UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Master degree in Astrophysics and Cosmology

Final dissertation

A new perspective on Kepler-68: a scaled-down solar system analog

Thesis supervisor

Prof. Luca Malavolta

Candidate

Marco Margini

Academic Year 2020/21

Contents

Abstract

Kepler-68 is a G1V old main sequence star known to host two transiting planets and a non-transiting one. With a magnitude in the Kepler band of 10.00 it is a suitable target for RVs follow ups and spectroscopic analysis. It is also part of the Gaia Early Data Release 3 (EDR3) that contains very precise coordinates and parallax. It is a low-activity star, both from the photometric and spectroscopic point of view.

The two transiting planets are named Kepler-68 b and Kepler-68 c. They have a short-period orbit around the host star of about 5.4 days and 9.6 days respectively and they are likely to be rocky Earths or super-Earths ($\sim 2.3\ R_\oplus$ and $\sim 1\ R_\oplus$ respectively). The non-transiting planet, named Kepler-68 d, has a period of about 630 days and, even if the radius is not known, it is likely to be a Jupiter analog, as its minimum mass is around 0.77 $M_{Jupiter}$.

In this work, we aim to study again and more deeply this system both from the photometric and spectroscopic point of view. The full Kepler light curve study with updated algorithms together with a new and thorough analysis of the stellar physical parameters led to a significant improvement in the planetary parameters determination. Thanks to new radial velocity data collected by the HARPS-N instrument at Telescopio Nazionale Galileo for this project, it was possible to increase the precision on the physical parameters also from the RV point of view. Together with the already existing public data from HIRES@Keck, we built an RV dataset with a temporal baseline longer than 11 years. Taking advantage of the new data, it was possible to infer the presence of a fourth brand new planet that turned out to have a period of about 1550 days and a minimum mass of $\sim 0.133\ M_{Jupiter}$. The name of Kepler-68 e has been given to this new planet. In addition we used the new TESS data to study, for the first time, possible transit time variations (TTVs) using ephemerides built on Kepler data. Based on our analysis, no long term TTVs are found.

We conclude with the determination of the density of planets b and c and their location on the Radius-Mass diagram, with Kepler-68 b being one of the planets with the most precise density measurement among the mini-Neptunes. Moreover, a brief discussion on the possible migration paths of the planets is addressed. All considered, this study constitutes the deepest and most updated study of the Kepler-68 planetary system, that is still far from being fully characterized and represents an amenable target for future photometric and spectroscopic measurements.

Sommario

Kepler-68 è una vecchia stella di sequenza principale di classe spettrale G1V conosciuta per ospitare due pianeti transitanti e uno non transitante. Avendo una magnitudine di 10.00 nella banda fotometrica di Kepler, la stella è un obiettivo abbordabile per ulteriori osservazioni con le velocità radiali e per analisi spettroscopiche. Essa fa anche parte della Gaia Early Data Release 3 (EDR3), la quale contiene misure molto precise delle sue coordinate e della sua parallasse. E' una stella che mostra bassa attività, sia dal punto di vista fotometrico sia spettroscopico.

I due pianeti transitanti sono chiamati Kepler-68 b e Kepler-68 c. Hanno un'orbita a corto periodo attorno alla stella ospitante di circa 5.4 giorni e 9.6 giorni rispettivamente ed è probabile che essi siano pianeti analoghi alla Terra o super-Terre ($\sim 2.3\ R_\oplus$ e $\sim 1\ R_\oplus$ rispettivamente). Il pianeta non transitante, chiamato Kepler-68 d, ha un periodo di circa 630 giorni e, anche se il raggio non è noto, è probabile che si tratti di un pianeta di tipo Gioviano dato che la sua massa minima si aggira intorno a $0.77\ M_{Giove}$.

In questo lavoro, il nostro obiettivo è di studiare ancora e più accuratamente questo sistema sia dal punto di vista fotometrico che spettroscopico. Lo studio dell'intera curva di luce di Kepler, usando i più moderni algoritmi, insieme ad una nuova e più accurata analisi dei parametri fisici della stella, ha portato ad un significativo miglioramento dei parametri planetari. Grazie alle nuove velocità radiali collezionate dallo strumento HARPS-N presso il Telescopio Nazionale Galileo specificatamente per questo progetto, è stato possibile incrementare la precisione anche dei parametri fisici derivati da esse. Insieme ai dati pubblici preesistenti di HIRES@Keck, siamo stati in grado di costruire un dataset con una lunghezza temporale di più di 11 anni. Grazie all'utilizzo dei nuovi dati è stato possibile inferire la presenza di un quarto nuovo pianeta che si è rivelato avere un periodo di circa 1550 giorni e una massa minima di circa $\sim 0.133\ M_{Giove}$. A questo nuovo pianeta è stato dato il nome di Kepler-68 e. Oltre a ciò abbiamo usato i dati provenienti da TESS per studiare, per la prima volta, possibili variazioni del tempo di transito usando le effemeridi costruite sui dati di Kepler. In base alla nostra analisi, non sono state trovate variazioni dei tempi di transito sul lungo termine.

Concludiamo con la determinazione della densità dei pianeti b e c e la loro posizione sul diagramma Raggio-Massa, sottolineando che Kepler-68 b è uno dei pianeti con la densità misurata più precisamente tra i mini-Nettuno. Inoltre, riportiamo una breve discussione sui possibili percorsi di migrazione seguiti dai pianeti. Alla luce di quanto esposto, questo lavoro costituisce il più profondo e aggiornato studio del sistema planetario Kepler-68, che è ancora ben lontano dall'essere completamente caratterizzato e rappresenta un obiettivo interessante per future osservazioni sia fotometriche che spettroscopiche.

Preface

This written piece is oriented to the presentation of the study of an exoplanetary system called Kepler-68. We will go through all the steps, aspects, details and issues of this kind of analysis. The *Chapter* 1 is intended to be a very fast overview of the state of the art of exoplanetary research. In *Chapter* 2 the steps from the photometric data acquisition to its preparation and analysis are accomplished. In *Chapter* 3 are presented the same kind of contents as before, but regarding spectroscopic data. The all-inclusive analysis is performed in *Chapter* 4, together with the presentation of the final results and their discussion.

Chapter 1

Introduction

1.1 Exoplanetary astrophysics

1.1.1 What is an exoplanet

The prefix *exo-* is indicating something that is *outside* with respect to something else. In this case, it is used to identify planets that live outside the Solar System. But the definition itself of *planet* has undergone some changes in the past and still today not all the scientists agree on it. The official definition given by the International Astronomical Union (IAU) in 2006 is the following:

> *A planet is a celestial body that: (a) is in orbit around the Sun, (b) has sufficient mass for its self gravity to overcome rigid body forces so that it assumes an hydro-static equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit.*

This definition encounters some problems when it is difficult to certify that the orbit is clear and the shape of a small rocky body is nearly round. But even bigger problems arise when we try to define the upper limit of the definition of a planet. There are mainly two school of thought: who is saying that mass should be used to discriminate between giant planets and brown dwarfs and who is invoking how the body originated in principle. Based on the mass threshold, all the bodies that are massive enough to ignite deuterium in their core, are to be considered brown dwarfs. Based on theoretical models and assuming the metallicity of our solar system, this threshold is set to more or less $13 M_{Jupiter}$ (e.g. Chabrier & Baraffe, 2000). For a moderate number of exoplanets though we do not have precise mass measurements: this is a big operative problem of this definition.

Still in those years Soter (2006) suggested another criterion for the definition of a planet at its high-mass tail: *A planet is an end product of disk accretion around a primary star or sub star.* Based on this assumption the mass threshold should be moved to $25 - 30 M_{Jupiter}$. This is what has been done by Schneider,

Dedieu, Le Sidaner, Savalle, and Zolotukhin (2011) when they set $25M_{Jupiter}$ as the upper limit for including objects in the Exoplanet Encyclopedia. People that are using this convention argue that there is no astrophysical reason why a planet could not burn its deuterium (Chabrier, Johansen, Janson, & Rafikov, 2014). For the sake of completeness it is worth to report also the mass threshold between deuterium-burning bodies and hydrogen-burning bodies: $0.08M_{\odot}$ or around $84M_{Jupiter}$ (e.g. Lang, 2013).

1.1.2 Present knowledge

Our today view of exoplanets is very partial and biased. The first exoplanet discovered around a solar-like star was *51 Pegasi b* (Mayor & Queloz, 1995). Since then, more than forty-five hundreds new planets have been discovered, still a tiny fraction of the total number of planets. Being hundreds of billions of galaxies out there, each of them having some hundred billion stars, the possibilities for exoplanets to form are countless. Some extrapolations reported in Perryman (2018) suggest that around 30% of stars or more can have at least one small planet. Together with these consideration, one should keep in mind also that planets exhibit a huge diversity, way greater than stars: planets' dimension is just one parameter, they can differ by the chemical composition and thus the density, by the atmosphere, by the surface temperature and so on. It is needless to say that we need much more completely characterized planets to have a sufficient comprehension of them.

Despite the low amount of data, we already started to classify planets based on their most easy-to-detect characteristics: radius and mass. Before going through this, it is worth to recall some useful conversions between solar system bodies that are often used as an unit of measure for extrasolar planets: the mass of the Sun is about $1040M_{Jupiter}$, while the mass of Jupiter is about $318M_{\oplus}$; the radius of the Sun is around $10R_{Jupiter}$ and the radius of Jupiter is around $11R_{\oplus}$. As reported in Perryman (2018), there are various choices that can be done for the classification, that has to be intended as arbitrarily and non-univocally defined. The classification in radii by Borucki et al. (2011) is the following:

- Earth-size $< 1.25R_{\oplus}$
- super-Earth-size $1.25 - 2R_{\oplus}$
- Neptune-size $2 - 6R_{\oplus}$
- Jupiter-size $6 - 15R_{\oplus}$

The classification in masses by Stevens and Gaudi (2013) is the following:

- sub-Earths $< 0.1M_{\oplus}$
- Earths $0.1 - 2M_{\oplus}$
- super-Earths $2 - 10M_{\oplus}$

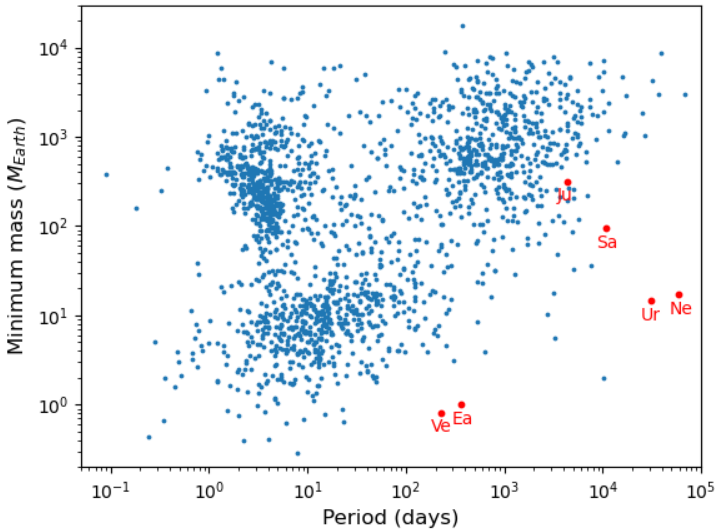

Figure 1.1: Confirmed exoplanets plotted by their period and mass. In red, planets of the Solar System are plotted for comparison.

- Neptunes $10 - 100 M_{\oplus}$ (or $0.31 M_{Jupiter}$)
- Jupiters $0.31 - 3.1 M_{Jupiter}$
- super-Jupiters $3.1 - 13 M_{Jupiter}$
- brown dwarfs $13 - 73 M_{Jupiter}$ (or $0.07 M_{\odot}$)
- stellar companions $> 0.07 M_{\odot}$

In the last years thanks to the rapid increase of the number of exoplanets discovered, an hybrid population that can not fit in these definitions has been discovered. Mini-Neptunes combine the size of a small Neptune ($2 - 3 R_{\oplus}$) and the masses of an Earth or super-Earth. This leads to a planet similar to the Earth in the sizes but gaseous-like in the density. Gaseous planets even lighter than this are not observed because their mass would not be enough to keep all the gas gravitationally bounded to the surface.

