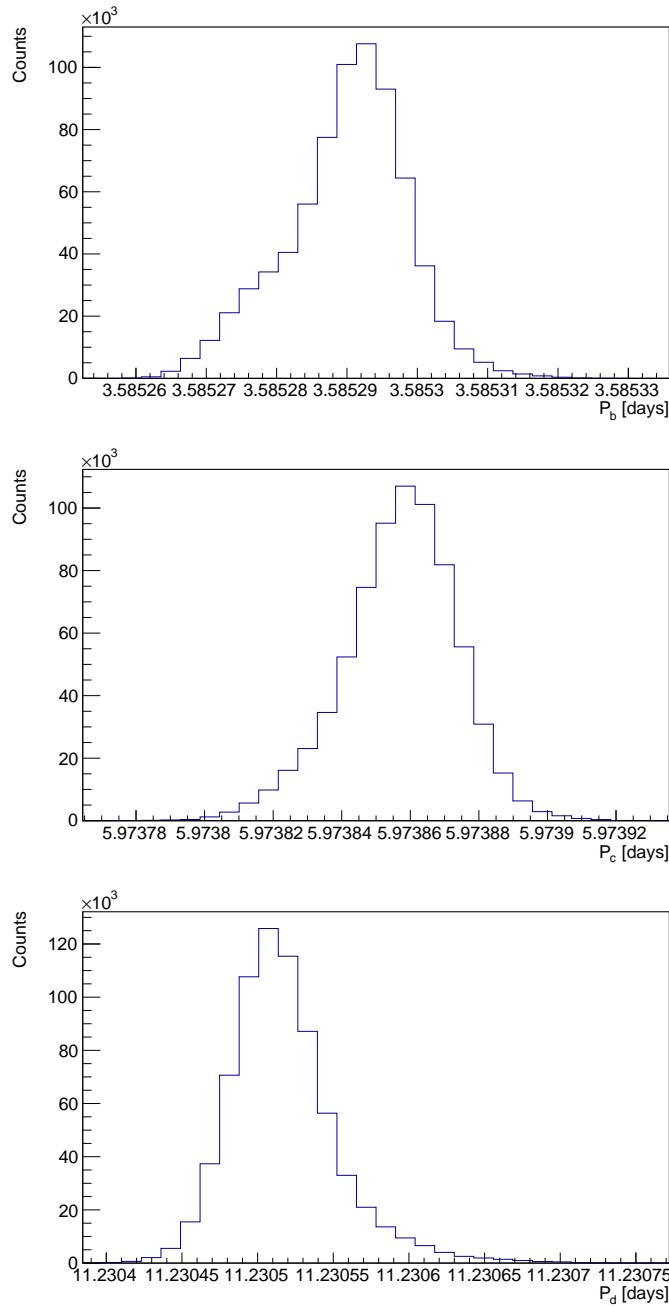


Figure 5.1: Posterior distributions for the orbital periods of TOI-396 b, c and d when assuming linear ephemerides.

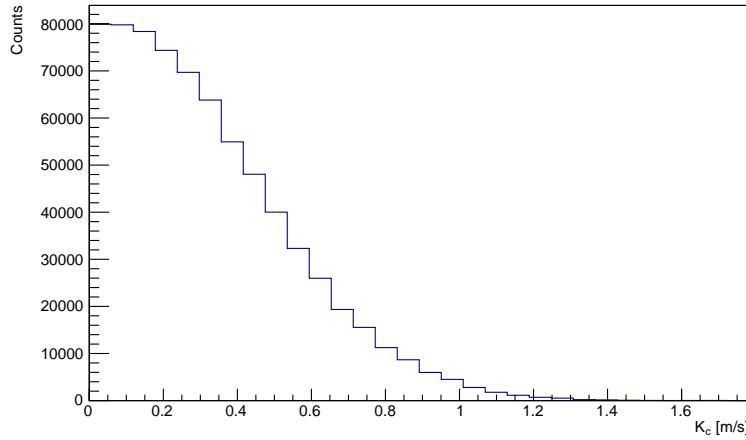


Figure 5.2: Posterior distribution of TOI-396 c RV signal semi-amplitude when assuming linear ephemerides.

detrended RV time series and for each planet I phase-folded it according to the corresponding orbital period after subtracting the Keplerian signals of the other two planets. For each of the three phase-folded RV time series, I then computed the GLS periodogram using the PYTHON 3 (Van Rossum & Drake, 2009) routine `gls.py`² (Zechmeister & Kürster, 2009). After this first run, I removed from the phase-folded RV time series the sinusoidal harmonic causing the highest peak of the periodogram and I executed the `gls.py` again (second run) to further check whether the “residual” RV time series contained additional significant peaks. The results of this analysis are plotted in Fig. 5.3, where both the periods of planet b and d are clearly detected after the first run with a FAP lower than 1% and 0.1%, respectively (first column; first and third panel). Instead, there is not any periodic signal compatible with the period of planet c, even after the second run of the `gls.py` routine (second row of Fig. 5.3).

Possible scenarios for which TOI-396 c is not detected via the RV technique are investigated in the following sections.

5.5.1 Hypothesis 1: TOI-396 c is a false-positive

To explain why planet c is not detected in the RV time series, one may wonder whether TOI-396 c is a false positive. Vanderburg et al. (2019) already conducted a false-positive analysis, but they finally classified TOI-396 b, c and d as validated planets. In particular, they tested the following scenarios:

- *TOI-396 is an eclipsing binary.* They obtained high-resolution reconnaissance spectroscopy and they ruled out large ($\sim \text{km s}^{-1}$) RV variations

²The `gls.py` routine is public and available at <https://github.com/mzechmeister/GLS>

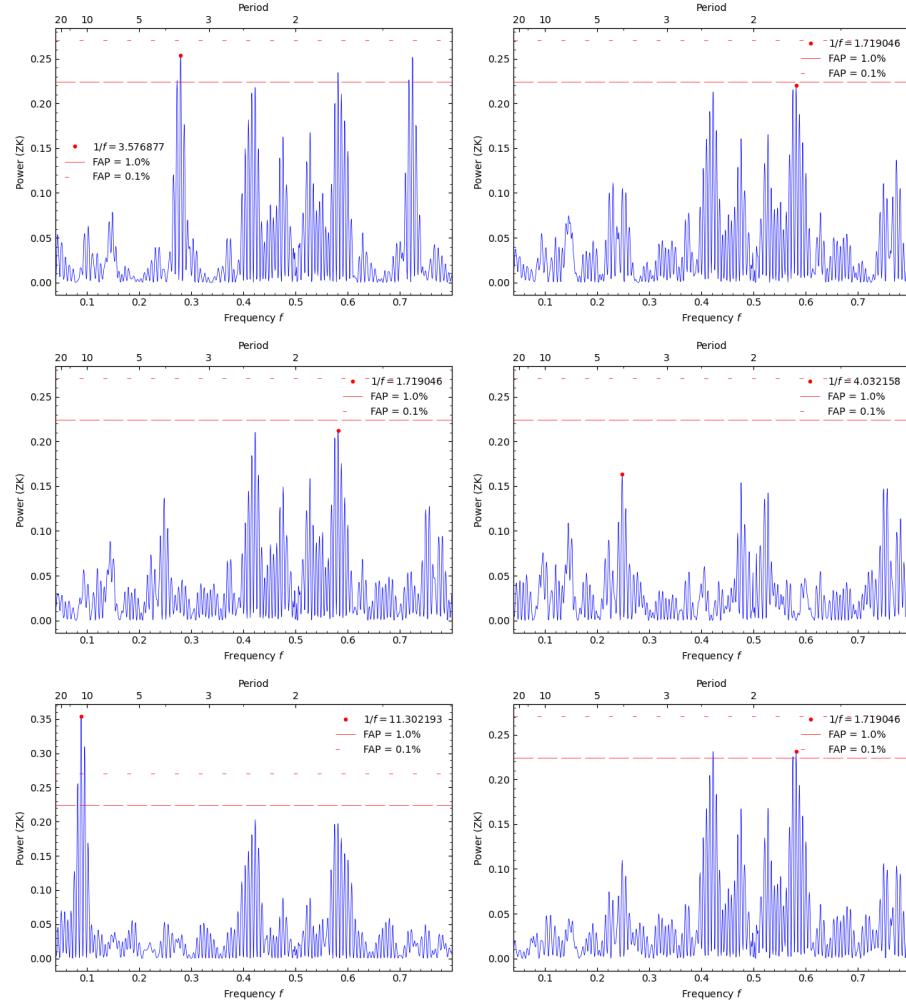


Figure 5.3: GLS periodograms computed from the detrended RV time series after phase-folding it according to the orbital periods of planet b (first row), planet c (second row), and planet d (third row). The second column show the GLS recomputation after subtracting the periodic signal having the highest peak from the RV time series.

which could indicate that the star is a close binary.

- *Light either from an unassociated eclipsing binary or a transiting planet system is blended with TOI-396.* After inspecting the sky region around TOI-396 using archival surveys, they found no evidence of stars that could be both close and bright enough to contribute to the transit signal.
- *Light from an associated companion (eclipsing binary or transiting planet system) is blended with TOI-396.* They exclude this possibility as the comoving companion, TOI-396 B, is too faint and there is no evidence of other resolved or unresolved companion the could be bright enough.

Since Vanderburg et al. (2019) could not exclude all the possible false-positive scenarios involving physically associated companions, they assigned a probabilistic metric to the false-positive scenarios using VESPA (Morton, 2012, 2015). This tool accounts for the constraints given by *TESS* LCs, imaging, and spectroscopy and computes false-positive probabilities (FPPs). For all the three planets VESPA found $\text{FPP} < 10^{-3}$, which suggests that their transiting signals are genuine rather than false-positive.

5.6 MCMC Analysis assuming TTV

So far we concluded that

- all the transit signals are genuine;
- TOI-396 b and c lie on resonance orbits, that is the P_c/P_b period ratio is commensurable.

Since gravitational interactions are expected between planets in resonant orbits, it is worth to investigate whether planet b and c exhibit TTVs, which may constitute a final proof that TOI-396 c belongs to the system as well, despite its non-detection in the RV time series. Therefore, I ran the MCMCI code again using the setup discussed in Sec. 5.3, but this time enabling the possibility of TTVs. To this end, I provided the `ttv01.dat`, `ttv02.dat`, and `ttv03.dat` input files as explained in Sec. 3.3.4, with the reference P and T_0 as coming from the run assuming linear ephemerides (Sec. 5.4). I performed four independent runs of 200,000 steps each and checked the convergence of all the jump parameters via the Gelman-Rubin test ($\hat{R} \lesssim 0.002$).