To show the extreme variability and diversity that goes far beyond what can be seen in our solar system, figures 1.1, 1.2 and 1.3 are reported. In figure 1.1 the distribution of masses versus period is shown. The population of hot Jupiters can be identified: are those with a period shorter than 10 days and a mass in the range written above for Jupiter-like planets. As can be seen, the vast majority of known exoplanets orbit nearer to their host star than solar system planets do. This is for sure strongly biased by the techniques that are used to

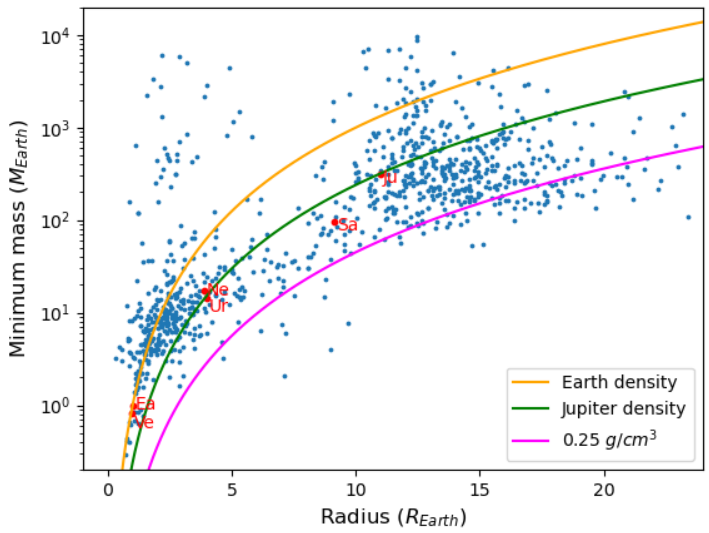

Figure 1.2: Confirmed exoplanets plotted by their radius and mass. Some iso-density curves are shown to let the plot be more clear. In red, planets of the Solar System are plotted for comparison.

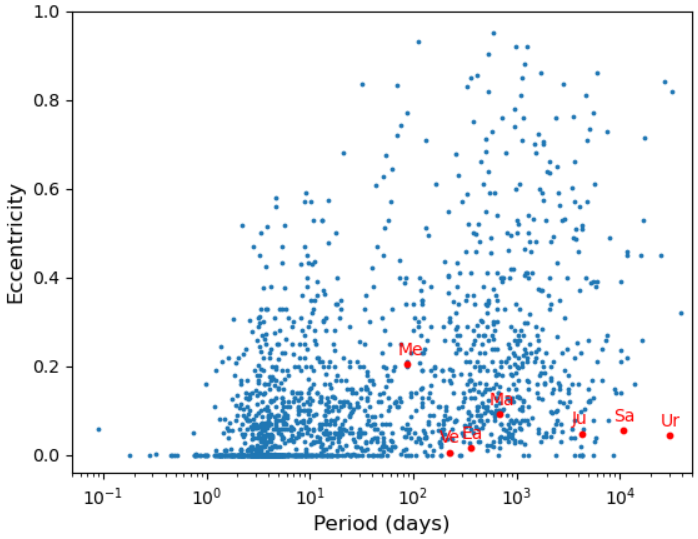

Figure 1.3: Confirmed exoplanets plotted by their period and eccentricity. In red, planets of the Solar System are plotted for comparison.

detect them (see next section), but at least now we know that there are many planetary systems that are way different from our solar system.

In figure 1.2 is clear that this diversity in the orbits is found also in the density and thus composition of exoplanets. The distinction between small thus rocky and large thus gaseous does not hold in the exoplanetary population. For a fixed radius, we can find up to three orders of magnitude of span in masses. Unfortunately, from the density it is impossible to determine the internal composition of these worlds: there is degeneracy. For example, a planet with a density of $\sim 2g\ cm^{-3}$ could be made out of an iron core and a thick atmosphere that lightens the overall density, or a lighter internal composition, e.g. silicon, with an huge planet-wide ocean of liquid water. More complex models try to constrain the composition of a planet by using not only the density, but still many degeneracies persist. The eccentricities distribution in figure 1.3 shows that planets with high eccentricity ($e > 0.5$) can exist even if in our Solar System we don't observe them.

Keeping in mind the observational biases and the incompleteness of the data, it is still worth to report some statistics on the exoplanets discovered until today, rounding numbers to the unity since they change day by day. Around the 33% of confirmed planets are Jupiters. About 21% are Neptunes and more or less 27% are super-Earths. Close to 18% of confirmed planets are Earths and finally the 2% are sub-Earths. Inside these categories hot, warm and cold planets can be identified based on their separation from the hosting star. Among discovered planets the most frequent are hot ones, then the warm and finally the cold ones. The hot-to-warm and warm-to-cold planets ratios are higher for small planets with respect to giant planets (Credit: The Planetary Habitability Laboratory @ UPR Arecibo (phl.upra.edu)).

In about 78% of the cases just one planet is found orbiting a certain star. Around 15% of planetary systems host two discovered planets. More or less 5% hosts three planets and the last 2% has four or more planets. Only one planetary system, called Kepler-90, has reached 8 exoplanets, just the same number of the Solar System (Credit: The Planetary Habitability Laboratory @ UPR Arecibo (phl.upra.edu)).

1.2 Detection techniques

In the following the techniques that are in use today are introduced, focusing specifically on the techniques that we will use in this thesis.

1.2.1 Radial velocities

Radial velocity technique (RV) is based on two physical phenomena. The first, is the common barycentre orbit: two bodies that are subject to their reciprocal gravitational force orbit both around the barycentre of the system. This is true for a star-planet system where, due to the large difference between masses, the

star performs a smaller orbit compared to the one that the planet does. In jargon the motion of the star is called reflex motion and it is the one that could be measured with this technique. The second phenomenon is the Doppler effect: if the motion of the star happens to be on a plane almost edge-on with respect to our line of sight, there will be a significant component of the velocity of the star directed towards us. Using Doppler shifts of the star's spectrum we are able to reconstruct the motion of the star and thus of the planet (e.g. Lovis & Fischer, 2010).

First of all, it is necessary to acquire a spectrum with very high precision (see section *Instrumental noise*). Subsequently there is the need to transform a shift on wavelengths to a shift in radial velocity through the Doppler effect formula, here in a simplified version for non-relativistic cases (Lovis & Fischer, 2010):

$$\lambda_B = \lambda_E \left(1 + \hat{k} \cdot \vec{V}_{star} c^{-1}\right) \tag{1.1}$$

where λ_B is measured in the barycentre of the solar system reference frame, λ_E is the wavelength emitted. $\hat{k}$ is the radial versor, $\vec{V}_{star}$ is the velocity of the star along its orbit as a function of time and c is the speed of light. Applying this formula to every observation, one may end up with a temporal series of radial velocity measurements. Mixed in radial velocity measurements there are signals coming from the stellar activity together with keplerian modulations given by the presence of planetary companions. In section *Astrophysical noise* we list ad summarize the main kind of signals that are not related to the keplerian motion of the planets around the star. For what concerns keplerian signals, they are fitted with an appropriate model, from which we want to extract the main planetary parameters. Among which, the most interesting are the period, the semi-amplitude and the eccentricity. The following formula is used to relate the semi-amplitude to the mass of the planet (Lovis & Fischer, 2010):

$$K = \left(\frac{2\pi G}{P}\right)^{\frac{1}{3}} \cdot (M_s + M_p)^{-\frac{2}{3}} \cdot (M_p sin(i)) \cdot \left(1 - e^2\right)^{-\frac{1}{2}} \tag{1.2}$$

where K is the RV semi-amplitude, P is the period of revolution of the planet. M_s is the mass of the star and M_p is the mass of the planet; e is of course the eccentricity of the orbit. i is the inclination angle between the orbital plane and the line of sight; it is measured in such a way that it is 0° when the system is seen face-on and the radial velocity signal is null and it is 90° when the system is edge-on and the RV signal is maximum. It can be seen that smaller periods lead to greater RV signal, and the same is true for the mass of the star. While on the other hand the bigger is the mass of the planet and the greater is the semi-amplitude, and the same is valid for the eccentricity. It is very important to notice that in principle the inclination i is unknown. Then, only the product $M_p sin(i)$, called minimum mass, can be measured. So it is impossible to determine the exact mass from RV only: just a minimum value for it is found.

Problems related to this kind of survey are mainly due to the fact that a relatively bright star is required to have enough signal-to-noise in all the used

parts of the spectrum, and to the fact that with current instrumentation we are only able to stare at a star at once. It is a very time-demanding technique and due to the upper limiting magnitude, in general it is not possible to study stars relatively far away from us or belonging to a very late stellar type (late M stars). As already anticipated before, there is also a bias in the orbital inclination of the orbits with respect to us. Although this is an important bias, it is affecting all the planetary types in the same way, while the preference for bright and quiet stars may introduce some biases in the diagrams that have been shown before. As it has been outlined during the explanation of the formula for the semi-amplitude, this technique favours high mass planets, orbiting near their low-mass star with a preferably high eccentricity. Until today close to one thousand extrasolar planets have been discovered with this technique.

Last but not least in this method there are also false positive events. They are mainly due to astrophysical noise, i.e. stellar activity that superimposes with keplerian signals in the series of data. Thus, they will be treated in the dedicated section.

1.2.2 Transits

The idea behind the transit technique is very simple: when a planet is passing in front of its star, it is blocking a certain amount of light and therefore we see a drop in the flux coming from the star. In other words, the alignment Earth-planet-star is called transit and the Earth-star-planet alignment is called occultation or secondary eclipse. Measuring the brightness of a star several times in the same night one can build up a temporal series of data called light curve (Winn, 2010). The depth of the dip in the light curve is called δ and it can be approximated by the formula:

$$\delta = \left(\frac{R_p}{R_s}\right)^2 \tag{1.3}$$

The light curve can be fitted with some complicate models, just like the RV curve, and several parameters can be found, assuming to know everything about the star from other studies. The most important parameter that can be retrieved is certainly the planetary radius. Then, also the inclination i of the orbit with respect to us and the period from multiple transits can be calculated. This technique led to the discovery of the majority of the confirmed exoplanets until today because it has one fundamental advantage with respect to radial velocities: we are able to observe thousands of stars at once and then easily cover the whole sky. It is also suitable for fainter stars with respect to those observable with RVs.

The main bias is that since the planet has to transit in front of the star, the inclination angle must be extremely close to 90°. Moreover there is an obvious preference for big planets because the dip is more easily detectable (see formula above). Also the period is playing an important role: long period planets are more difficult to be detected because the probability for them to transit is lower,

the explanation is coming from simple geometric considerations. Additionally, they transit more rarely and this has two consequences: it is more difficult to see them more than once and their transits are so long that are difficult to model. Thanks to the possibility of performing observations of a lot of stars at once, this is the most prolific detection technique, with more than three thousands exoplanets discovered.

The false positive scenario here is more linked to phenomena that can mimic the transit of an exoplanet rather than noises coming from the star itself. For a complete description belonging to the Kepler survey, see Fressin et al. (2013). An example of a false positive scenario is a grazing binary, a binary where there are transits and occultations but the projected disks of the two stars are only partially superimposed. Or another possibility is to have an eclipsing binary in the same field of view of the star target of the observations.

In all those cases it is important to take follow-up observations carried with other techniques, one above all radial velocity. The synergy with RVs is very important to address the problem of false positives and to solve the degeneracy in the inclination angle. From the modelling of the transit light curve it is possible to retrieve the angle i that can be used to calculate the true mass of the candidate exoplanet. Knowing the mass and the radius it is possible to calculate the density. As already mentioned before, from the density is not possible to derive the exact chemical composition of the planet, but it is at least conceivable to understand whether it is a rocky planet or a gaseous one.

1.2.3 Transit time variations

Transit time variations (TTV) occur when a transiting planet feels the gravitational attraction of another body other than the host star. Its orbit is slightly perturbed and with very high precision transit data it is possible to see those very small variations. It is worth to notice that the perturber, which can be another planet or even a brown dwarf or a stellar companion in an outer orbit, could not transit. In this case, from the analysis of TTV signals it is possible to retrieve some new information about the perturber and also to discover new planets (e.g. Agol, Steffen, Sari, & Clarkson, 2005). The signal is strongly enhanced if the hidden planet is locked in a low-order orbital resonance with the transiting one, like 1:2 or 2:3. Nevertheless the order of magnitude of this effect is in the best case a few minutes. Around forty exoplanets have been confirmed using this technique, combining both transit and occultation time variations.

There are other related effects that are transit duration variations (TDV) and transit depth variations (TδV). TDVs are possibly due to the presence of an exomoon: the planet-moon system is orbiting its own barycentre and this is causing an alteration in the velocity of the planet during its orbit. TDVs are always followed by TTVs too. This is an even smaller effect to be measured: it can rise up to only some tens seconds of amplitude. It is very difficult to measure and even if nowadays we have some hints of its presence in some systems, the claim of the first exomoon discovered is yet to come (Kipping, 2009). TδVs are variations of the transit depth and they may be caused by apsidal and nodal

precession (Perryman, 2018). There are some cases where they were found but it is still an open research field.

1.2.4 Others

There are some other important techniques used to detect the presence of exoplanets. They are listed below together with a brief explanation.

Direct imaging Direct imaging is a powerful tool that allows to retrieve astrophysical information directly from the planet. To take a picture containing the planet is a tough achievement: almost all the light coming from the star must be masked in order for the faint planet to be detectable. In order to do this, the planet must have a minimum angular separation from the star and must be bright enough. For this reason young and hot gaseous planets are preferred, in a very long orbit that lies in a plane perpendicular to the line of sight (face-on systems). It is also a good choice to observe in near infrared wavelength, where the ratio between the brightness of the planet and that of the star is maximum (Lagrange et al., 2010).

Until today more than fifty exoplanets have been discovered with this method but since the orbits are very long (many tens of years on average) and the observational temporal baseline is still short, almost all of them do not have well defined orbital parameters.

Microlensing Microlensing is an effect predicted by general relativity. GR says that the light must bend under gravity influence, because space-time itself is curved. One natural manifestation of this behaviour is indeed microlensing: when two stars, the lens and the source, occasionally align with the Earth, the light of the source is bent by the lens in such a way that we see the source magnified from our perspective. The lens must be in the foreground with respect to the source, i.e. the distance between the two must be large. The theoretical magnification, expressed in the case of perfect alignment by the formula below, is infinite, although in the real case the alignment can't be perfect and the two stars aren't point sources (Gaudi, 2010).

$$A = \frac{u^2 + 2}{u \cdot \sqrt{u^2 + 4}} \tag{1.4}$$

where u is the angular separation between the two stars in units of θ_E, a scale angle of the phenomenon. If the lens star has an exoplanet, this additional body is perturbing the otherwise symmetric magnification function. It can be modeled to retrieve the mass of the planet and the angular separation from its host star, the lens. If the distance from us to the lens is known, we are able to determine even the true physical separation. Unfortunately this is not always true because the lens and the source are not spatially resolved for the months or years before and after the event, due to their extreme alignment.

It is a very powerful technique since it allows us to observe large portions of the sky at once, as for transits, and to discover otherwise invisible planets. Its biggest flaw is that they are one-time never-repeating events and they are not amenable for follow up programs because very often the lens star is too far from us for the other techniques. Nevertheless something more than one hundred exoplanets have been discovered with this method.

Astrometry The concept of astrometry is based on the motion of the star and the planet around the common barycentre. An instrument precise enough would then be able to record the very faint wobble of the star projected in the plane of the sky. Unlike radial velocity though, astrometry is effective when the system is seen face-on, i.e. the inclination angle i is close to 0°. It is necessary to remember also that the reflex motion of the star increases with the mass of the planet and with the semi-major axis of the orbit of the latter. Last but not least, of course it is distance-dependent: the nearer is the star the greater would be the effect. The measured amplitude of the wobble as an angular quantity in the sky is (Malbet & Sozzetti, 2018):

$$\alpha_s = 0.33 \left(\frac{M_p}{M_\oplus}\right) \left(\frac{a_p}{AU}\right) \left(\frac{M_s}{M_\odot}\right)^{-1} \left(\frac{d}{10pc}\right)^{-1} \mu as \tag{1.5}$$

Where subscript s stands for *star* and p for *planet*. It is clear that a very high precision is required. In the best cases a signal amplitude of 1 milli-arcsecond is expected, while an Earth-like planet around a Sun-like star seen from 10 parsec will give a 0.33 micro-arcsecond signal. It is also important to be able to detrend proper motions of the star in the plane of the sky. The Gaia spacecraft already has the precision required to see some of those tiny wobbles but we haven't confirmed planets yet because of the long temporal baseline needed (Ranalli, Hobbs, & Lindegren, 2018). As for the direct imaging technique, long period planets are favored, but then long observations are needed to constrain the orbital parameters. Moreover, if the mass of the star is known it is possible to retrieve the mass of the planet itself.

Until today, no planets have been detected with this technique below the deuterium-burning mass threshold. However, it is expected to have the discovery of thousands of planets from Gaia in the next years.

Timing variations around pulsars Pulsars are the most precise clocks the Universe has ever made. Since the period of the pulse is so precise it is possible to recognise in its variations some perturbations. For example, if a planet is orbiting around the pulsar, the reflex motion of the stellar remnant is bringing it near to us in some parts of the period an far from us in other parts. So the distance from the Earth to the pulsar is changing slightly during one period and due to the finite speed of light this is translated to a timing variation of the arrival of the pulse. From the analysis of the signal it is possible to deduce the minimum mass of the planet and of course its period (Kramer, 2018).

An important thing to remember is that the motion of the Earth around the Sun is adding another unwanted, but well known, signal. The strength of this method is that it works independently from the distance of the pulsar, while its weakness is that pulsars are uncommon and very particular objects. They are not suitable for life as we know it and thus less interesting from this point of view. Nonetheless it would be interesting to understand what kind of formation and evolution did those planets undergo and how did they survive to the end of the life of their host star, or if they arrived later. Until today, only seven planets have been discovered with this technique.

Pulsation timing variations No star is completely static. There is always some activity on the surface of them. If a star shows regular pulsations, they can be detected and recorded from Earth. For the same principle of the timing variations around pulsars, if there is a planet around a pulsating star it can produce a delay or an advance on the arrival of the pulse (Hermes, 2018).

Also in this case it is important to detrend the motion of the Earth around the Sun. As far as the star is bright enough to be observed with sufficient precision, this method doesn't suffer a decreasing efficiency with distance. As of today there are only two planets discovered with this method.

1.3 Kepler-68

Kepler-68 is the star that will be studied in this work. It will be introduced more into details from the photometric and spectroscopic point of view in the relative chapters. Its coordinates in the sky are $RA\ 19h24m07.754084(80 \pm 58)s\ \ DEC\ +49°02'24.7610(77 \pm 11)''$ (Gaia Collaboration et al., 2021), between the constellations of the Cygnus and the Lyra. It is a G1V old main sequence star (Grieves et al., 2018) known to host two transiting planets and one non-transiting (Gilliland et al., 2013). With a magnitude in the Kepler band of 10.00 it is a suitable target for RVs follow ups and spectroscopic analysis. Thanks to the Gaia early data release 3 (EDR3) also a precise measurement of the parallax and the distance are known: $6.930 \pm 0.010\ mas$ and $144.30 \pm 0.21\ pc$ (Gaia Collaboration et al., 2021). From the activity point of view, it is a very quiet star, as its age and Sun-like mass would suggest.

Kepler-68 hosts a multi-planetary system. In this work, we aim to study this system both from the photometric and spectroscopic point of view. The first two planets, named Kepler-68 b and Kepler-68 c, were discovered with the transit method by the Kepler spacecraft during its main mission. Thanks to a radial velocity follow-up performed with the HIRES instrument at Keck Telescopes, it was possible to infer the presence of a third planet, named Kepler-68 d. All three planets were presented in the literature for the first time by Gilliland et al. (2013). Planet b and c have short periods, being around 5.4 and 9.6 days respectively, while planet d has a Mars-like period of about 630 days. In more recent days new radial velocity data were collected by the HARPS-N instrument at Telescopio Nazionale Galileo. These data are unpublished and they have been

used in this work for the first time. Taking advantage of the new data, it was possible to infer the presence of a fourth brand new planet and to calculate its period of about 1550 days. To this new planet, it has been given the name of Kepler-68 e.

Thanks to the careful light curve re-analysis, the usage of the Gaia EDR3 data, of state-of-the-art computational tools for stellar parameter determination and of the new HARPS-N observations, this work constitutes the most recent expansion of the knowledge that we had on the Kepler-68 planetary system.

Chapter 2

Analysis of photometric data

The target star of this study is Kepler-68. In this chapter a presentation of the steps required to analyse photometric data is given. From their reduction from the instrument pipeline to the transit fitting procedure.

2.1 Introduction to photometry

2.1.1 Photometry

Photometry is about the measurement of the intensity of light radiated by astronomical objects. In this case, the target is a star and thus will be treated as a point source. The wave fronts that reach us are planes and they are focused on the focal plane by the optics of the telescope. In this position there is a detector, in the most cases a Charged Couple Device (CCD), to catch the photons and convert them into an electric current. This electronic signal is then saved as a flux of electrons per second. A CCD is capable of catching photons with a variety of energies across the electromagnetic spectrum. In order to have a control on the photons' wavelength that are detected, a filter is always used. Each filter has its own well-defined band-pass in the wavelength domain. In the exoplanetary field the most used band is the visible but, on top of this however each instrument could have slightly different filters resulting in slightly different bands.

Photometry can be performed both from ground and from space. Space-based telescopes undoubtedly have much advantages, among which the ability to stare to a certain spot on the sky for months or even years almost without interruptions. This is a very important point when it comes to find planetary signals with the transit method. The underlying idea of the Kepler mission is exactly this, as it is explained in the upcoming subsection.

2.1.2 Kepler mission

The photometry of this study is provided by the Kepler mission (Borucki et al., 2010). For about four years between 2009 and 2013 it has observed the same portion of the sky, located between Cygnus and Lyra constellations. This choice led to the unprecedented opportunity of a long temporal baseline observation of a moderate amount of stars. This permitted the discovery of several long-period transiting planets, previously missing in the exoplanet astrophysics landscape. To avoid problems given by the presence of the Earth and the Moon like visual impairment and reflected light, it has been launched on an Earth-trailing heliocentric orbit. Since it is important that the satellite keeps its solar panels always oriented towards the Sun, it has to do some periodical maneuvers. Namely, once every three months or so, it has to stop observing and re-orienting itself to face the Sun with the solar panels. For this reason the data are divided into segments, called *quarters* because there are four of them in one year. On top of this, every quarter is divided in three chunks of one month each* and in between them there is a small break used to send data back to Earth. This is because otherwise the amount of data to store in the on-board memory waiting to be downloaded would have been too much.

2.1.3 TESS mission

Launched in 2018, TESS is the Transiting Exoplanets Survey Satellite. It has a completely different approach with respect to Kepler. It has four cameras, each one linked to a different telescope (Ricker et al., 2015). The squared fields of view of the four cameras are built to be consecutive in the sky, drawing a stripe. Each camera has a field of view of 24x24 degrees, thus the stripe is 96x24 degrees. This field of view is amazingly large if compared with the one of Kepler. It can contain hundreds of thousands of observable stars. Another difference with Kepler is the fact that every more or less 27 days TESS changes the pointing to stare at a different portion of the sky. In this way, it is able to observe one ecliptic hemisphere every 13 pointings, pinning on the ecliptic pole for its movements. Each area observed by the cameras is called "*sector*".

In about two years TESS has been able to observe almost the whole sky. There were left only some areas around the position of the Earth and the Moon, that obstruct the observations with their reflected light and visual impairment, and a thin stripe around the ecliptic equator. In the extension of the mission, which is still ongoing, it has been decided to re-observe the sky with the same strategy but with slightly different pointings to try to fill the gaps between the sectors, especially near the ecliptic. This strategy has been chosen to allow the observation of the largest possible amount of stars in the neighborhood of the Sun. The mission has a total of about two hundreds thousand primary targets, meaning that these stars are observed at a short cadence. While all the other stars that fall in the TESS' field of view are integrated over a longer time interval, as it is explained in the dedicated section.