The three panels of Fig. 5.4 show the TTV amplitudes of TOI-396 b, c and d, whose details are reported in table 5.5. For each TTV set of measurements ($i = 1, \dots, N_{\text{tr},j}$) referring to the j -th planet, I then computed the corresponding reduced- χ^2_j as:

$$\chi^2_j = \frac{\sum_{i=1}^{N_{\text{tr},j}} (\text{TTV}_{j,i}/\sigma_{j,i})^2}{N_{\text{tr}}}, \quad (5.1)$$

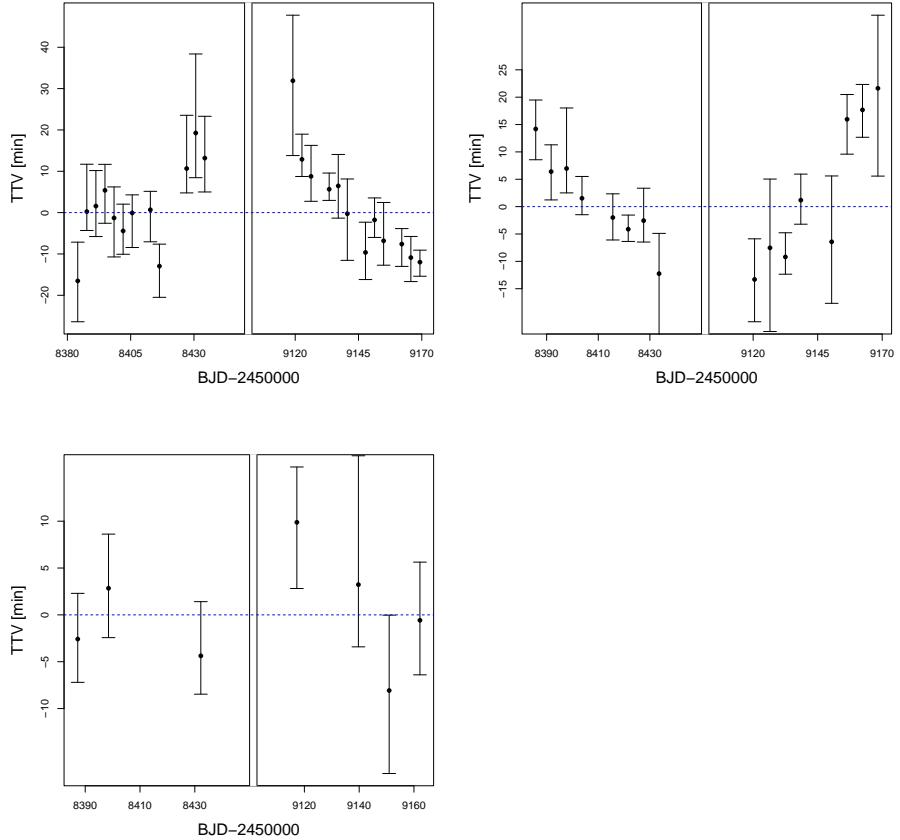


Figure 5.4: TTV amplitudes distribution for TOI-396 b, c and d.

where $\text{TTV}_{j,i}$ is the TTV amplitude of the i -th transit, $\sigma_{j,i}$ is its respective uncertainty, and $N_{\text{tr},j}$ is the number of transits of the j -th planet. By applying Eq. (5.1) to the three planets I found:

$$\chi_b^2 = 2.2 \quad \chi_c^2 = 3.0 \quad \chi_d^2 = 0.67, \quad (5.2)$$

where χ_b^2 and χ_c^2 sensibly differ from unity, hence the registered TTVs of TOI-396 b and c are statistically significant.

Moreover, the TTVs of TOI-396 b and c show an anti-correlation pattern, which is visible from the first two panels of Fig. 5.4 and is even more evident from Fig. 5.5, where I superimposed the TTV measurements. This kind of trend gives a solid proof that TOI-396 c belongs to the system, as that TTV pattern is theoretically expected in case of exoplanets in mean motion resonance. Indeed, by computing the orbital period ratio using the P_b and P_c values derived in

Table 5.5: TTV amplitudes for each transit of TOI-396 b, c and d, which are identified through the epoch N , as computed from Eq. (3.11).

TOI-396 b		TOI-396 c	
N	TTV amplitude (min)	N	TTV amplitude (min)
-7	-16.5 ^{+9.4} _{-9.9}	-5	14 ⁺⁵ ₋₆
-6	0.2 ⁺¹¹ _{-4.6}	-4	6.4 ^{+4.9} _{-5.2}
-5	1.6 ^{+8.6} _{-7.4}	-3	7.0 ⁺¹¹ _{-4.5}
-4	5.4 ^{+6.3} _{-8.0}	-2	1.5 ^{+4.0} _{-3.0}
-3	-1.3 ^{+7.5} _{-9.4}	0	-2.0 ^{+4.3} _{-4.1}
-2	-4.4 ^{+6.5} _{-5.6}	1	-4.1 ^{+2.6} _{-2.2}
-1	0.0 ^{+4.4} _{-8.4}	2	-2.6 ^{+5.9} _{-3.9}
1	0.7 ^{+4.5} _{-7.7}	3	-12 ⁺⁷ ₋₁₂
2	-13 ⁺⁵ ₋₈	118	-13 ⁺⁷ ₋₈
5	11 ⁺¹³ ₋₆	119	-8 ⁺¹³ ₋₁₅
6	19 ⁺¹⁹ ₋₁₁	120	-9.2 ^{+4.4} _{-3.2}
7	13 ⁺¹⁰ ₋₈	121	1.2 ^{+4.8} _{-4.4}
198	32 ⁺¹⁶ ₋₁₈	123	-6 ⁺¹² ₋₁₁
199	13 ⁺⁶ ₋₄	124	16 ⁺⁵ ₋₆
200	8.8 ^{+7.5} _{-6.0}	125	18 ⁺⁵ ₋₅
202	5.7 ^{+3.9} _{-2.7}	126	22 ⁺¹³ ₋₁₆
203	6.5 ^{+7.6} _{-7.8}	TOI-396 d	
204	-0.3 ^{+8.4} ₋₁₁	N	
206	-9.6 ^{+7.3} _{-6.6}	TTV amplitude (min)	
207	-1.8 ^{+5.3} _{-4.2}	-2	-2.6 ^{+4.9} _{-4.6}
208	-6.8 ^{+9.3} _{-5.9}	-1	2.8 ^{+5.8} _{-5.3}
210	-7.6 ^{+3.7} _{-5.4}	2	-4.4 ^{+5.8} _{-4.1}
211	-11 ⁺⁵ ₋₆	63	9.9 ^{+5.9} _{-7.1}
212	-12 ± 3	65	3.2 ⁺¹⁴ _{-6.6}
		66	-8.1 ^{+8.0} _{-8.9}
		67	-0.6 ^{+6.2} _{-5.8}

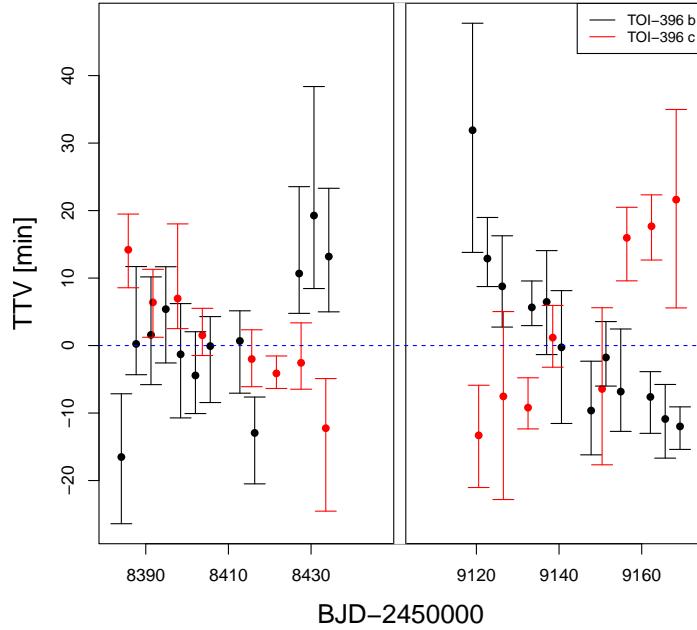


Figure 5.5: Overlapping of TTV amplitude for planets TOI-396 b and c.

Sec. 5.4, I found:

$$\frac{P_c}{P_b} \simeq 1.666213, \quad (5.3)$$

which differs from a 5:3 ratio by $\Delta = \frac{5/3 - P_c/P_b}{5/3} \simeq 0.03\%$.

In the following sections I will discuss in detail the RV and photometric results. They represent the adopted outcomes characterising the TOI-396 exoplanet system (see table 5.6).

5.6.1 Radial velocity

The RV results obtained in the final analysis of my work when considering TTVs are displayed in Figures 5.6, 5.7, and 5.8, which show the detrended RV time series phase-folded according to the orbital periods P_b , P_c , and P_d , respectively. To show the single Keplerian contribution of one planet in each Figure, the Keplerian signals of the other two planets have been removed.

I derived RV semi-amplitudes $K_b = 1.31 \pm 0.34 \text{ m s}^{-1}$, $K_c = 0.28_{-0.19}^{+0.29} \text{ m s}^{-1}$, and $K_d = 1.80_{-0.40}^{+0.41} \text{ m s}^{-1}$ for TOI-396 b, c, and d, respectively. These measurements imply planet masses $M_b = 3.34 \pm 0.88 M_\oplus$, $M_c = 0.84_{-0.59}^{+0.88} M_\oplus$, and

Table 5.6: Planetary parameters as retrieved from the MCMCI run when assuming TTVs. These results hold for circular orbits, which was the configuration favoured by the BIC. P and T_0 were derived in the MCMCI run when assuming linear ephemerides. The 3σ upper limits for the mass and the density of TOI-396 c are $M_{\oplus}^{\text{up}} = 3.84M_{\oplus}$ and $\rho_c^{\text{up}} = 2.54 \text{ g cm}^{-3}$, respectively.