*Only the short cadence is affected by this additional subdivision, see next pages.

2.1.4 Kepler-68

Kepler-68 is one of the stars observed during this long period of time. It is located at $RA\ 19h24m07.754084(80 \pm 58)s\ \ DEC\ +49°02'24.7610(77 \pm 11)''$ (Gaia Collaboration et al., 2021). All the photometric observations of Kepler-68 performed by the Kepler and TESS spacecrafts are available for download on the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute (STScI). Funding for the Kepler mission is provided by the NASA Science Mission Directorate. STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–26555. They can be downloaded in *.fits* format, a format that contains all the information needed such as the number of the quarter, the date of the observations, the filter used and so on. They covers all the 4-years-long observing time span peculiar of the Kepler observations.

2.2 False positive scenarios

There are some astrophysical events that can mimic transit signals in the light curve. They can lead to false positive detection if not properly individuated and recognized. Hence in this section the most common false positive scenarios are presented, together with the most common ways to detect them and thus validating real transiting planets. Where not explicitly said, the reference for this section is Perryman (2018) (pp. from 196 to 199).

2.2.1 Gravitationally bound companions

A possible source of false positive events can be stellar companions of the target star. Objects that falls into the same photometric aperture and thus their light is indistinguishable from the light of the target. The first scenario is a simple stellar companion doing grazing eclipses. If its presence is unknown, it is normal to consider the sum of the light of the two stars as the light coming from a single hypothetical star. When one of the two stars passes slightly in front of the other it blocks a certain percentage of light. If the inclination is adequate, this eclipse can simulate very accurately a full transit of a planet much smaller than the true size of the companion.

Another scenario can be the presence of an eclipsing binary in a hierarchical triple system. Let the target star be the central object of a hierarchical triple system, meaning it has a binary orbiting around it. In these cases the companions are often much fainter than the central star, meaning that the block of all the light from one of them is causing a shallow dip in the light curve. Being the light curve the result of the light coming from all the stars involved in the aperture, the binary can be totally eclipsing and still provide a plausible false positive signal.

The last scenario is involving a planet transiting in front of a stellar companion, instead of the target star. This case is more subtle than the previous two, because a planet is indeed there, but it orbits a different star. Since the two stars falls in

the same photometric aperture, the transit appears diluted. (NOTE: dilution is present also if the planet is orbiting the main star of the binary.) Moreover, it is easy to retrieve wrong planetary parameters if in the analysis of the light curve we use stellar parameters of the target star.

2.2.2 Background and foreground stars

Another possible source of false positive events can be stars in background or foreground, but that coincidentally appear to be aligned with the target and thus they fall in the same photometric aperture. In particular, there can be background or foreground eclipsing binaries. The principle is very similar to the one explained for hierarchical triple systems: there is an eclipsing binary composed of stars fainter or much fainter than the target star. In this case though, the binary is not gravitationally related to the target star.

Also in this case, there can be true planetary transits, but around stars that lay in the background or foreground with respect to the target star. Planets are transiting stars similar to the intended one, but with different physical parameters. For this reason, the conclusions on planetary parameters extracted from a light curve analysis may be completely wrong.

2.2.3 Validation techniques

There is a lot of interest in the study and accurate detection of false positive signals. There are works that try to predict their impact on present and future exoplanetary surveys, like the one based on simulations of TESS data by Sullivan et al. (2015). Here we present the main techniques and arguments used to rule out possible false positive scenarios. An example of very well done false positive probability estimation is in Barclay et al. (2013), for the Kepler-69 system.

The most effective way to distinguish a planetary and a stellar companion is to study radial velocities. To read more in detail on the radial velocity technique, see chapter 3. Here the only thing that matters is that from this technique we can retrieve the minimum mass of the companion. Talking about radial velocity amplitudes, the difference between a planetary and a stellar signal is as evident as 10 km/s against 0.1 km/s for a Jupiter-mass companion (Evans & Sackett, 2010). Since we are talking about transits, it is highly probable that the inclinations involved are close to 90 degrees, constraining the minimum mass to be considerably near the true mass. Since the masses of a planetary or a stellar companion are very different, this is a very effective technique. The main problem is that Kepler targets are often too faint to be amenable targets for spectroscopic follow-ups.

In the previous subsections we spoke about unresolved companions that fall into the same photometric aperture. The best technique to infer the presence of stellar companions is the high resolution direct imaging. It is often performed from ground, using adaptive optics techniques. This technique can detect companions down to some tenths of arcsecond, up to a difference in magnitude of about 6. Once the exact number of stars involved is determined, it is much

easier to treat the possible degeneracy of the system. Moreover, monitoring stellar companions (i.e. in a hierarchical triple system), it is possible to detect the eclipse that is causing the signal, being able though, to recognize it as an eclipsing binary. High-resolution imaging can be implemented also to observe possible changes in the centroid position correlated with brightness changes. If the companion has a star-like mass, and it is too near to be detected directly even with high-resolution imaging, there is still the possibility to see a tiny wobbling in the PSF. While the center of mass of an unperturbed binary system is at rest, its luminosity centroid may not. If these changes are correlated with the candidate exoplanet's signals, it is highly probable that they are due to an unresolved stellar companion doing grazing transits.

Another very useful technique, easy to setup, is the analysis of odd and even transits. If the signal is due to a real transiting planet the two subsets of transits must be equal. If instead the signal is caused by a grazing stellar eclipse, and the two stars are not perfect twins, even transits would appear slightly different from odd ones. This is true because, alternately, the light coming from one or the other star is blocked.

In some cases it is possible to obtain light curves at different pass bands. The shape of the transit varies with wavelength due to limb darkening effects. But if the transit is emulated by stellar companions, this change is bigger and recognizable.

Other minor techniques are based on a) multi-planet statistics, b) presence of TTVs, c) dynamical stability calculations. The first is based on the fact that, empirically, if an additional planet is discovered in a system where there are already some validated, it is highly probable that the new planet will be validated too. This is true because, statistically speaking, if there has been a false positive, it would have already been seen. The second is based on the fact that from TTVs it is possible to estimate the masses of the involved bodies. And thus, the vast majority of stellar companions can be excluded. The last technique is based on numerical simulations of the orbital evolution of planetary candidates. Giving as inputs the retrieved orbital parameters with the corresponding errors it is possible to simulate back and forth in time the motion of the object. If the outputs are stable over a reasonable amount of time, the orbit can be statistically validated, or not.

2.3 Instrumental noise

The measurements of flux are subject to noise. Going into space, as with TESS and Kepler, solves a lot of problems but the instruments are not perfect of course. It is of extreme importance to know in the best possible way how the instruments behave and what are all the source of noise and thus of error in the observing process. This is the first step needed to be able to subtract or de-trend them. For this reason in this section a list of the main sources of instrumental noise is presented.

2.3.1 Thermal noise

Electrons that come from a Charge-Coupled Device (CCD) are not only originated from the photons that hit it, called photoelectrons. There are some electrons originated thermally in the silicon structure of the CCD, called thermal electrons. Their abundance is strongly function of the temperature, hence it is possible to limit this effect by cooling down the CCD (e.g. Ma et al., 2014). The variance is proportional to the integration time, since the thermal electrons are continuously produced: $\sigma_{th}^2 = D \cdot t$, where D is a function of temperature and other CCD characteristics. Nowadays this contribution to the CCD noise is strongly limited by cooling CCDs with liquid Nitrogen easily to -100 degrees Celsius. In short exposures, it can be limited to a few electrons per image and thus neglected.

2.3.2 Photon noise

The arrival of photons on the CCD is not predictable. Even if the source is perfectly stable, the number of photons caught by the detector has a random component. On top of this, considering the subdivision of the CCD into pixels, the arrival of a photon in a certain pixel follows a Poissonian distribution (e.g. Frenkel, Sartor, & Wlodawski, 1997). Photoelectrons generated in each pixel fluctuates randomly as a consequence. From the Poissonian roots of this process we derive that the associated variance is exactly the incident signal intensity, S. Moreover, also this noise is proportional to the exposure time, leading to the final expression of $\sigma_{ph}^2 = S \cdot t$. Due to the temporal dependence, this type of noise is stronger in long exposures, but apart from this, it is always present since it relies directly on the incident flux.

2.3.3 Readout noise

The readout noise is a fluctuation in the pixel value caused by electrical noise added to the CCD signal (e.g. Fernández Moroni et al., 2012). In practice, electrons that come out from a pixel are amplified and red through specific electric circuits, becoming a number between 0 and 255 in pixel intensity. These processes are noisy, even if nowadays readout circuits are becoming better and better. The main contributor is the CCD amplifier, but also the readout system, the bias and clocks signals are to be taken in consideration (Fernández Moroni et al., 2012). Since this type of noise is added to the image once it is taken, it does not depend on the exposure time. It is a constant value instead, called N_r, so the equation is simply $\sigma_{rn}^2 = N_r^2$. Typical values are $\sim 10e^-$ at 1 MHz of readout rate. Since it is constant, its influence can be reduced increasing the integration time. Apart from this, modern camera architectures are able to dramatically reduce its magnitude.

Summarizing, the signal to noise ratio (SNR) can be expressed like that:

$$SNR = \frac{S \cdot t}{\sqrt{S \cdot t + D \cdot t + N_r^2}}$$

2.4 Preliminary work

This section is meant to show what are the first things that one must always do when it comes to the analysis of photometric data.

2.4.1 Search for field stars

From the Kepler Input Catalog (KIC) there is evidence of a star about 11" away from Kepler-68: KIC 11295432. In Gilliland et al. (2013) they have studied the problem and concluded that the transit signals are without any doubt belonging to Kepler-68. Some years after, in a lucky imaging survey, Ginski et al. (2016) re-observed the star and concluded that it is likely a gravitationally bounded companion of Kepler-68. Moreover, in 2019, when the TESS Input Catalog (TIC) was revised and completed, it has been observed again and confirmed (Stassun et al., 2019). It has been given the name TIC 417676617 and, of course, Kepler-68B.

Even if the pixel resolution of Kepler is 4", thus better than 11", we wanted to calculate the dilution factor of the stellar companion in order to be completely sure that even if a small amount of its light would have fallen in the same pixels as our target, it would not be a problem. Since for Kepler-68B a magnitude in the Kepler band is not provided, we had to go through approximations using V band and TESS band. Kepler-68 has a V magnitude of $m_V = 10.08$ and a TESS magnitude of $m_{TESS} = 9.509$. Its stellar companion has a V magnitude of $m_V = 17.01$ and a TESS magnitude of $m_{TESS} = 15.975$. From these data it has been possible to calculate the dilution factor $D.F. = 10^{0.4(m_A - m_B)}$ equal to 0.0017 for the visible band and to 0.0026 for the TESS band. Since the Kepler band is in between those two it is safe to conclude that the dilution factor for it will be very similar to those calculated. Hence, it is completely negligible.

2.4.2 Types of flux

There are two types of flux coming out from the Kepler pipeline: SAP flux and PDCSAP flux. SAP stands for *Simple aperture photometry* and it is the raw flux coming from the pixels in which the star's light falls. This set of pixels is the so called aperture and it is decided by the pipeline. The flux inside the aperture's pixels is summed and the background is subtracted, to obtain finally the SAP flux. PDCSAP stands for *Pre-search data conditioned SAP* and it is a more elaborated version of the SAP flux. In this set of data the instrumental trends are removed. It is the best-effort product of the mission pipeline and it should be used for the scientific analysis. See figure 2.1, left panel, for comparison.