Parameter	b	c	d
P (days)	$3.585291^{+0.000007}_{-0.000010}$	$5.973858^{+0.000015}_{-0.000017}$	$11.230513^{+0.000034}_{-0.000028}$
T_0 (BJD _{TBD})	$8409.18801^{+0.00060}_{-0.00071}$	$8415.6352^{+0.0014}_{-0.0011}$	$8409.7334^{+0.0015}_{-0.0020}$
δ_{tra} (ppm)	$216.3^{+7.2}_{-7.1}$	$208.8^{+9.0}_{-8.7}$	223 ± 11
b	$0.662^{+0.023}_{-0.024}$	$0.752^{+0.016}_{-0.017}$	0.768 ± 0.020
K_p (m s ⁻¹)	1.31 ± 0.34	$0.28^{+0.29}_{-0.19}$	$1.80^{+0.41}_{-0.40}$
R_b/R_s	0.01471 ± 0.00024	$0.01445^{+0.00031}_{-0.00030}$	$0.01495^{+0.00036}_{-0.00037}$
a/R_s	$7.78^{+0.15}_{-0.16}$	$10.93^{+0.21}_{-0.22}$	$16.65^{+0.33}_{-0.34}$
W (h)	$2.707^{+0.048}_{-0.046}$	$2.844^{+0.053}_{-0.048}$	$3.43^{+0.10}_{-0.11}$
a (AU)	$0.04738^{+0.00076}_{-0.00083}$	$0.0666^{+0.0011}_{-0.0012}$	$3.4248^{+0.039}_{-0.043}$
i_p (degrees)	$85.11^{+0.25}_{-0.26}$	$86.06^{+0.15}_{-0.16}$	87.36 ± 0.10
T_{eq} (K)	1573 ± 20	1327 ± 17	1075^{+14}_{-13}
R_p (R_{\oplus})	2.104 ± 0.043	$2.067^{+0.052}_{-0.050}$	2.137 ± 0.058
M_p (M_{\oplus})	3.34 ± 0.88	$0.84^{+0.88}_{-0.59}$	6.7 ± 1.5
ρ_p (g cm ⁻³)	$1.99^{+0.55}_{-0.54}$	$0.53^{+0.56}_{-0.36}$	$3.79^{+0.96}_{-0.87}$

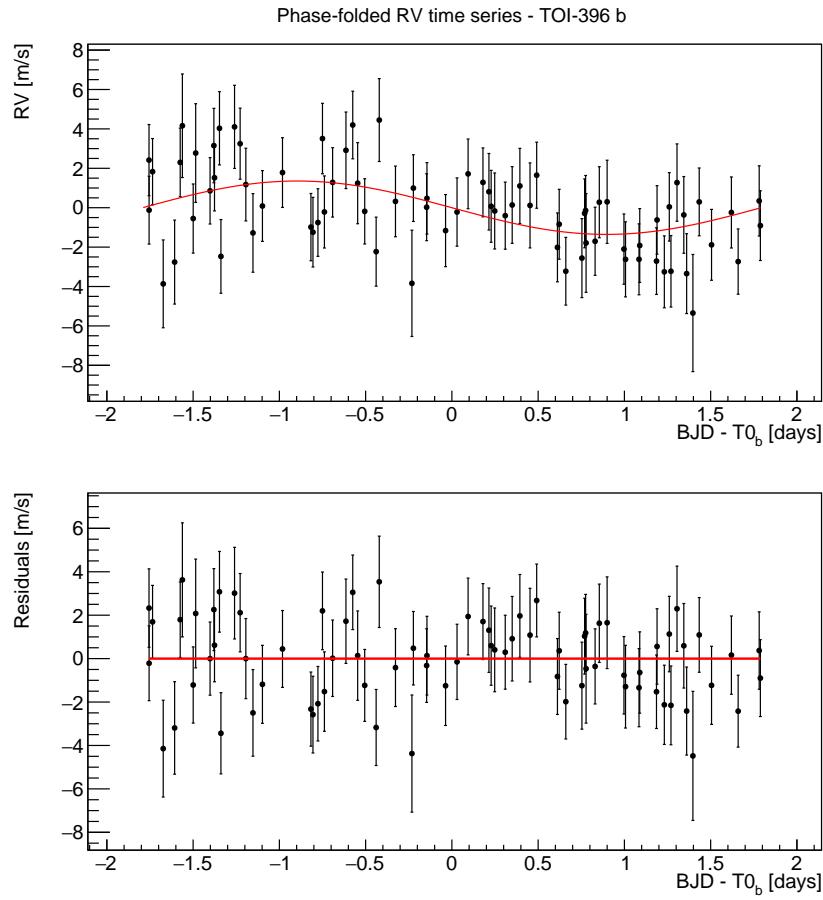


Figure 5.6: *Top*: phase-folded RV signal produced by TOI-396 b with the best-fit Keplerian model superimposed in red. *Bottom*: residuals, that is the difference between the detrended RV signal and the Keplerian model. The error bars also contain the jitter contribution.

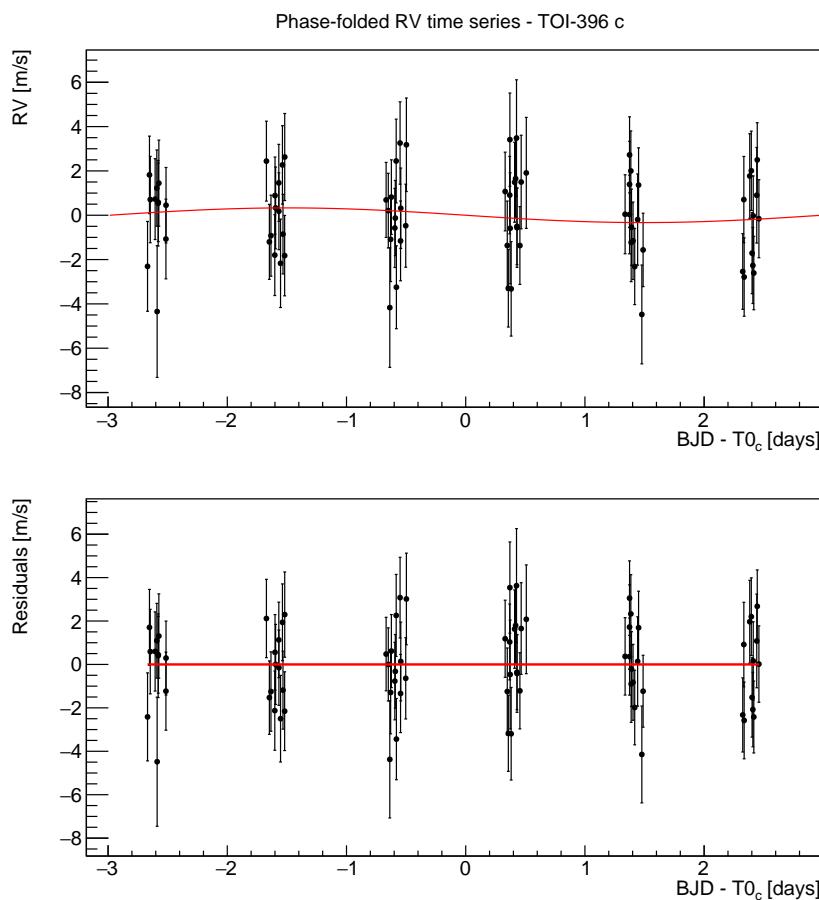


Figure 5.7: Same as Fig. 5.6, but for TOI-396 c.

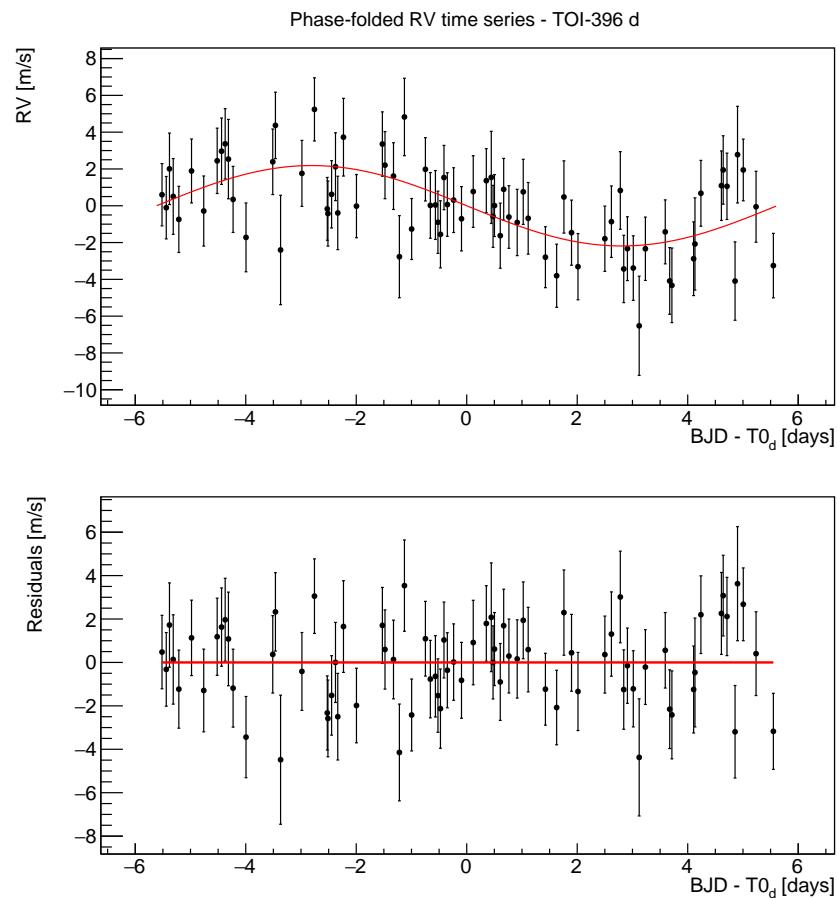


Figure 5.8: Same as Fig. 5.6, but for TOI-396 d.

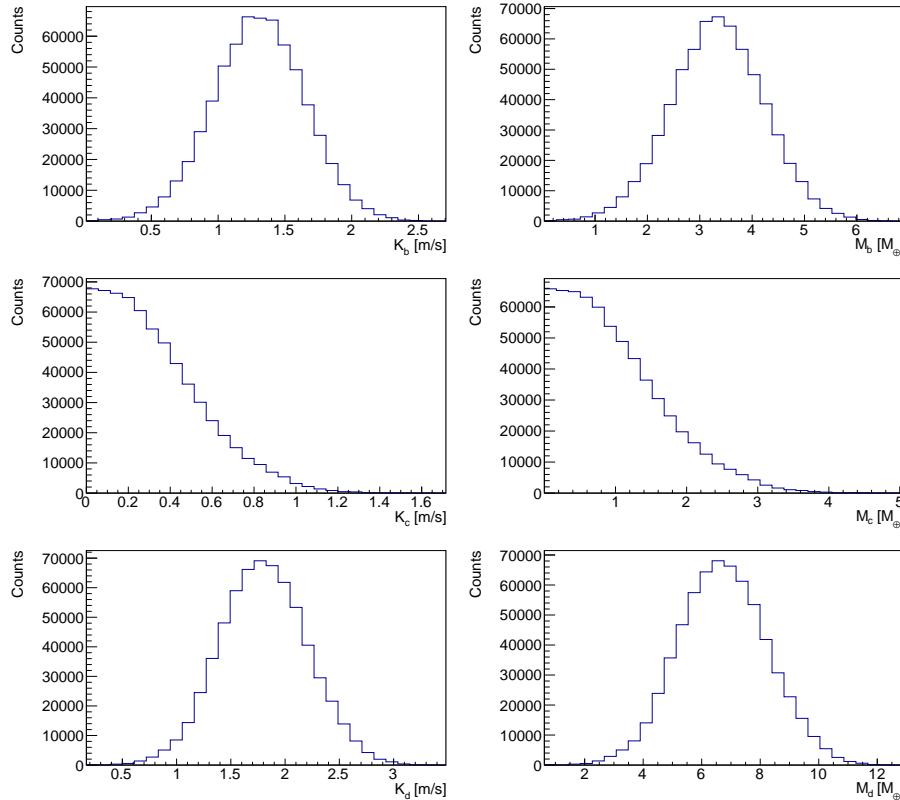


Figure 5.9: Posterior distributions of the RV semi-amplitudes (*first column*) and of the planetary masses (*second column*) for TOI-396 b (*first row*), TOI-396 c (*second row*), and TOI-396 d (*third row*).

$M_d = 6.7 \pm 1.5 M_\oplus$, respectively. We can conclude that the SN-fit-based extraction of RV data and the break-point detrending strategy lead to firmly detect the masses of TOI-396 b and d (at the 3.8σ and 4.5σ levels, respectively). Instead, as previously mentioned, the Keplerian signal of TOI-396 c is not detected. Indeed, by looking at the posterior distributions of the RV semi-amplitudes and of the masses (first and second column of Fig. 5.9, respectively) TOI-396 c posterior distributions are asymmetric and consistent with zero.

5.6.2 Photometry

On the photometric side, Figures 5.10, 5.11 and 5.12 display the phase-folded and detrended light curves with the transits of TOI-396 b, c and d, respectively, as derived from the MCMCI run. I obtained $R_b = 2.104 \pm 0.043 R_\oplus$, $R_c = 2.067^{+0.052}_{-0.050} R_\oplus$, and $R_d = 2.137 \pm 0.058 R_\oplus$; see Fig. 5.13 that shows the posterior distributions of the transit depth (first column) and of the radius (sec-

ond column) for each planet. The consequent radius uncertainty is 2.0%, 2.5%, and 2.7% for planet b, c, and d, respectively, which marks a precision improvement of a factor ~ 1.5 with respect to the results from Vanderburg et al. (2019). In fact, Vanderburg et al. (2019) obtained the following estimates of the planet radii: $R_b^{\text{lit}} = 2.085_{-0.064}^{+0.068} R_{\oplus}$, $R_c^{\text{lit}} = 1.939 \pm 0.069 R_{\oplus}$, and $R_d^{\text{lit}} = 2.164_{-0.083}^{+0.086} R_{\oplus}$ for TOI-396 b, c and d respectively. Although less precise, I note that R_b^{lit} and R_d^{lit} are consistent well within 1σ with the estimates in this work, while R_c^{lit} is consistent at the 1.5σ level.

The combination of the radius estimates with the mass values presented in Sec. 5.6.1 yields to the following mean planetary densities: $\rho_b = 1.99_{-0.54}^{+0.55} \text{ g cm}^{-3}$, $\rho_c = 0.53_{-0.36}^{+0.56} \text{ g cm}^{-3}$, and $\rho_d = 3.79_{-0.87}^{+0.96} \text{ g cm}^{-3}$. Given the RV non-detection of planet c, the consequent ρ_c estimate is unreliable as also discussed in Sec. 5.5. The 3σ upper limit I derived for the mean planet density of TOI-396 c is $\rho_c^{\text{up}} = 2.54 \text{ g cm}^{-3}$.

5.6.3 TOI-396 in the exoplanetary context

To put the TOI-396 system into context, I built a mass-radius (MR) diagram to compare TOI-396 b, c and d with other exoplanetary systems studied in the literature. Specifically, I collected the parameters of those systems³ that consist of at least three planets, as it is the case of TOI-396. Since some of the planetary systems were characterised by different research groups, I arbitrarily chose the most recent reference for each planet. I then kept only those planets whose precision on mass and radius was better than 45% and 15%, respectively.

The resulting MR diagram is shown in Fig. 5.14, where the markers are colour-coded with respect to the planet equilibrium temperature when available. TOI-396 b and d are located in a crowded region. On the other hand, TOI-396 c is almost isolated in a region characterised by an extrely low mean planetary density, which derives from the unreliable median mass value, as previously discussed.

Starting from the same sample of exoplanets, I also plotted a period-density ($P-\rho$) diagram (Fig. 5.15), where the planets belonging to the same system are connected by straight lines to display possible ρ trends as a function of P . I note that the densest planet is the outermost one (i.e. TOI-396 d), which is quite atypical within the exoplanet population (e.g. Ciardi et al., 2013, Weiss et al., 2018, Mishra et al., 2023).

5.7 Hypothesis 2: Stellar activity covers the RV signal of TOI-396 c

The discussion in both Sec. 5.5.1 and 5.6 yields to the conclusion that TOI-396 c indeed orbits TOI-396 in ~ 6 days. However, the reason for its RV non-detection

³Source: NASA Exoplanet Archive website; <https://exoplanetarchive.ipac.caltech.edu/>

5.7. HYPOTHESIS 2: STELLAR ACTIVITY COVERS THE RV SIGNAL OF TOI-396 C73

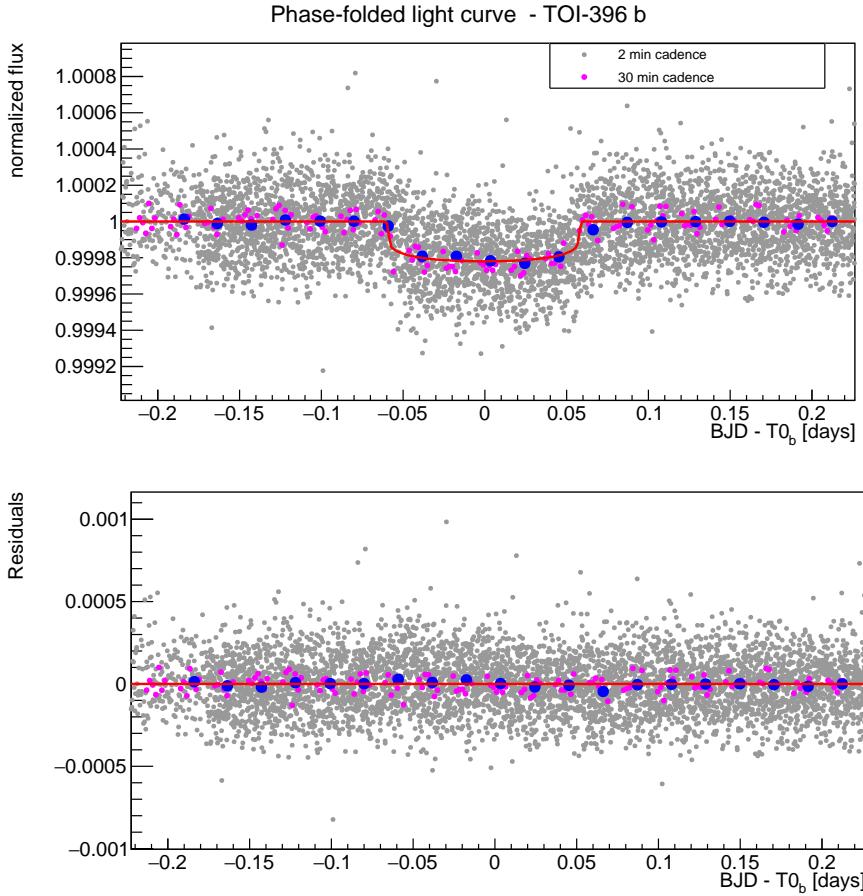


Figure 5.10: *Top:* phase-folded and detrended LC showing the transit of TOI-396 b with the best-fit transit model superimposed (in red). Grey dots represent the observations from Sectors 4, 30 and 31 (2 min cadence), while pink dots the observations from Sector 3 (30 min cadence). The bigger blue dots are the result of a 30 min binning of the entire set of data points. *Bottom:* LC residuals.

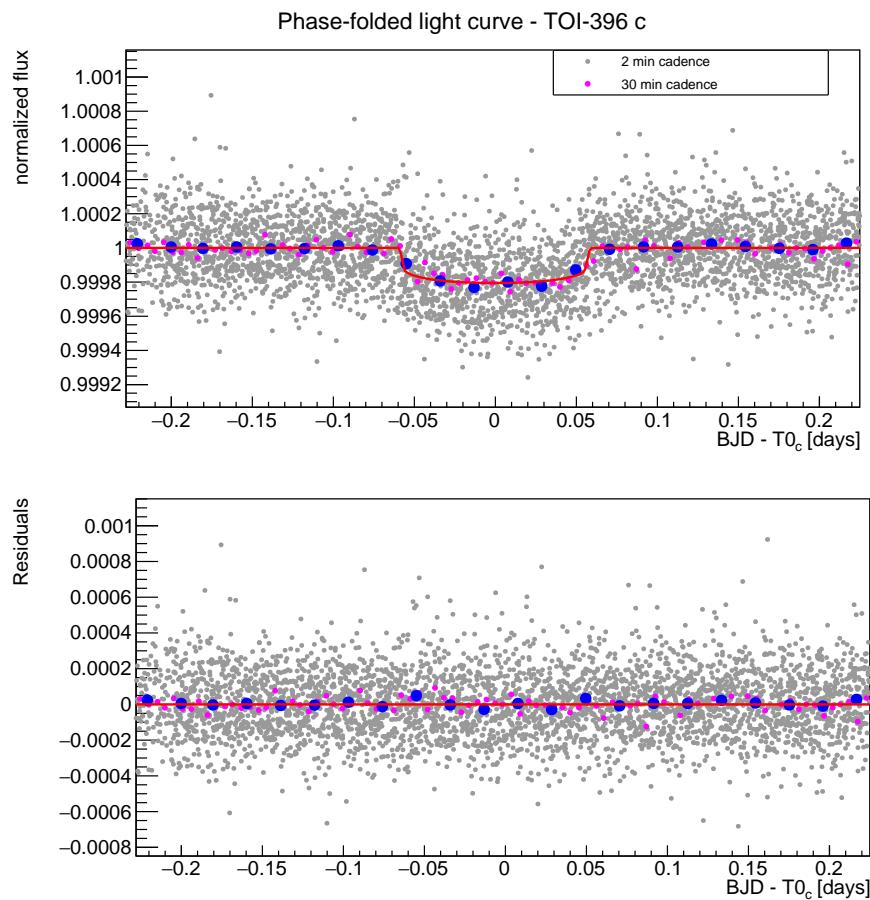


Figure 5.11: Same as Fig. 5.10, but for TOI-396 c.

5.7. HYPOTHESIS 2: STELLAR ACTIVITY COVERS THE RV SIGNAL OF TOI-396 C75

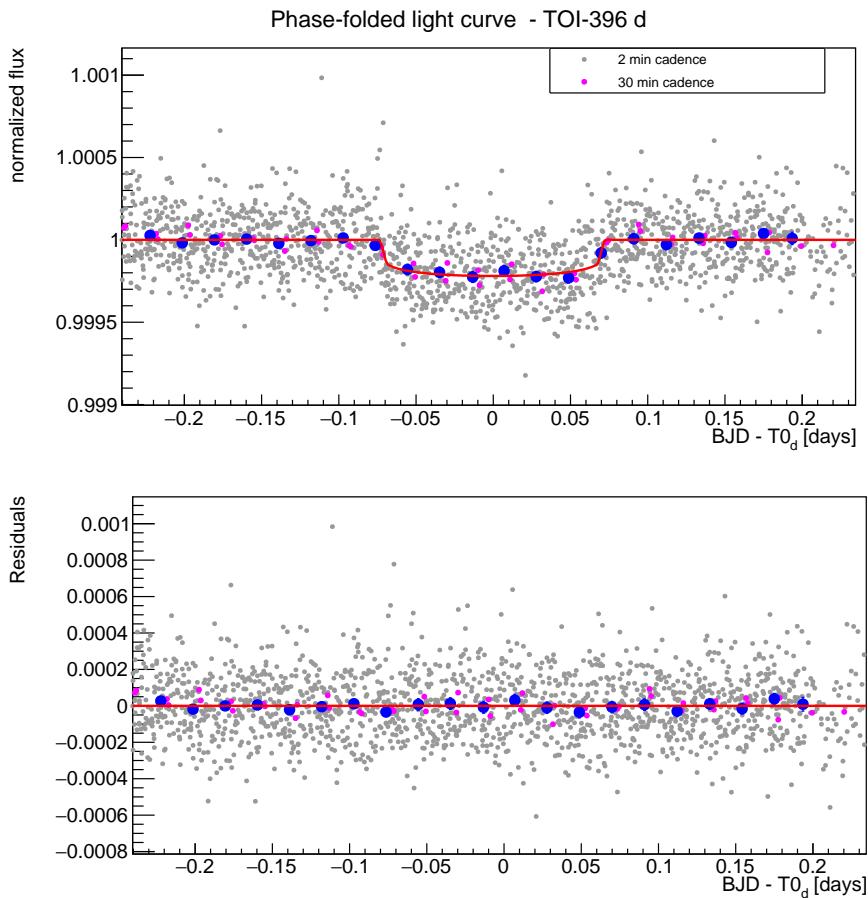


Figure 5.12: Same as Fig. 5.10, but for TOI-396 d.