The light curves comes in two different formats: short cadence and long cadence (Gilliland et al., 2010). Short cadence has an integration time of about 1 minute while long cadence has an integration time of almost 30 minutes. The stars for which the short cadence is available are a few compared to the total number of Kepler targets. This is normal because short cadence data occupy much more on-board memory with respect to long cadence, hence it is not possible

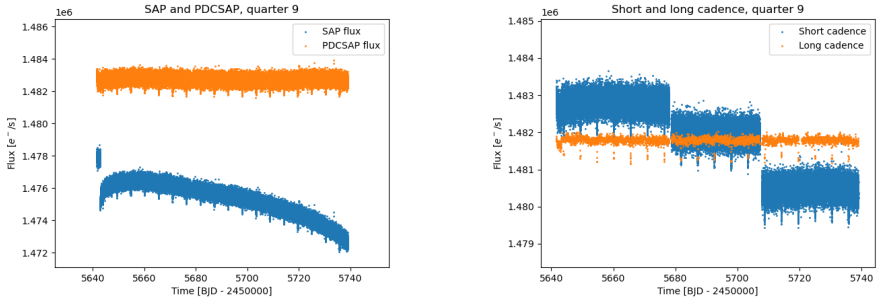

Figure 2.1: Left panel: comparison between SAP and PDCSAP flux. Quarter number 9 is shown as an example.
Right panel: comparison between short and long cadence. Quarter number 9 is shown as an example.

to store it for all the stars. Short cadence is important for asteroseismology targets and planet candidate targets for which one may want to calculate TTVs. This holds because these light curves have better time-sampling, but unavoidably they have a lower signal-to-noise ratio due to the shorter integration time. In addition they may be more affected by systematics than the long cadence light curves (Christiansen et al., 2010). On the other hand long cadence is used for the bulk of core mission science. See figure 2.1, right panel, for an example. It is clearly visible the lower dispersion of the long with respect to the short cadence and also that the subdivision in chunks is affecting only the short cadence.

2.4.3 Quality flags and gaps

The Kepler pipeline is providing a quality flag per each flux measurement. It is meant to be a schematic description of what were the conditions of the instrument and/or what happened at the moment of acquisition of a given flux value. In table 2.1 are reported all the possible outcomes. A quality flag can be the sum of two or more values, if more than one issue has been registered at the same time. Moreover, the time series of data is continuously interrupted by gaps, when the instrument was maneuvered or suspended in the *safe mode*. It is common that after a pause and a re-pointing, the values of flux undergo a discontinuity: the same star changes its measured flux from a quarter to the other. All those behaviours (quality flags, gaps and jumps) must be properly treated in order to have a clean dataset before starting the scientific analysis.

2.4.4 Preparation of the light curve

Long cadence is available for all the quarters from 0 to 17. Short cadence instead is available for the last chunk of the second quarter and for quarters from 5 to 17. The first thing to do is to remove points with flag different from

Value	Description
1	Attitude tweak
2	Safe mode
4	Coarse point
8	Earth point
16	Zero crossing
32	Desaturation event
64	Argabrightening
128	Cosmic ray in optimal aperture
256	Manual exclude
1024	Sudden sensitivity dropout
2048	Impulsive outlier
4096	Argabrightening on CCD
8192	Cosmic ray in collateral data
16384	Detector anomaly
32768	No fine point
65536	No data
131072	Rolling band in optimal aperture
262144	Rolling band in full mask
524288	Possible thruster firing
1048576	Thruster firing

Table 2.1: All the possible quality flags in Kepler data.

zero, which means some problems have been registered. The most common flags in this set of data are *cosmic rays in optimal aperture*, *cosmic rays in collateral data* and *zero crossings*. Sometimes it could happen that some clearly flawed flux measurements survive to the flags, so it is important to keep an eye on the data anyway. In this case, we saw by eye that quarter 8 had a bad behaviour that was visible both in the SAP and PDCSAP. We decided to manually select and discard the first part.

At this point, for a clean and safe analysis of the PDCSAP flux, we proceeded removing NaNs. Now the last thing to do is to let the different quarters match the flux of the previous and the following, removing jumps. This is done by re-scaling the flux of a quarter by dividing for the right factor. The factor is simply the ratio between the mean flux of the last piece of the quarter on the left with the mean flux of the first piece of the quarter on the right. It is fundamental to remember to re-scale also the error bars in order to keep it all proportioned. It is not important to have a millimetric precision, the flattening algorithm described later on will refine this. We can always re-scale the flux because for that type of analysis we will only need the normalized flux, not the absolute one.

2.5 Stellar activity and parameters

Before to proceed with the search of transit signals in the light curve it is important to study properly the host star. We are interested in its activity, namely its rotation period, in order to be able to distinguish such periodic signal from the ones caused by the presence of the planets. To search for the stellar activity it is better to use the PDCSAP flux, since it does not contain instrumental trends that could not be treated properly by the below-mentioned algorithms. Moreover, for the subsequent fit of the light curve, it will be of fundamental importance to know the physical parameters of the star. We first search in the literature what are the state-of-the-art results for the directly measurable quantities, then we use specific tools to retrieve all the others.

2.5.1 Rotation of the star

It is usually possible to determine the rotation of the star around its own axis from high precision photometric time series, such that of our dataset. This effect is mainly due to the presence of dark spots and faculae on the surface of the star. By appearing and disappearing due to the rotation, they are capable to modulate the observed flux. To search for such features in the set of data there are several tools among which to choose.

One of them is the auto-correlation function (ACF) (McQuillan, Aigrain, & Mazeh, 2013). Its strength lies in the fact that it does not need any function a priori, because no fits are performed. The algorithm shifts the signal of a certain amount along the time axis and computes the correlation with the original signal. Doing this for every possible non-trivial shift it obtains a set of correlation factors

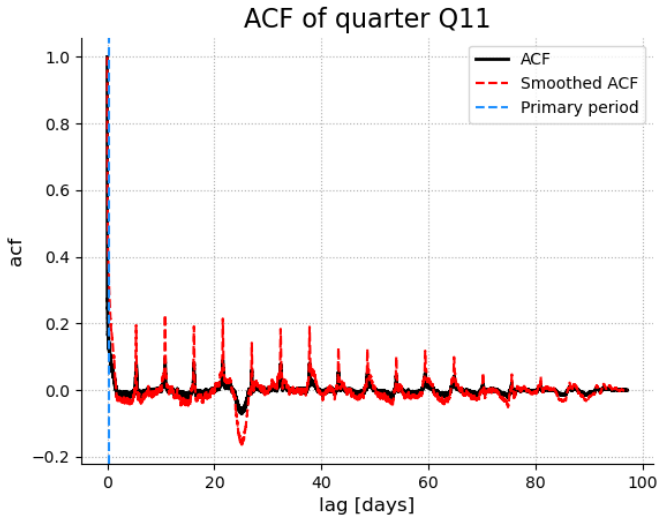

Figure 2.2: Auto-correlation function for quarter number 11.

as function of the shift itself. If there is a periodicity in the dataset, when the shift equals the corresponding period a peak in the ACF is visualized.

Another powerful algorithm to find periodic signals in the light curve is the Generalized Lomb-Scargle periodogram (GLS) (Zechmeister & Kürster, 2009). This makes use of a sinusoidal function to try to fit the data. The goodness of the fit is measured by its χ^2: the lower is this value the better is the fit. So in this case the goal is to find the minimum of the χ^2 as function of the period of the sinusoidal function used.

We decided to analyse all the quarters of short cadence separately. It was a good compromise between the computational time implied and the possibility to study relatively long periods. With ACF the maximum period that could be found is more or less half of the duration of the data; while with GLS it grows up to the entire length of the dataset. As it has already been said, each quarter is around 90 days long. Furthermore, from the literature we expect this star to be quite evolved, thus we do not expect a fast rotation period. That being said, a plot showing the results (either ACF factor or χ^2) as function of the period is produced, for each quarter.

In figure 2.2 is presented a common output from ACF analysis. In figure 2.3 a representative quarter for the GLS technique is shown. We can see in both cases that harmonics of $\sim 5.4d$ period are found. That correspond to the deepest transit signal in the light curve and it has nothing to do with the star itself. There is no evidence of a signal hinting the presence of a rotation period of the star. As we already knew from the previous studies this is a very quiet star (Gilliland et al., 2013). Now we are confirming with our new analysis, with

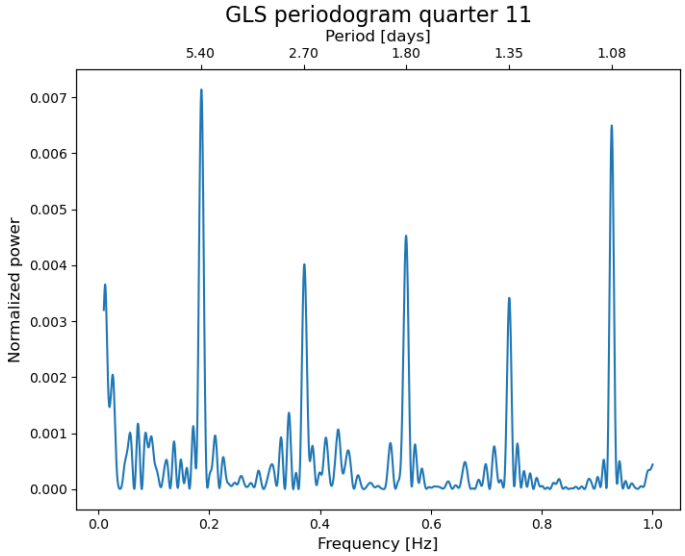

Figure 2.3: Generalised Lomb-Scargle periodogram for quarter number 11.

two different and independent techniques, that Kepler-68 is not displaying a measurable period.

It could be argued that the period is not present because the star is rotating around an axis almost perpendicular to the plane of the sky. For us this hypothesis is very unlikely because of the evidences of at least three planets orbiting it edge-on. It would be very unlikely to have such an abundance of planets all of them in a polar orbit.

2.5.2 Stellar parameters from literature

Kepler-68 is a pretty well-studied target. The first relevant result is coming from the paper presenting for the fist time the planetary system (Gilliland et al., 2013). They cross-correlated three new spectra against a library of synthetic model stellar spectra. Then, they applied the Spectroscopy Made Easy (SME) tool to an HIRES (at the W. M. Keck Observatory) spectrum (Valenti & Piskunov, 1996; Wehrhahn, 2021). Moreover, they used asteroseismology techniques to further constrain the value of surface gravity and at the end added in quadrature some systematics to their estimate of the error.

Another deep study of Kepler-68 has been carried out using spectroscopic tools SME and SpecMatch (SM) (Petigura et al., 2017). In particular, they used an updated version of SME, called SME@XSEDE. While the spectra used are coming from the HIRES instrument.

On top of this, for this study, we use additional information coming from different algorithms and completely new spectroscopic data. The spectra used

T_{eff} (K)	$log(g)$	[Fe/H] (dex)	Reference
5793 ± 75	4.28 ± 0.10	0.12 ± 0.07	Gilliland+2013
5773 ± 75	4.12 ± 0.10	0.13 ± 0.05	Petigura+2017 (SME@XSEDE)
5805 ± 75	4.28 ± 0.10	0.14 ± 0.05	Petigura+2017 (SM)
5908 ± 75	4.28 ± 0.12	0.16 ± 0.05	this work (ARES)
5785 ± 75	4.22 ± 0.10	0.04 ± 0.08	this work (SPC)
5849 ± 75	4.12 ± 0.21	0.12 ± 0.05	this work (CCFpams)

Table 2.2: Results of the spectroscopic analyses on Kepler-68 spectra. Where the errors were too low, we decided to adopt a lower threshold: 75 K for effective temperature, 0.10 for the surface gravity and 0.05 dex for metallicity.

Band	Magnitude	Band	Magnitude
B	10.73 ± 0.04	W1	8.539 ± 0.023
V	10.08 ± 0.03	W2	8.583 ± 0.020
J	8.975 ± 0.046	W3	8.539 ± 0.023
H	8.662 ± 0.029	W4	8.223 ± 0.157
K	8.588 ± 0.022	/	/

Table 2.3: Magnitudes of the target star reported from the online catalogs.

are the same used for the radial velocity analysis, thus coming from the HARPS-N spectrograph at Telescopio Nazionale Galileo (see dedicated section). In particular, we analysed the new spectra with the Automatic Routine for line Equivalent widths in stellar Spectra (ARES) (Sousa, Santos, Israelian, Mayor, & Monteiro, 2007); with Stellar Parameter Classification (SCP) (Buchhave et al., 2012, 2014); with CCFpams, a tool that makes use of cross-correlation functions (Malavolta, Lovis, Pepe, Sneden, & Udry, 2017). The results of all the analyses introduced are reported in table 2.2. As can be seen from the table, the values are more or less in agreement between each other. If in a few cases there are values outside one sigma from the mean, that is not a problem for the subsequent analysis.