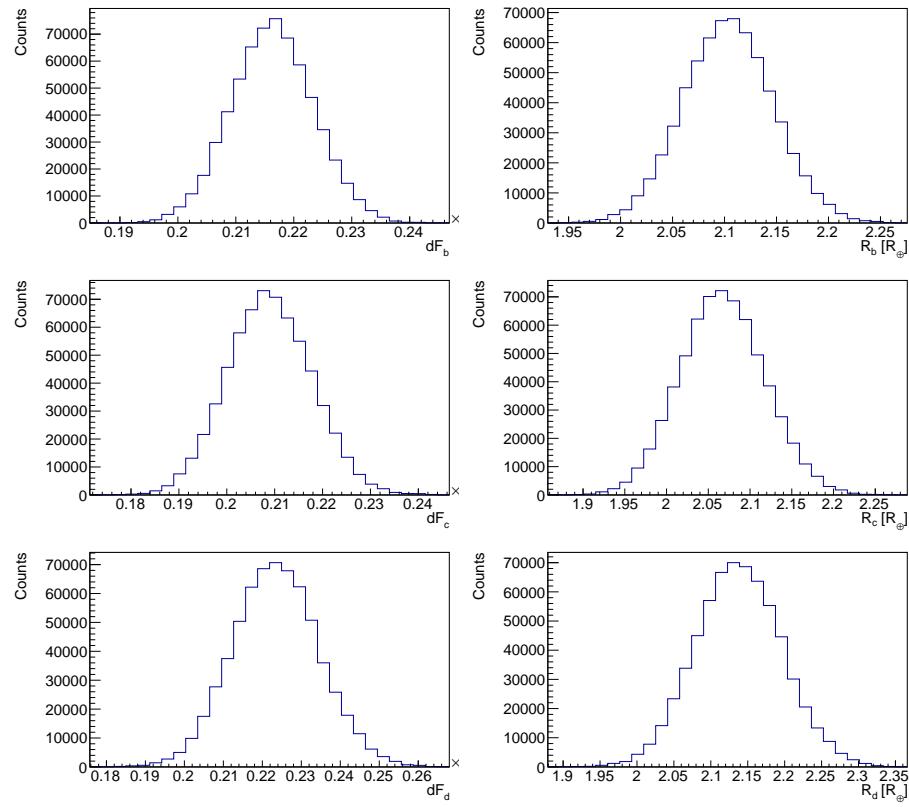


Figure 5.13: Posterior distributions of the transit depths (*first column*) and of the planetary radii (*second column*) for TOI-396 b (*first row*), TOI-396 c (*second row*), and TOI-396 d (*third row*).

5.7. HYPOTHESIS 2: STELLAR ACTIVITY COVERS THE RV SIGNAL OF TOI-396 C77

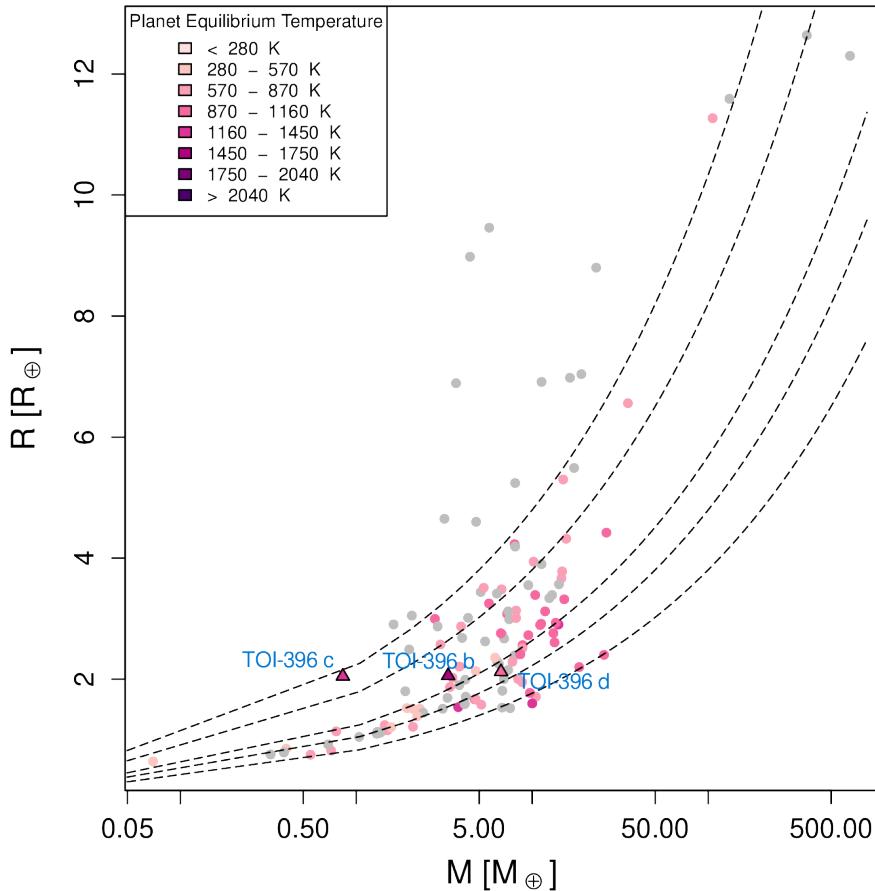


Figure 5.14: Mass-Radius diagram of those exoplanets within multi-planet systems which contains at least three planets. The mass and the radius of the three inner exoplanets is known with a precision better than 45% and 15%, respectively. The triangular markers represent the TOI-396 planets. All the markers are colour-coded with respect to the planet equilibrium temperature (T_{eq}) when available, otherwise they are grey. The dashed lines are loci of points with the same mean density, which is equal to 0.5, 1.0, 3.0, 5.0, and 10.0 g cm^{-3} going from top to bottom.

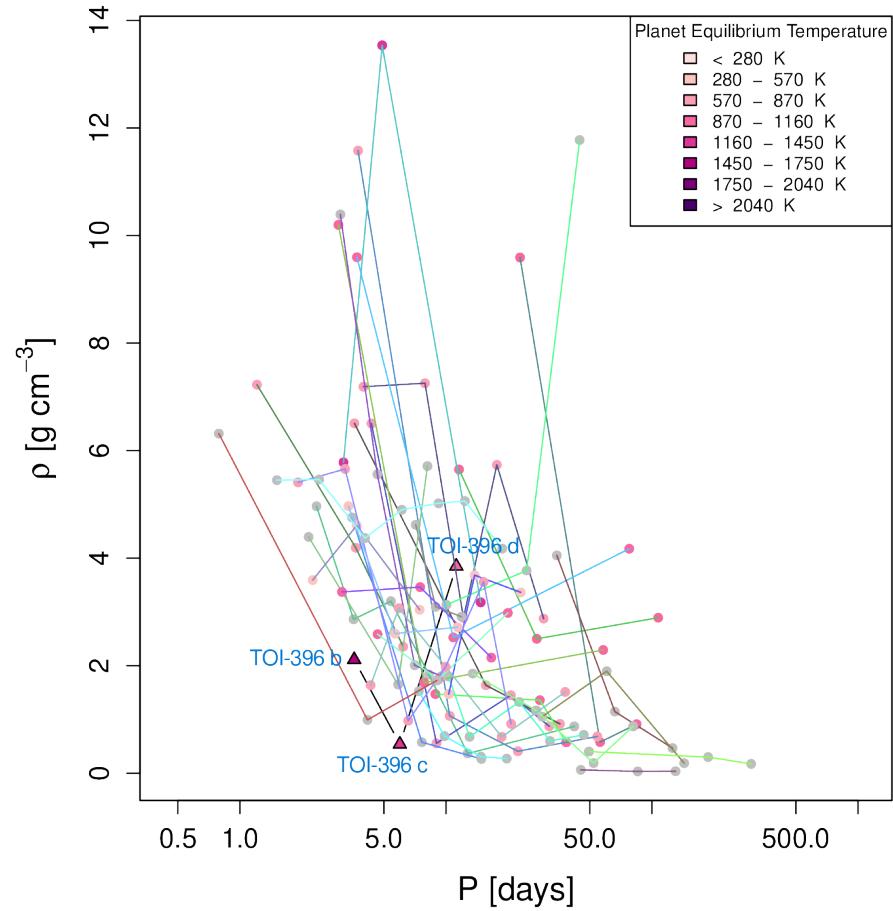


Figure 5.15: Period-Density diagram of those exoplanets following the same selection criteria specified for Fig. 5.14, with straight lines connecting planets belonging to the same system. All the markers are colour-coded with respect to the planet equilibrium temperature (T_{eq}) when available, otherwise they are grey, while the line colours are randomly selected, being not related to any physical parameter. Triangular markers display the TOI-396 planets as indicated by the text labels.

5.7. HYPOTHESIS 2: STELLAR ACTIVITY COVERS THE RV SIGNAL OF TOI-396 C79

is still not clear.

As broadly discussed in Sec. 2.4, stellar activity generates RV variations which may mimic planetary signals if they are stable enough (e.g. Suárez Martínez et al., 2017). A possible explanation is that stellar activity might have destructively interfered with the Keplerian signal of TOI-396 c, hence cancelling out the RV signal of the planet. This may happen if the star has a rotation period comparable to the orbital period of its planet (e.g. Vanderburg et al., 2016). Indeed, in those cases, disentangling the planet signal from stellar noise and retrieving the RV Keplerian signal is really challenging (e.g. Dragomir et al., 2012, Kossakowski et al., 2022).

To explore this possibility, I computed the GLS periodograms of those activity indicators both outputted by the *HARPS* pipeline (Fig. 5.16) and extracted when performing the skew-normal fit onto the CCFs (Fig. 5.17). In fact, those periodograms are expected to show significant peaks at the rotation period of the star.

Possibly after removing the harmonic responsible of the highest peak in the periodograms, all the three SN-fit-related activity indicators (namely the contrast, the FWHM_{SN} , and γ) peak in the range [7.0; 7.5] days (power of 0.20% on average). The agreement in the peak locations of these three indicators may suggest an estimate for stellar rotation period P_{rot} such that $7 \lesssim P_{\text{rot}} [\text{d}] \lesssim 7.5$. Instead, the *HARPS*-derived activity indicators lead to not very significant (power $\lesssim 15$) and dishomogeneous GLS peaks, which indicates their lower performance in tracing the stellar activity.