Additional information needed for the analysis explained in the next pages are magnitudes in several bands and the parallax. Luckily, we had at our disposal the magnitudes in B and V from the Johnson-Morgan photomeric system (Høg et al., 2000), J, H and K from the 2MASS system (R. M. Cutri et al., 2003) and W1, W2, W3 and W4 from the Wide-field Infrared Survey Explorer (WISE) photometry (R. M. e. Cutri, 2012). WISE is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. The values of the magnitudes are reported in table 2.3. The parallax from Gaia EDR3 6.930 ± 0.010 *mas* was used for the first time in this work to refine the estimates of the stellar physical parameters (Gaia Collaboration et al., 2021).

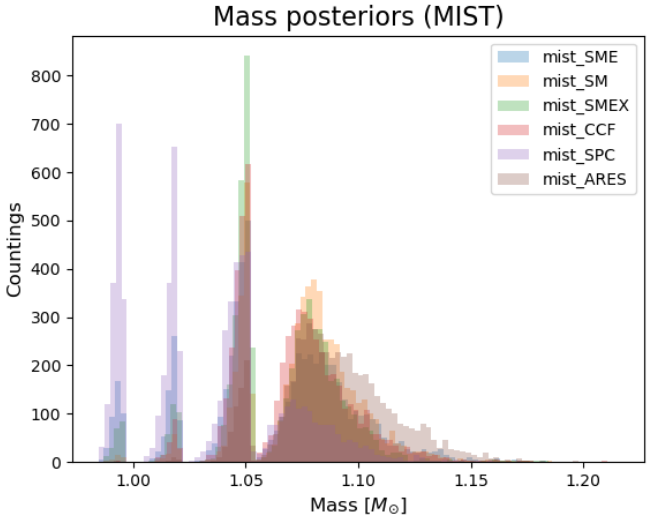

Figure 2.4: mass posteriors for the six datasets using MIST isochrones.

2.5.3 Isochrones analysis

From the quantities derived from spectroscopy above we can directly infer more information about the star. We can derive all the physical parameters that we need for the analysis of the light curve. In order to do so we used isochrones fitting.

The tool used is the *isochrones* package, written in Python (Morton, 2015). The fitting tool needs a library of theoretical isochrones in order to find the one that best represents the data. We decided to use two different databases, i.e. libraries of isochrones models: DARTMOUTH Stellar Evolution Database (Dotter et al., 2007, 2008) and MESA Isochrones and Stellar Tracks (MIST), where MESA means Modules for Experiments in Stellar Astrophysics (Dotter, 2016; Choi et al., 2016; Paxton et al., 2011, 2013, 2015).

Each of the six sets of spectroscopic parameters has been fitted separately, giving back a set of probability distributions (i.e. posteriors), one for each physical quantity. The most important quantities that can be calculated this way are the stellar radius, mass and density, together with the age and the distance. The two models have not given rise to exactly the same results of course, but they agree very well on all the posteriors except for the mass.

As can be seen in figure 2.4, all the six MIST posteriors for the mass have a strange behaviour on the left hand side. They split in a multi-modal distribution. This can be attributed to a degeneracy of the isochrones or to a more serious problem due to the absence of an isochrone in the parameter-space grid of the model. In any case, since we had a lot of data already, and excluding the MIST results from all the posteriors did not change the results significantly, we decided

Quantity	Median	σ_-	σ_+
Radius ($R_\odot$)	1.2564	0.0084	0.0084
Mass ($M_\odot$)	1.057	0.024	0.029
Density ($\rho_\odot$)	0.534	0.020	0.022
Age (Gyr)	6.84	1.04	0.90
Distance (pc)	144.31	0.21	0.21

Table 2.4: Results from the isochrones fit using only DARTMOUTH models.

to use only DARTMOUTH calculations for scientific purposes.

The final distribution is obtained putting together all the six distributions calculated separately. The results listed in table 2.4 are the medians of the corresponding distributions, i.e. the 50 percentiles. The associated errors are $\sigma_- = 50p - 15.865p$ and $\sigma_+ = 84.135p - 50p$ respectively, where p stands for percentile.

2.5.4 Limb Darkening

Limb darkening is a phenomenon that occurs when observing a star: the measured intensity drops moving from the center to the limbs. This happens because the optical depth changes according to the angle between the normal to the stellar surface and the line of sight to the observer. Near the limbs the radiation measured is coming from a rather external layer of the structure of the star, hence cooler than an inner one. This decrease in luminosity from the center to the limbs can be modeled with some functions of $\mu = cos(\theta)$, where θ is the angle described above. The most commonly used is the quadratic limb darkening in the form:

$$I(\mu) = I(1) \cdot \left[1 - a \cdot (1 - \mu) - b \cdot (1 - \mu)^2\right]$$

The limb darkening coefficients a and b are functions of the effective temperature, surface gravity, metallicity and microturbulence. Since the microturbulence parameter is difficult to retrieve and it has a small impact on a and b, it is very often fixed or neglected. In this case, we fixed also the metallicity to the solar one.

We decided to use the tables computed by Claret because of their completeness and proven reliability (Claret, Hauschildt, & Witte, 2013; Claret, 2018). In the first paper he and his collaborators calculate the limb darkening coefficients for the quadratic law (the one used above) and with quasi-spherical models. In particular, they are calculated with two different methods: Least Squares (LS) and Flux Conservation (FC). While in the more recent paper he managed to calculate the same coefficients using fully spherical models. We decided to calculate the coefficients in all the three fashions and to compare them at the end.

The tables that are available in the cited papers are covering a grid of values in the temperature-gravity space. Of course our star did not match exactly one

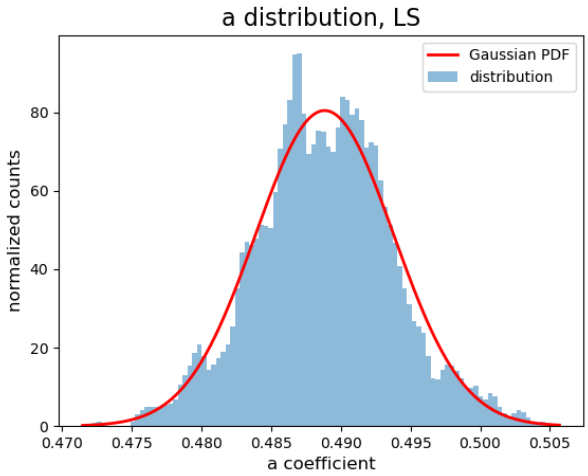

Figure 2.5: Limb darkening coefficient "a" posterior, in the case of quasi-spherical model with LS method.

Coefficient	quasi-sph LS	quasi-sph FC	spherical
a	0.4888 ± 0.0049	0.5123 ± 0.0050	0.4648 ± 0.0051
b	0.1815 ± 0.0024	0.1357 ± 0.0026	0.2003 ± 0.0028

Table 2.5: Results from the interpolation of the limb darkening tables.

of the tabulated $(T_{eff}, log(g))$ pairs, so we had to interpolate the table to get a result. We interpolated them with a simple yet robust bilinear spline function. In order to calculate an error for those results we decided to apply a Monte Carlo approach. From the posteriors of the effective temperature and logarithmic surface gravity of the previous analysis we extracted N values randomly. For all those values we calculated the corresponding a and b coefficients and we built a distribution for them too. It showed up that it was nearly a Gaussian, like the one in figure 2.5. We decided to use the standard deviation of this Gaussian as the error on the median. In the table 2.5 we present the results, divided by type of the model used. We can see that they do not agree on a one-sigma scale. This is not a problem: limb darkening is always difficult to estimate and the theoretical error bars are always inflated to let them more realistic. For now, we did not choose between them. Sometimes it could happen that a more reliable estimate of the limb darkening coefficients comes from the fit of the light curve. So we postpone the discussion to the section where the fit will be presented.

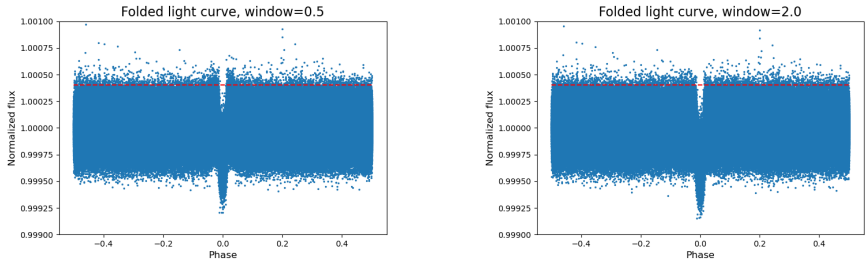

Figure 2.6: Left panel: flattening and folding with a temporal window of 0.5 days. Right panel: flattening and folding with temporal window of 2 days.

2.6 Flattening and transit detection

The model used for the fit of the light curve is perfectly flat outside the transit. For this reason we need to correct the data to reflect this fact. Then, we need to know exactly how many planetary signals are present in the light curve and to find them all.

2.6.1 Algorithms for flattening

The so called flattening of the light curve is the process where trends and non-transit signals are removed from the time series. In order to do so the best way is to use a time-windowed slider based on Tukey's biweight (Tukey, 1977). This is implemented and optimized in the Wōtan package in Python (Hippke, David, Mulders, & Heller, 2019).

There are a lot of parameters to play with, though we mainly focused on the window length. The window is an interval on the temporal axis that is moving along the dataset and inside which the average is calculated. The wider is the sliding window the smoother will be the results: all the features smaller than the window will be averaged out. The thinner is the window the more the average follows the data: in the limiting case where the window contains only one point at a time, the algorithm does nothing. After some trials, we decided to opt for a window length of 2 days. It is several times the duration of the transits (less than 4 hours each) so it is able to completely bypass them but it is not so wide to neglect also long-term trends. It is clearly visible in figure 2.6 the difference between an half-day-wide filter and a two-day-wide filter. In the first one the transit is partially modeled and the outside-of-transit part is modified while in the second one the transit is correctly left untouched.

In the plots the light curves are folded with a periodicity equal to the period of the planet b. This is easily done with some lines of code and allows us to see if the flattening worked or not. We also experimented some combinations of the parameters *break_ tolerance* and *edge_ cutoff* (Hippke et al., 2019). They take care of the gaps in the data and adapt the average by consequence.

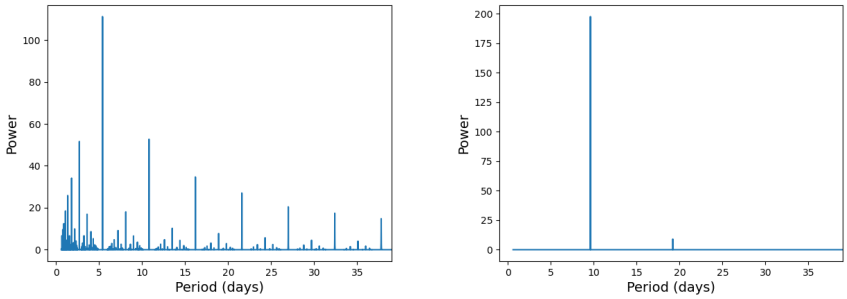

Figure 2.7: Left panel: periodogram obtained with TLS; a strong signal corresponding to the transits of planet b is present, together with its harmonics.
Right panel: periodogram obtained with TLS after masking planet b's transits; a signal corresponding to the transits of planet c is present.