To check whether we find a periodic signal compatible with ~ 7 days also in the photometric data, I computed the GLS periodograms of the *TESS* LCs (Fig. 5.18) after removing all those temporal windows containing transit events. In detail, from each Sector I built a time series made of the raw stellar flux (FLUX_RAW), that is the stellar flux as measured by the instrument and not corrected via the PDCSAP *TESS* pipeline⁴. In fact, I want to shed light on possible periodicities of the stellar activity, which might have been cancelled out by photometric corrections applied at the PDCSAP level. Possibly after removing the harmonic responsible of the highest periodogram peak, the GLS periodograms computed from Sector 3, 30, and 31 show a significant peak at ~ 6.7 , 7.1, and 6.6 days, respectively. Apparently, the GLS periodogram from Sector 4 (second row in Fig. 5.18) does not show a peak consistent with those from the other sectors, but the peak at ~ 6 days (second row, first column in Fig. 5.18) is actually an harmonic of the highest peak in the plot, as it disappears in the plot aside where the harmonic responsible of the highest peak has been removed.

Summing up, both the RV-related activity indicators and the photometric flux from *TESS* consistently show periodic signals between 6 and 7.5 days, which may be ascribed to the stellar rotation period. Therefore, the orbital period of TOI-396 c ($P_c \sim 6$ d) is likely close to the stellar rotation period,

⁴I recall that for building the `phot` files from Sector 4, 30, and 31 I used the PDCSAP flux, instead.

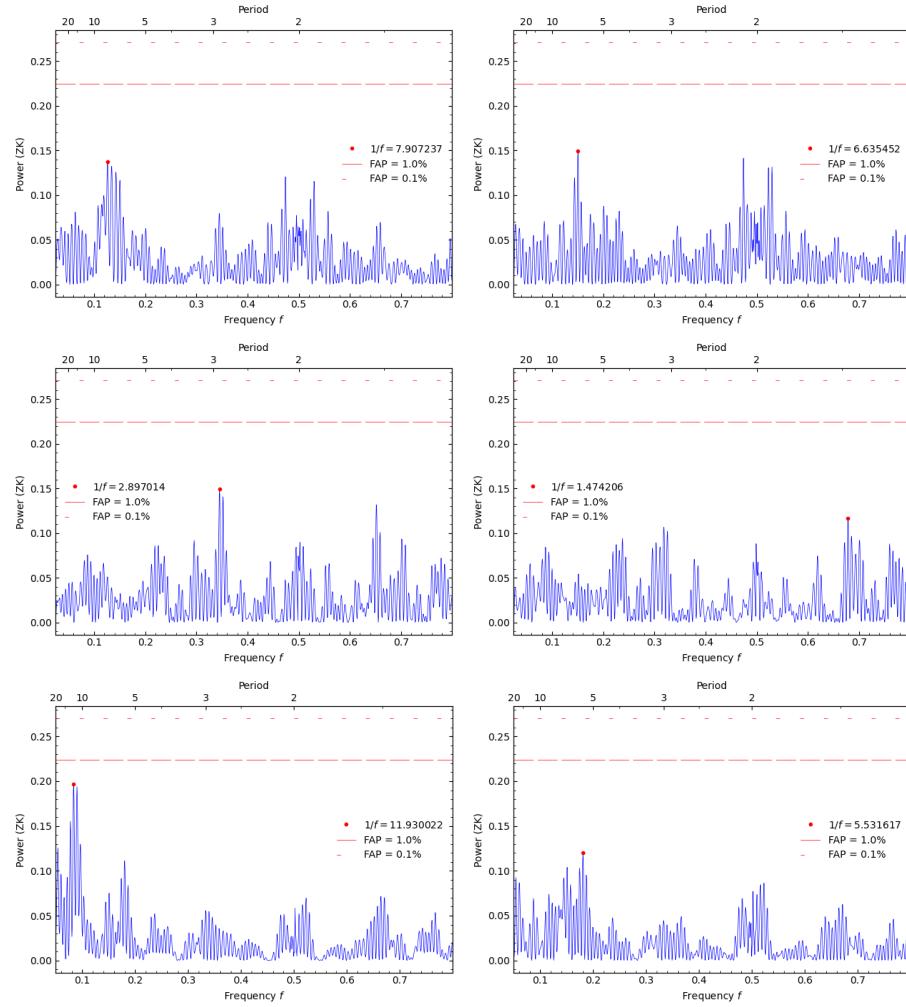


Figure 5.16: *Left column:* GLS periodograms computed from the activity indicators as derived by the *HARPS* pipeline, namely the contrast (*first row*), the FWHM (*second row*), and $\log R'_{\text{HK}}$ (*third row*). In the *right column* the periodograms have been recomputed after subtracting from the time series the sinusoidal wave responsible of the highest peak seen in the *left column*. The peaks are not significant and their location differs from one row to the other, therefore the *HARPS*-derived indicators are not very effective in tracing the stellar activity. See text for further details.

5.7. HYPOTHESIS 2: STELLAR ACTIVITY COVERS THE RV SIGNAL OF TOI-396 C81

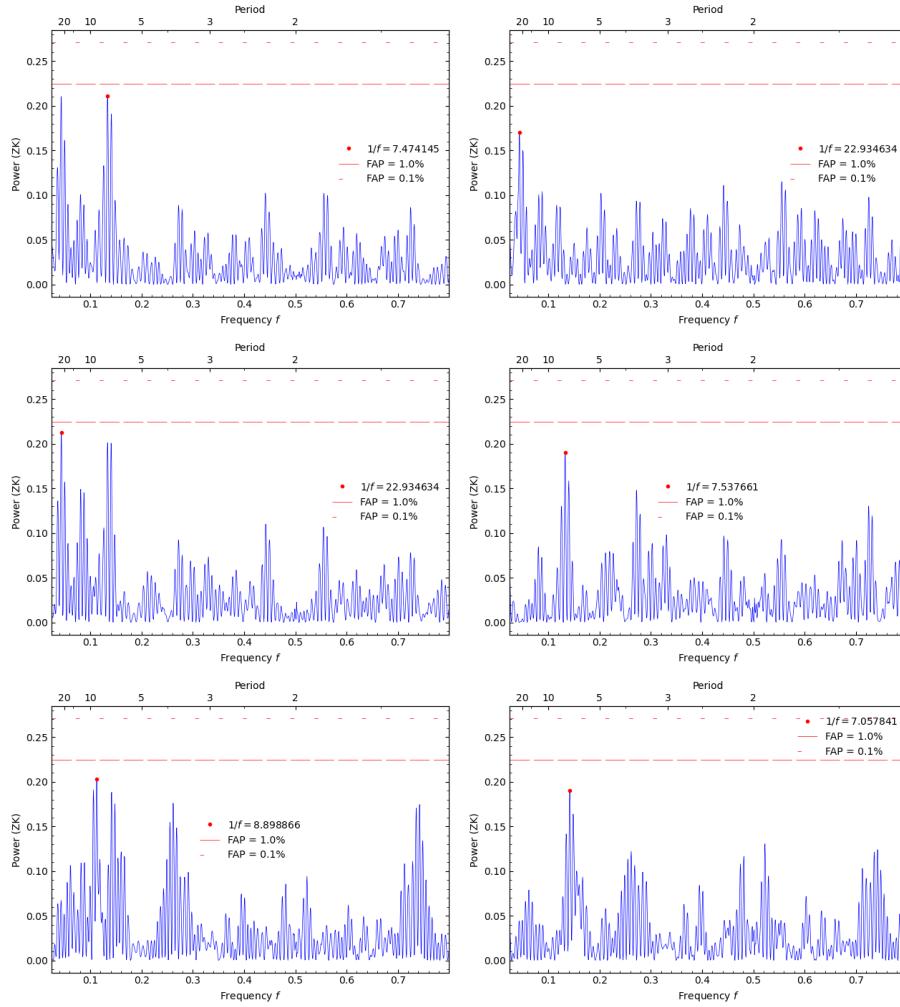


Figure 5.17: *Left column:* GLS periodograms computed from the activity indicators as derived from the SN-fit onto the CCF, namely the CCF-contrast (*first row*), the FWHMSN (*second row*), and the CCF-skewness γ (*third row*). In the *right column* the periodograms have been recomputed after subtracting from the time series the sinusoidal wave responsible of the highest peak seen in the first column. The periodograms from the three activity indicators have consistent peaks in the range [7.0; 7.5] d, which may give hints about the stellar rotation period. See text for further details.

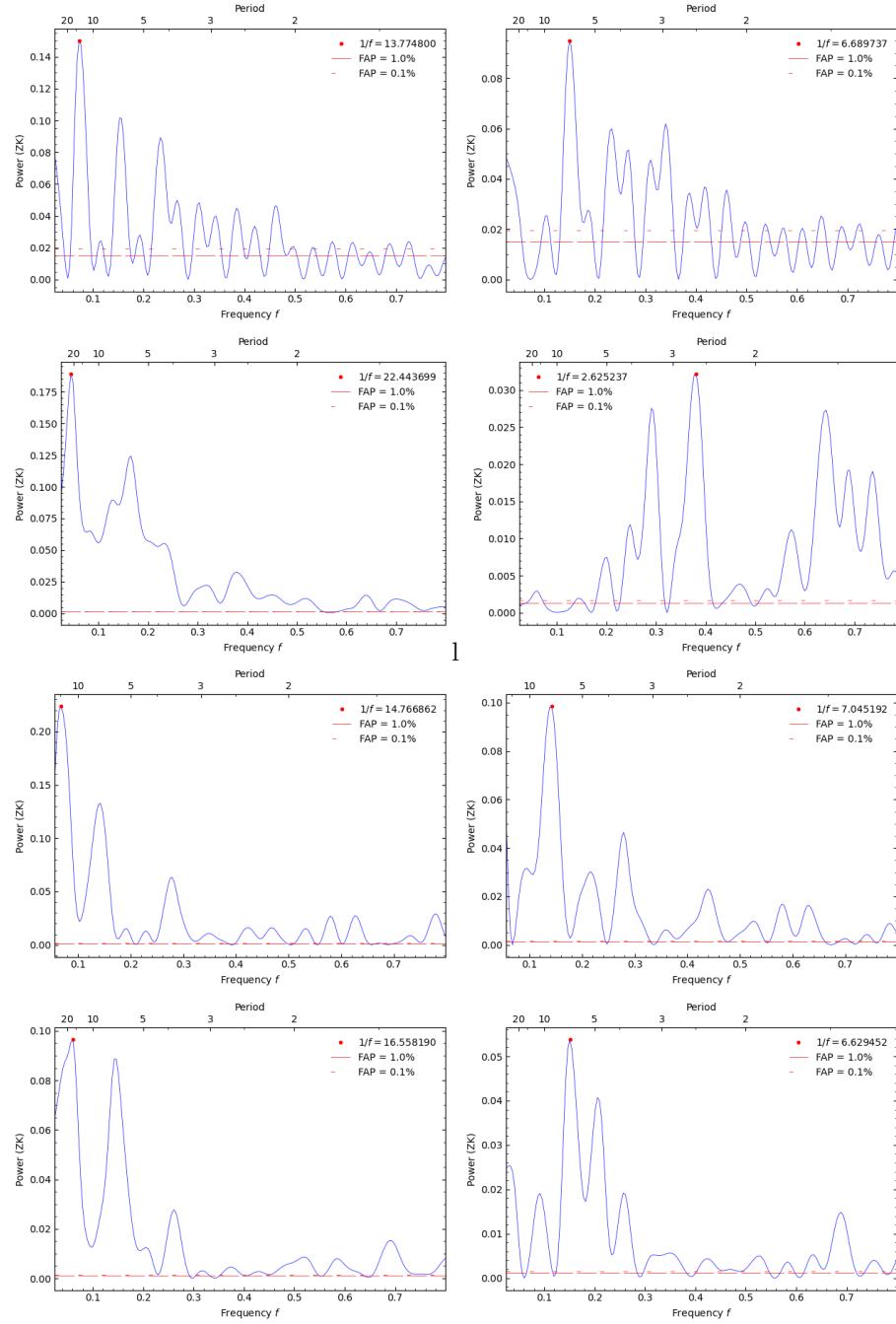


Figure 5.18: *Left column:* GLS periodograms as computed from the *TESS* raw photometric flux of Sector 3 (*first row*), Sector 4 (*second row*), Sector 30 (*third row*), and Sector 31 (*fourth row*). In the *Right column* periodograms have been recomputed after subtracting from the time series the harmonic responsible of the highest peak seen in the *left column*.