2.6.2 Algorithms for transit detection

There are two algorithms designed for the detection of transiting planetary signals. Box Least Squares (BLS) is the simplest one (Kovács, Zucker, & Mazeh, 2002). It is composed by a flat out-of-transit part and a box-shaped transit profile. It means that ingress and egress are of zero duration and the in-transit part is flat too. Transit Least Squares (TLS) is rather more complex (Hippke & Heller, 2019). The out-of-transit part is still flat but the transit shape is more similar to a true one. The ingress and egress are smoothed to mimic limb darkening effects and the fit is more realistic. BLS has the advantage to be faster because of its simplicity. TLS instead, is better in recognizing planetary signals, i.e. it has a lower false negative rate. In this work we tried both methods and they both succeeded in finding the two already known planets, b and c. Two periodograms produced with the TLS technique are shown in figure 2.7, planet b on the left hand side and planet c on the right hand side.

Transits of planet b are way deeper and more marked than c's transits, so it has been necessary to remove b's transits on order to let c's arise in the periodograms. Since in this work the full Kepler light curve for Kepler-68 has been analysed for the first time, we also searched for new, undetected, long-period transiting planets. To do that, we had to mask both the transits of b and c and to search on the whole dataset, thus including also single-transit signals. The resulting TLS periodogram in shown in figure 2.8. The graph is showing a few sporadic peaks, but with a low associated power. It turned out that, from this study, no other transiting planets are present in the 4-years-long Kepler light curve. In table 2.6 we report the numerical results from the TLS analysis. We decided to use them because of the intrinsic advantages that the TLS model has with respect to BLS; by the way, the results are so similar that they are completely interchangeable. They consist in the periodicity of the transits, their duration, their depth and their central time. Those data are fundamental for

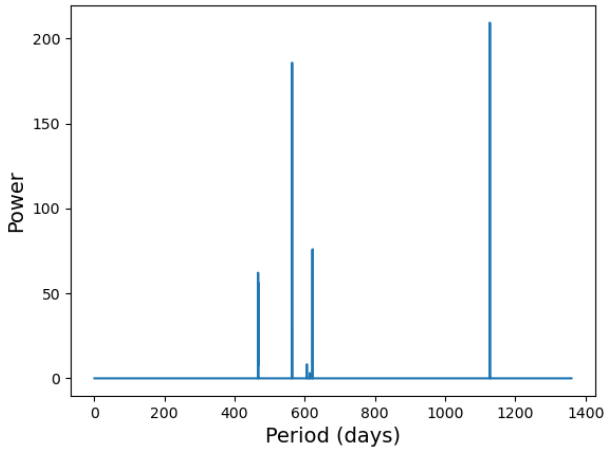

Figure 2.8: Periodogram obtained with TLS after masking planet b's and planet c's transits; no strong evidences of an additional transiting planet are found.

Quantity	planet b	planet c
period (days)	5.3987	9.6050
transit time (BKJD)	233.2465	232.4349
duration (hours)	2.46	1.91
depth (ppm)	345	48

Table 2.6: Results from the TLS analysis for the two planets. BKJD is a peculiar Kepler temporal reference, it corresponds to BJD - 2,454,833 days.

the light curve fit that we had to perform right after, because they provide the priors necessary for the analysis to start. They do not have an error bar, but this is not a problem: we will use an expedient later on.

2.7 Bayesian analysis of photometry

In this section we explain how we performed the first preliminary fits of the light curve. They were used to test the input parameters, commonly know as priors, and to find the best configuration for the final fit that will be discussed in chapter 4. Here still only the photometry has been used.

2.7.1 PyORBIT code

PyORBIT is a modular code that can be used to model planetary signals in both photometric and spectroscopic data (Malavolta, 2016). It has a lot of possibilities in it: more than one planet can be included simultaneously, more

than one parametrization between which to choose for the orbital parameters or for the limb darkening coefficients. It can accept even more than one dataset for the same analysis, each one of them having its own offset (in case of radial velocities). It is also designed to accept both plain boundaries and various distributions as priors. The core of the code is based on a Bayesian approach: there is no more a precise value for every single parameter, equipped with an error bar. A probability density distribution for each parameter is built instead, the so called posteriors. This means that the true value has a certain probability to fall in a given interval.

The code can explore the parameter space and thus find the posteriors in two different ways, based on what the user choose, that are Markov-chain Monte Carlo or Nested Sampling algorithms. Markov-chain Monte Carlo algorithm is based on the presence of walkers, that at each step explore the parameter space starting from the priors and evolving independently. The posteriors are built with the values explored by the walkers, after they reached convergence. The number of steps to be performed varies according to the complexity of the problem. In PyORBIT, this algorithm is implemented through *emcee*, a very strong and widely used software (Foreman-Mackey et al., 2019).

Nested Sampling is the alternative to the Markov-chain Monte Carlo. It is more focused to the comparison between different models: it will be used in chapter 3 and thus it will be explained better in a dedicated subsection.

The parameters of the PyORBIT analysis are passed to the algorithm through a *.yaml* file; how to compile it can be found in the PyORBIT documentation. Finally, it is worth to cite the solver used to perform the fits, that is the differential evolution-based Python package PyDE (Storn & Price, 1997).

2.7.2 Fits with only planet b

The goal was to understand how many and which priors were to be used for stellar parameters and planetary ones. Since we had two transiting planets, the options were either to use only one of them or to use both. As we already had the opportunity to see when we folded the light curve, one of the two planets has a much deeper transit with respect to the other. So we decided to use only planet b in this phase, as it is a relatively deep transit with no evident problems, while including planet c will have not improved the results. Those considerations set planet b to be a perfect tool to work out the parameters that we need from this kind of analysis.

After having converted times from BKJD to the more common BJD-2,450,000 and having selected with a proper mask only the transits of planet b, we performed the first of four fits. Some priors were present in all of them, some others varied. Flat priors on period, central time of transit and eccentricity, together with Gaussian prior on stellar radius (which is only used to convert planetary radius into physical units) were always present. On the other hand, the priors which we wanted to investigate were on the eccentricity and stellar density. Thus we performed four fits: all the combinations of presence and absence of those two priors. We left limb darkening coefficients free because we wanted to see which

Quantity	prior type	1st value	2nd value
Period (days)	$\mathcal{U}$	5.30	5.50
Transit time (BJD-2,450,000)	$\mathcal{U}$	5066.20	5066.30
Eccentricity	$\mathcal{U}$	0.0	0.5
Stellar radius ($R_\odot$)	$\mathcal{N}$	1.251	0.010
Eccentricity	$\mathcal{HN}$	0.0	0.083
Density ($\rho_\odot$)	$\mathcal{N}$	0.540	0.021

Table 2.7: Priors used in the first four fits. $\mathcal{U}$ stands for uniform prior, $\mathcal{N}$ stands for Gaussian prior and $\mathcal{HN}$ stands for half-Gaussian prior. In the case of uniform prior, the two values reported are the two extremes of the interval; in the case of Gaussian or half-Gaussian prior, the two values are median and standard deviation.

Origin	coefficient a	coefficient b
Claret+2013 LS	0.4888 ± 0.0049	0.1815 ± 0.0024
Claret+2013 FC	0.5123 ± 0.0050	0.1357 ± 0.0026
Claret+2018	0.4648 ± 0.0051	0.2003 ± 0.0028
Fit 1	0.409 ± 0.027	0.306 ± 0.043
Fit 2	0.406 ± 0.027	0.304 ± 0.046
Fit 3	0.407 ± 0.028	0.306 ± 0.045
Fit 4	0.405 ± 0.027	0.308 ± 0.045

Table 2.8: Limb darkening results from the photometric fits compared to the theoretical ones found in previous sections.

values are retrieved from a data-driven approach. The prior on eccentricity is taken from a study on multiple planetary systems discovered by Kepler (Van Eylen et al., 2019). In table 2.7 the used priors are shown, both the fixed and the variable. The interval on the period is calculated subtracting and adding the transit duration to the period found with the TLS technique, since we are sure that the code could not get it wrong by a quantity larger than this. The interval on the central time of transit is calculated in the same way, but subtracting and adding half of the duration, for the same reason.

In the fits without priors on the density its resulting value is still coincident with the prior given to the other two analyses, but with more variance. This can be seen in figure 2.9 and it means that the photometric data are in agreement with theoretical models of stellar parameters. Limb darkening coefficients found in all the four analysis are around 0.40 for coefficient a and 0.30 for coefficient b. From table 2.8 it is clear that they do not agree with any of the theoretical values that we have found before. The behaviour of eccentricity is following the prior that it is given. If only the uniform prior is given, the posterior is almost flat occupying all the interval. If also the half Gaussian prior is given, the posterior is giving back almost the same shape. This is an hint that our observations are not constraining in any way the eccentricity.

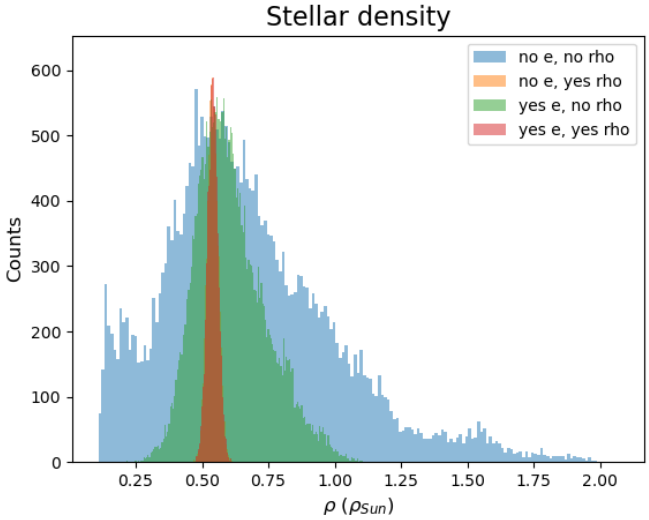

Figure 2.9: Posteriors of the stellar density. Their median values are almost indistinguishable. Red and orange histograms are superimposed

Based on what we have just described, we decided to use the theoretical prior on density in the subsequent analysis. We also decided to use the prior on eccentricity when performing photometry-only fits, leaving the spectroscopic discussion for later. This choice came after simple considerations on the reciprocal position of the two planets. We observed that, assuming that the two orbits are co-planar, even with an eccentricity of 0.3 they start to overlap. Considering this and considering the relatively old age of the star (so the planets must occupy stable orbits), we concluded that most likely the orbit are nearly-circular and so it is worth to constrain that with the result by Van Eylen et al. (2019). Lastly, we decided to trust the Kepler photometric data for limb darkening, considering them more reliable than theoretical ones.

2.7.3 Fits with planets b and c

We used what we have just learned to perform a fit including both planet b and planet c, thus using priors on stellar density and orbit eccentricities but not on limb darkening. Firstly it was necessary to select only the transits of the two planets, discarding the vast majority of the out-of-transit part. This is because that part is not carrying any information and deleting it saves a lot of computational time. From this fit we derived the orbital and physical parameters of the planets that are typical of the transit technique; they are listed in table 2.9.

Since the PyORBIT code is based on a Bayesian approach, the results are probability distributions, called posteriors, that tell us what is the probability

Quantity	planet b	planet c
Period (days)	$5.398751(95^{+77}_{-79})$	$9.6050(43 \pm 15)$
Radius (R_*)	$0.017(19^{+11}_{-13})$	$0.007(14 \pm 13)$
Duration T_{14} (hours)	$3.451^{+0.114}_{-0.081}$	$2.98^{+0.18}_{-0.15}$
Transit time (BJD-2,450,000)	$5066.244(94 \pm 15)$	$5065.42(91^{+14}_{-13})$
Semi-major axis (R_*)	$10.54^{+0.13}_{-0.13}$	$15.48^{+0.20}_{-0.20}$
Impact parameter (R_*)	$0.508^{+0.038}_{-0.052}$	$0.788^{+0.025}_{-0.032}$
Eccentricity	0.059	0.087

Table 2.9: Results for the planetary parameters from the fit of the entire light curve containing both planet b and c. The value reported for the eccentricity is an upper limit as described in the text.

that a certain parameter is falling between a certain interval. In this work-frame it is not contemplated that a variable assumes a specific value with an associated error, which is typical of a frequentist approach. Nevertheless it is useful for displaying purposes to extract a single number from the posteriors in order to be able to compare it with other studies in the existing literature. As the representative value of each posterior we have chosen the median, i.e. the $50.0th$ percentile, as the positive half of the error bar we used the $\sigma_+ = 84.135p - 50.0p$ and for the negative half $\sigma_- = 50.0p - 15.865p$, where p stands for percentile. We decided to keep the error asymmetric to try to preserve as much information as possible of the original distribution, using only three numbers. The eccentricity is the only exception because its posterior is squashed toward zero, hence it is not feasible to use the median with the errors. The solution is to use an upper limit corresponding to the $68.27th$ percentile, as it has been done in table 2.9. This means that there is a 0.6827 probability that the value of eccentricity is lower than that.