5.7. HYPOTHESIS 2: STELLAR ACTIVITY COVERS THE RV SIGNAL OF TOI-396 C83

Table 5.7: RV semi-amplitude K_{out} as retrieved from an MCMC analysis of that RV time series obtained by adding a Keplerian signal with period P and semi-amplitude K_{in} to the original RV time series stored in the `rv0001.txt` input file.

P (days)	K_{in} (m/s)	K_{out} (m/s)
$P_b = 3.5852921$	$K_b = 1.31$	$2.65 \pm 0.33 (\simeq 2K_b)$
$P_d = 11.230513$	$K_d = 1.79$	$3.53 \pm 0.42 (\simeq 2K_d)$

which may explain the non detection of planet c within the RV time series, as detailed before. Going along this line, I averaged out the locations $P_{\text{activity}}^{\max}$ of the consistent and most significant peaks in the GLS periodograms, hence obtaining the following estimate for the stellar rotation period:

$$P_{\text{rot}} = 6.8 \pm 1.1 \text{ days}, \quad (5.4)$$

where the uncertainty has been computed as the half maximum dispersion between the GLS estimates.

By combining P_{rot} and the stellar radius R_s , we get an estimate of the rotational speed of the star:

$$v = \frac{2\pi R_s}{P_{\text{rot}}} = 9.8 \pm 1.6 \text{ km s}^{-1}, \quad (5.5)$$

where the uncertainty on v comes from propagating the errors on both R_s and P_{rot} . Our v estimate is consistent with its lower limit given by the projected rotational velocity of the star ($v \sin i = 8.30 \pm 0.50 \text{ km s}^{-1}$; Vanderburg et al., 2019). Thus, we can estimate the inclination of the stellar spin axis, that is

$$i = 57.9 \pm 2.6 \text{ deg} \quad (5.6)$$

5.7.1 Test1: adding artificial RV signals with $P = P_c$

To test how a planetary RV signal having the same period as P_{rot} may be cancelled out by stellar activity, I decided to add to the RV time series Keplerian signals of the following form

$$RV_{\text{art}} = K_{\text{art}} \sin \left[\frac{2\pi}{P} (T_0 - t) \right] \quad (5.7)$$

I first added a signal with $K_{\text{art}} = K_b$, $P = P_b$, and $T_0 = T_{0,b}$ and after running the MCMCI code I retrieved a planet signal, whose semi-amplitude was the double of K_b , as expected. Similarly, I repeated this sanity check for planet d, and again the output semi-amplitude was the double of K_d (see Table 5.7).

As the sanity check involving planets b and d performed as expected, I then focused on the core test involving planet c. In detail, I generated four different

Table 5.8: Same as table 5.7, but this time the different artificial Keplerian signals with $K = K_{\text{in}}$ have always period $P = P_c$. Column 2 and 3 translate the injected signals into the corresponding physical parameters of the planet, while column 5 quantifies the detection level of K_{out} .

K_{in} (m/s)	ρ_p (g cm $^{-3}$)	M_p (M_{\oplus})	K_{out} (m/s)	Detection level
0.51	~ 1.0	~ 1.6	0.68 ± 0.36	$\sim 2\sigma$
0.77	~ 1.5	~ 2.3	0.93 ± 0.42	$\sim 2\sigma$
1.0	~ 1.9	~ 3.0	1.15 ± 0.38	$\sim 3\sigma$
$K_d = 1.79$	~ 3.5	~ 5.4	1.94 ± 0.38	$\sim 3\sigma$

RV time series by separately adding to the original one Keplerian signals following Eq. (5.7) with $P = P_c$, $T_0 = T_{0,c}$, and $K_{\text{art}} = K_{\text{in}}$, where K_{in} are four different amplitude values listed in the first column of table 5.8. The results of the four different MCMCI run are reported in table 5.8, which specifically highlight the retrieved planet semi-amplitude K_{out} . I note that the resulting $K_{\text{out}} \approx K_c + K_{\text{in}}$, where $K_c = 0.27^{+0.29}_{-0.19}$ is the RV semi-amplitude of planet c as derived from the analysis on the original RV time series. As I essentially retrieved what I inputted in the RV time series, I may conclude that the destructive interference between the RV signals induced by the star and by planet c has already occurred and any further RV signal added to the RV time series is actually detected.

5.7.2 Test2: adding artificial RV signals with $P = P_{\text{rot}}$

How robust is then our P_{rot} estimate? To validate it, I repeated the same kind of analysis outlined in Sec. 5.7.1, but this time the artificial Keplerian signals I injected into the original RV time series have a period $P = P_{\text{rot}}$. The K_{out} values outputted by the MCMC analyses in correspondence of the different K_{in} values are reported in table 5.9. I note that K_{out} are systematically and significantly smaller than K_{in} ; for example, by injecting a Keplerian signal with $K_{\text{in}} = K_d$ and $P = P_{\text{rot}}$, the MCMCI analysis is able to retrieve a planet signal whose amplitude is half than expected. Things get better when K_{in} increases, however we still underestimate K_{out} . These tests prove that it is hard to retrieve planetary signals with $P = P_{\text{rot}}$, which seems to confirm that P_{rot} is a reliable estimate of the stellar rotation period. In fact, stellar activity is responsible of generating spurious RV signals whose harmonics also contain the stellar rotation period. As a consequence, I confirm that it is hard to reliably detect planets with orbital periods comparable to the stellar rotation period via the RV technique and I quantified the magnitude of this effect in table 5.9.

5.7. HYPOTHESIS 2: STELLAR ACTIVITY COVERS THE RV SIGNAL OF TOI-396 C85

Table 5.9: Same as table 5.8, but this time the Keplerian signals with $K = K_{\text{in}}$ have all period $P = P_{\text{rot}}$. In addition, column 5 gives the relative difference (in percentage) between the outputted RV semi-amplitude (K_{out}) and the expected one (K_{in}), while column 6 expresses the $K_{\text{out}} - K_{\text{in}}$ difference in terms of the 1σ uncertainty of K_{out} .

K_{in} (m/s)	ρ_p (g cm $^{-3}$)	M_p (M_{\oplus})	K_{out} (m/s)	$\frac{\Delta K}{K}$	$\frac{\Delta K}{\sigma}$
1.0	~ 2.0	~ 3.1	$0.29 \pm 0.28(\sim 1\sigma)$	-71%	-2.5
$K_d = 1.79$	~ 3.6	~ 5.6	$0.80 \pm 0.39(\sim 2\sigma)$	-55%	-2.5
3.0	~ 6.1	~ 9.4	$1.90 \pm 0.40(\sim 5\sigma)$	-37%	-2.8
10.0	~ 20.2	~ 31.4	$9.02 \pm 0.36(\sim 25\sigma)$	-10%	-2.7

Chapter 6

Conclusions

The work of my Master thesis focuses on the characterization of the planetary system transiting the naked-eye F-type star TOI-396. Using the transit method, this system was discovered by Vanderburg et al. (2019), who unveiled the presence of three small planets transiting the star. In the course of my thesis work, I studied the system using both the radial velocity technique, and the transit method, through a joint analysis that uses a Markov Chain Monte Carlo code called MCMCI developed by Bonfanti & Gillon (2020). The RV analysis uses 78 *HARPS* spectra collected in 2019 as part of a Doppler follow-up program of small transiting planets carried out by the Physics Department of the University of Turin. The RV time series, with the relevant ancillary parameters, were extracted by performing a skew-Normal fit to the *HARPS* cross-correlation functions (Simola et al., 2019), while the RV detrending used the breakpoint technique (Simola et al., 2022) to remove the Doppler signals induced by stellar activity. The photometric data came from four different *TESS* sectors, namely Sector 3 and 4 – observed in 2018 –, and Sector 30 and 31 – observed in 2020.

The most relevant results of my thesis work are summarised in table 5.6. Specifically, I measured – with high precision – the radius and the mass of the three small planets transiting TOI-396, with the only exception for the mass of TOI-396 c, for which only an upper limit was obtained. The null detection of a Keplerian signal at $P_c \approx 6$ d is likely due to the orbital period of TOI-396 c being close to the rotation period of the star, which makes it quite difficult to detect exoplanets via the RV technique.

By combining my reliable mass and radius estimates for both TOI-396 b and d, I derived the planetary bulk density of the inner and the outer planets. I found that TOI-396 b has a mean density of $\rho_b = 1.99^{+0.55}_{-0.54}$ g cm⁻³, whereas TOI-396 d of $\rho_d = 3.79^{+0.96}_{-0.87}$ g cm⁻³. This suggests that TOI-396 b is an icy planet, while TOI-396 d is richer in silicates. I would like to stress that the masses of these three planets were previously unknown, and that the results presented in my thesis work have been possible thanks to the SN-fit and the

breakpoint methods, which are very promising techniques for the analysis of RV time series affected by stellar activity.

I also found that TOI-396 b and c exhibit transit timing variations (TTVs), which is expected given that they are in a commensurable 5:3 mean motion resonance. A TTV-based derivation of the mass of TOI-396 c could be a promising approach to finally determine the mean density of planet c, and infer its composition and internal structure. In this respect, additional transit observations are required to better constrain the TTV amplitude.

The TOI-396 system is also quite unusual, as its densest transiting planet is the outermost one (i.e. TOI-396 d). To further unveil the system architecture, a thorough characterisation of the three exoplanets would require to study the presence of an atmosphere, which can be investigated with, e.g., the James Webb Space Telescope (JWST).

As a final remark, the analysis carried out during my Master thesis work will be presented in a paper to be submitted to a scientific journal, and that will include myself, Irene Amateis, as co-author.

Bibliography

- Adams, J. C. 1846, MNRAS, 7, 149
- Agol, E. & Deck, K. 2016, ApJ, 818, 177
- Agol, E. & Fabrycky, D. C. 2018, in Handbook of Exoplanets, ed. H. J. Deeg & J. A. Belmonte (Cambridge University Press), 7
- Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567
- Anderson, D. R., Collier Cameron, A., Hellier, C., et al. 2011, ApJ, 726, L19
- Arellano-Valle, R. B. & Azzalini, A. 2008, Journal of Multivariate Analysis, 99, 1362, special Issue: Multivariate Distributions, Inference and Applications in Memory of Norman L. Johnson
- Azzalini, A. 1985, Scandinavian Journal of Statistics, 12, 171
- Bai, J. & Perron, P. 2003, Journal of Applied Econometrics, 18, 1
- Baluev, R. V. 2018, Astronomy and Computing, 25, 221
- Barragán, O., Gandolfi, D., & Antoniciello, G. 2019, MNRAS, 482, 1017
- Bellman, R. & Roth, R. 1969, Journal of the American Statistical Association, 64, 1079
- Betancourt, M. 2017, arXiv e-prints, arXiv:1701.02434
- Boisse, I., Bouchy, F., Hébrard, G., et al. 2011, A&A, 528, A4
- Bonfanti, A., Delrez, L., Hooton, M. J., et al. 2021, A&A, 646, A157
- Bonfanti, A. & Gillon, M. 2020, A&A, 635, A6
- Bonfanti, A., Ortolani, S., & Nascimbeni, V. 2016, A&A, 585, A5
- Bonfanti, A., Ortolani, S., Piotto, G., & Nascimbeni, V. 2015, A&A, 575, A18
- Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
- Casella, G. & George, E. I. 1992, The American Statistician, 46, 167

- Chatzistergos, T., Ermolli, I., Krivova, N. A., et al. 2022, *A&A*, 667, A167
- Christensen-Dalsgaard, J. 2004, *Sol. Phys.*, 220, 137
- Ciardi, D. R., Fabrycky, D. C., Ford, E. B., et al. 2013, *ApJ*, 763, 41
- Claret, A. & Bloemen, S. 2011, *A&A*, 529, A75
- Csizmadia, S., Pasternacki, T., Dreyer, C., et al. 2013, *A&A*, 549, A9
- Dawson, R. I. & Johnson, J. A. 2018, *ARA&A*, 56, 175
- Deck, K. M. & Agol, E. 2016, *ApJ*, 821, 96
- Dragomir, D., Kane, S. R., Henry, G. W., et al. 2012, *ApJ*, 754, 37
- Dumusque, X. 2016, *A&A*, 593, A5
- Duncan, D. K., Vaughan, A. H., Wilson, O. C., et al. 1991, *ApJS*, 76, 383
- Eastman, J., Gaudi, B. S., & Agol, E. 2013, *PASP*, 125, 83
- Eastman, J., Siverd, R., & Scott Gaudi, B. 2010, *Publications of the Astronomical Society of the Pacific*, 122, 935
- Eastman, J. D., Rodriguez, J. E., Agol, E., et al. 2019, EXOFASTv2: A public, generalized, publication-quality exoplanet modeling code
- Espinoza, N. & Jordán, A. 2015, *MNRAS*, 450, 1879
- Espinoza, N., Kossakowski, D., & Brahm, R. 2019, *MNRAS*, 490, 2262
- Figueira, P., Santos, N. C., Pepe, F., Lovis, C., & Nardetto, N. 2013, *A&A*, 557, A93
- Fischer, D. A., Anglada-Escude, G., Arriagada, P., et al. 2016, *PASP*, 128, 066001
- Fisher, W. D. 1958, *Journal of the American statistical Association*, 53, 789
- Ford, E. B. 2005, *AJ*, 129, 1706
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *Publications of the Astronomical Society of the Pacific*, 125, 306
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2022, arXiv e-prints, arXiv:2208.00211
- Gaudi, B. S. & Winn, J. N. 2007, *ApJ*, 655, 550
- Gelman, A. & Rubin, D. B. 1992, *Statistical Science*, 7, 457
- Gillon, M., Demory, B. O., Triaud, A. H. M. J., et al. 2009, *A&A*, 506, 359

- Giménez, A. 2006, ApJ, 650, 408
- Gray, R. O., Corbally, C. J., Garrison, R. F., et al. 2006, AJ, 132, 161
- Günther, M. N. & Daylan, T. 2021, ApJS, 254, 13
- Guthery, S. B. 1974, Journal of the American Statistical Association, 69, 945
- Haghighipour, N. 2013, Annual Review of Earth and Planetary Sciences, 41, 469
- Harten, R. H., Grosbol, P., Greisen, E. W., & Wells, D. C. 1988, A&AS, 73, 365
- Hartmann, L., Soderblom, D. R., Noyes, R. W., Burnham, N., & Vaughan, A. H. 1984, ApJ, 276, 254
- Hastings, W. K. 1970, Biometrika, 57, 97
- Hatzes, A. P. 2016, in Astrophysics and Space Science Library, Vol. 428, Methods of Detecting Exoplanets: 1st Advanced School on Exoplanetary Science, ed. V. Bozza, L. Mancini, & A. Sozzetti, 3
- Hatzes, A. P. 2019, The Doppler Method for the Detection of Exoplanets (Institute of Physics Publishing)
- Haywood, R. D. 2015, PhD thesis, Saint Andrews University, UK
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
- Holman, M. J., Winn, J. N., Latham, D. W., et al. 2006, ApJ, 652, 1715
- Kipping, D. M. 2010, Monthly Notices of the Royal Astronomical Society, 408, 1758
- Kossakowski, D., Kürster, M., Henning, T., et al. 2022, A&A, 666, A143
- Le Verrier, U. J. 1877, Annales de l'Observatoire de Paris, 14, A.1
- Lindgren, L., Bastian, U., Biermann, M., et al. 2021, A&A, 649, A4
- Lomb, N. R. 1976, Ap&SS, 39, 447
- Lovis, C. & Fischer, D. 2010, in Exoplanets, ed. S. Seager (Tucson, AZ: University of Arizona Press), 27–53
- Mandel, K. & Agol, E. 2002, ApJ, 580, L171
- Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
- Maxted, P. F. L. 2018, A&A, 616, A39
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger, 114, 20

- Mayor, M. & Queloz, D. 1995, *Nature*, 378, 355
- Meunier, N., Desort, M., & Lagrange, A. M. 2010, *A&A*, 512, A39
- Middelkoop, F. 1982, *A&A*, 107, 31
- Mishra, L., Alibert, Y., Udry, S., & Mordasini, C. 2023, *A&A*, 670, A68
- Morton, T. D. 2012, *ApJ*, 761, 6
- Morton, T. D. 2015, VESPA: False positive probabilities calculator, Astrophysics Source Code Library, record ascl:1503.011
- Murray, C. D. & Correia, A. C. M. 2010, in *Exoplanets*, ed. S. Seager (University of Arizona Press), 15–23
- Nardetto, N., Mourard, D., Kervella, P., et al. 2006, *A&A*, 453, 309
- Nesvorný, D. 2009, *ApJ*, 701, 1116
- Nesvorný, D. & Beaugé, C. 2010, *ApJ*, 709, L44
- Nesvorný, D. & Morbidelli, A. 2008, *ApJ*, 688, 636
- Parmentier, V. & Crossfield, I. J. M. 2018, in *Handbook of Exoplanets*, ed. H. J. Deeg & J. A. Belmonte (Cambridge University Press), 116
- Pepe, F., Molaro, P., Cristiani, S., et al. 2014, *Astronomische Nachrichten*, 335, 8
- Perryman, M. 2018, *The Exoplanet Handbook*, 2nd edn. (Cambridge University Press)
- Pont, F., Zucker, S., & Queloz, D. 2006, *MNRAS*, 373, 231
- Queloz, D., Henry, G. W., Sivan, J. P., et al. 2001, *A&A*, 379, 279
- R Core Team. 2019, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, ed. J. Oschmann, Jacobus M., M. Clampin, G. G. Fazio, & H. A. MacEwen, 914320
- Robertson, P., Endl, M., Henry, G. W., et al. 2015, *The Astrophysical Journal*, 801, 79
- Robinson, R. D., Cram, L. E., & Giampapa, M. S. 1990, *ApJS*, 74, 891
- Robrade, J., Schmitt, J. H. M. M., & Favata, F. 2012, *A&A*, 543, A84

- Scargle, J. D. 1982, ApJ, 263, 835
- Schwarz, G. 1978, The Annals of Statistics, 6, 461
- Scuflaire, R., Montalbán, J., Théado, S., et al. 2008, Astrophysics and Space Science, 316, 149
- Seager, S. & Mallén-Ornelas, G. 2003, ApJ, 585, 1038
- Simola, U., Bonfanti, A., Dumusque, X., et al. 2022, A&A, 664, A127
- Simola, U., Dumusque, X., & Cisewski-Kehe, J. 2019, A&A, 622, A131
- Sissa, E., Gratton, R., Desidera, S., et al. 2016, A&A, 596, A76
- Smith, A. M. S., Gandolfi, D., Barragán, O., et al. 2017, MNRAS, 464, 2708
- Struve, O. 1952, The Observatory, 72, 199
- Suárez Mascareño, A., Rebolo, R., González Hernández, J. I., & Esposito, M. 2017, MNRAS, 468, 4772
- Ter Braak, C. J. F. 2006, Statistics and Computing, 16, 239
- Triaud, A. H. M. J. 2018, in Handbook of Exoplanets, ed. H. J. Deeg & J. A. Belmonte (Springer International Publishing), 2
- van den Burg, G. J. J. & Williams, C. K. I. 2022, An Evaluation of Change Point Detection Algorithms
- Van Rossum, G. & Drake, F. L. 2009, Python 3 Reference Manual (Scotts Valley, CA: CreateSpace)
- Vanderburg, A., Huang, C. X., Rodriguez, J. E., et al. 2019, ApJ, 881, L19
- Vanderburg, A., Plavchan, P., Johnson, J. A., et al. 2016, MNRAS, 459, 3565
- Varvoglis, H. 2006, AIP Conference Proceedings, 848
- Vaughan, A. H., Preston, G. W., & Wilson, O. C. 1978, PASP, 90, 267
- Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018, AJ, 155, 48
- Wells, D. C., Greisen, E. W., & Harten, R. H. 1981, A&AS, 44, 363
- White, O. R. & Livingston, W. C. 1981, ApJ, 249, 798
- Wilson, O. C. 1978, ApJ, 226, 379
- Winn, J. N. 2010, arXiv e-prints, arXiv:1001.2010
- Winn, J. N. & Fabrycky, D. C. 2015, ARA&A, 53, 409
- Wright, J. T., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2004, ApJS, 152, 261

- Zamaro, M. 2011, Tesi di laurea in ingegneria industriale, Politecnico di Milano
- Zechmeister, M. & Kürster, M. 2009, A&A, 496, 577
- Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, A&A, 609, A12
- Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. 2003, Computational Statistics & Data Analysis, 44, 109