To visualize the results is also useful to plot the folded light curve for the two planets. Using the period found by PyORBIT it is possible to fold the dataset on itself to highlight the transit shape. When folding the curve for planet b it is better to subtract the signal of planet c and vice versa, to better see the result. They are shown in figure 2.10.

We compared these results with the existing literature, finding that they are in agreement with it. This is already a very good result, considering that we did not introduce radial velocities yet. That will be the topic of the next chapter.

2.8 TESS analysis

In this section the analysis of the TESS data is shown. The same procedure used for Kepler data has been used here. For this reason some explanations are not repeated, the ones given before are valid also in this section. The aim of

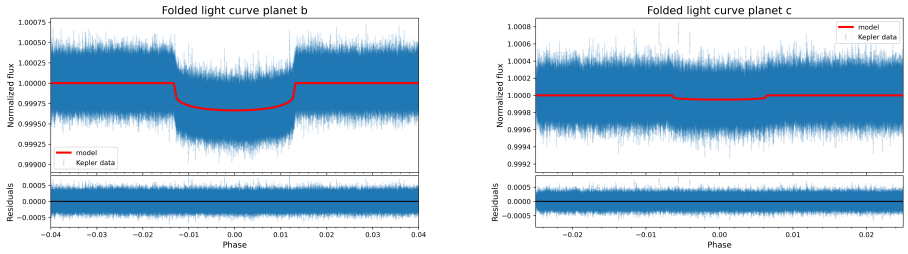

Figure 2.10: Left panel: folded light curve for planet b. Right panel: folded light curve for planet c.

this analysis is to be able to compare some physical parameters of the planets calculated with two different data sets and in two different epochs.

2.8.1 Preparation of the light curve

Kepler-68 has been observed in sectors 14 and 15, from July to September 2019. Also in this case, as it was true for Kepler, there are Simple Aperture Photometry (SAP) flux and Pre-search Data Conditioned SAP (PDCSAP) flux. The short cadence data are available every 2 minutes only for the primary targets of the mission, while every 30 minutes a Full Frame Image (FFI) is downloaded, containing much more stars. The target of this study is part of the selected stars for which the short cadence is available. Together with the time stamp and the flux each measurement comes with a quality flag. If it is zero it means that the mission control did not find any reason to exclude that point, while if it is a positive integer, it means that the mission control wanted to put a warning note on it. Different numbers have different meaning, in this data set the flags found are 4096, meaning *Argabrightening on CCD* and 32768, meaning *No fine point*. All the points with flag different from zero have been removed from the data set. In figure 2.11 the PDC flux without bad flags is reported.

Note that even the transits of planet b are not clearly visible by eye, as it was for the Kepler light curve. This is because Kepler fluxes are much more precise than the TESS' ones, resulting in more noise for the latter. Quantitatively, in figure 2.11, a scattering of $\pm 2‰$ from the average flux is present. The drawback of this fact is that more transits of the same planet are needed by TESS to actually see it.

2.8.2 Flattening and transit detection

In order to perform the flattening the Wōtan package is used, as described in a previous section (Hippke et al., 2019). The biweight method with a 1-day window is used (Tukey, 1977), together with a *break_tolerance* of 0.1 days. For the transit detection Transit Least Squares (TLS) is again the preferred choice. Here in figure 2.12 the periodogram of the TLS analysis is presented.

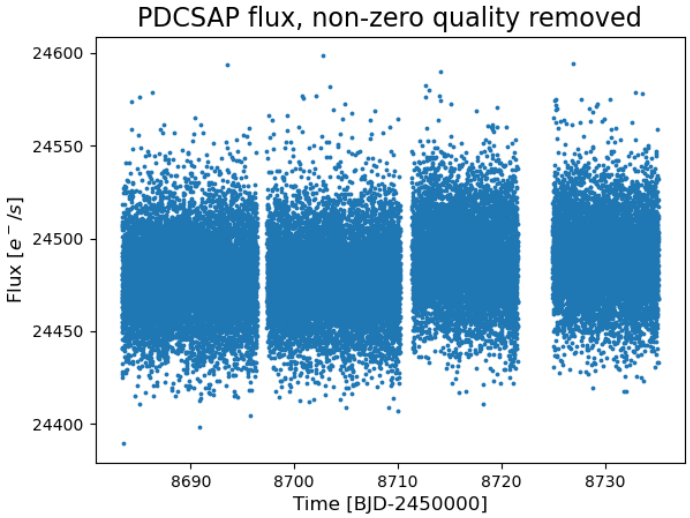

Figure 2.11: PDCSAP flux of sectors 14 and 15 of the star Kepler-68, where all the points with flags different from zero have been removed.

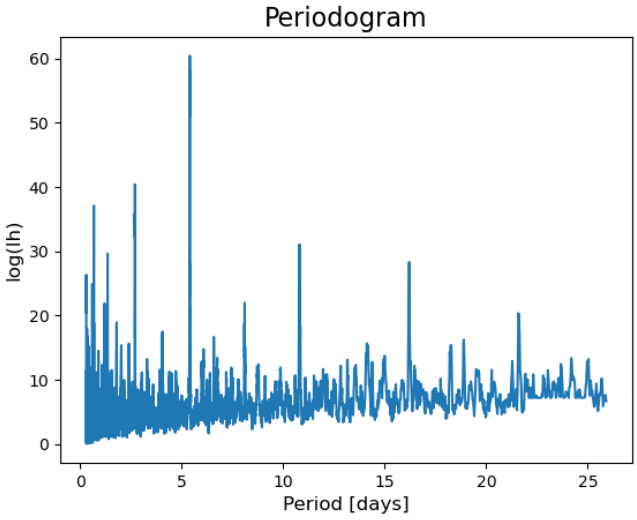

Figure 2.12: Periodogram of the TESS data done using TLS. The signal at 5.4 days corresponding to the first planet is showing up.

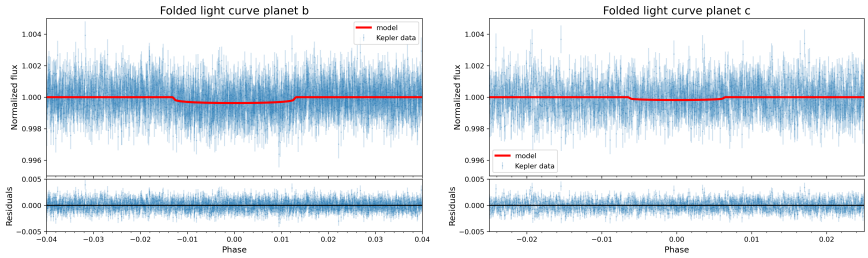

Figure 2.13: Left panel: folded light curve for planet b. Right panel: folded light curve for planet c.

It can be seen clearly a peak at 5.4 days with its harmonics. It correspond to the planet b, of course. Removing the transits of planet b by masking them, it is possible to perform another TLS analysis to search the transits of planet c. From the results of this analysis, no other transit signals are found. This is not unexpected since the depth of the transit of planet c (a bit more than 50 ppm) is much fainter than the noise in TESS data. So finally, only one planet out of two is recovered in TESS data. In the next subsection, the complete analysis of the light curve using PyORBIT is presented.

2.8.3 PyORBIT fits

Even if planet c has not been found by TLS, it is still included in the PyORBIT analysis. We took his decision because it is known that planet c is there from the Kepler analysis. So it is correct to include it, reserving the right to see from the posteriors if it is found by TESS too. In the configuration file are present large and uninformative boundaries on period and time of transit, taken from the TLS analysis for planet b and from the Kepler analysis for planet c. Moreover, from the Kepler analysis a prior on the impact parameter is imposed, since this quantity is intrinsic of the system and it is reasonable to assume that it is not changed significantly in a few years. Finally, the same prior on the eccentricity used for Kepler is used here too: a half-Gaussian of median zero and standard deviation 0.083 (Van Eylen et al., 2019). About the star, priors on the radius and the limb darkening coefficients are set. As a first approximation the limb darkening parameters are taken from the Kepler analysis, since the TESS data are too noisy for their direct determination and theoretical models are not reliable for this target.

The resulting posteriors found by the analysis are shown in the figure 2.13, for both planets b and c. As it can be seen, planet c is not identifiable from the scatter plot, nonetheless the algorithm has been able to constrain its period, radius and time of transit. As it is reported in figure 2.14, the corner plot containing the most important physical parameters of planet c, the posterior are not very sharp and symmetric but they outline clear results.

These results are summarized in table 2.10: they are consistent with the ones

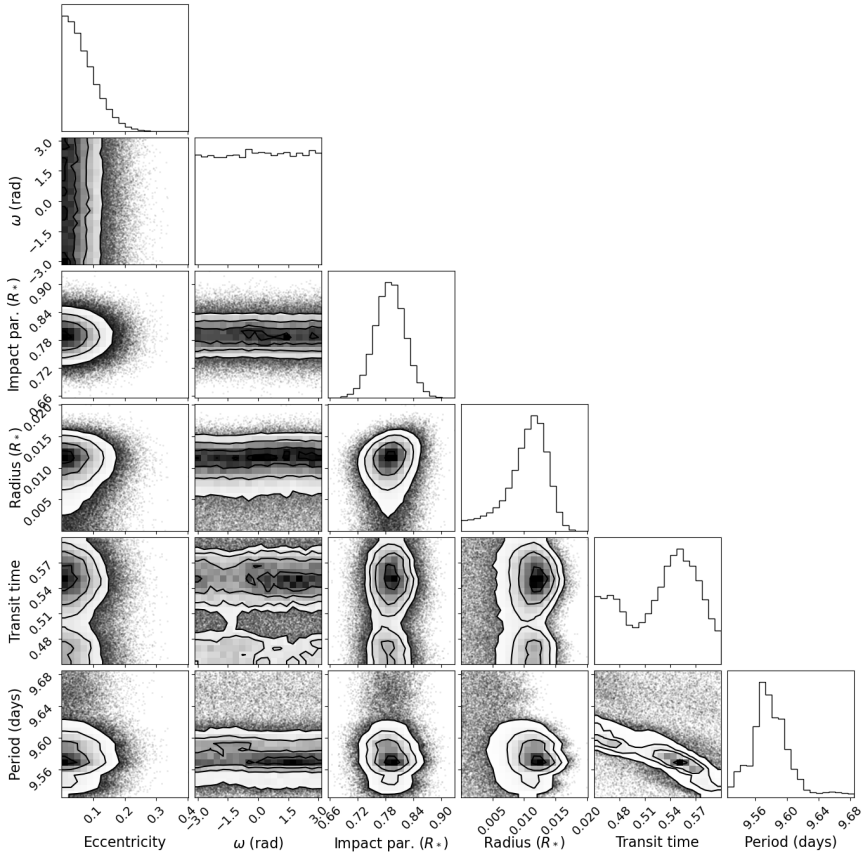

Figure 2.14: Corner plot of the posteriors of planet c. As it can be seen, they are peaked while showing broad wings that explore all, or almost all, the parameter space allowed. The transit time is in units of BJD-2,458,686 , only for displaying purposes.

Quantity	planet b	planet c
Period (days)	$5.39(86^{+18}_{-16})$	$9.5(77^{+24}_{-19})$
Radius (R_*)	$0.017(73^{+85}_{-87})$	$0.01(09^{+25}_{-38})$
Duration T_{14} (hours)	$3.481^{+0.106}_{-0.114}$	$3.00^{+0.18}_{-0.20}$
Transit time (BJD-2,450,000)	$8683.40(82^{+99}_{-96})$	$8686.5(37^{+33}_{-63})$
Semi-major axis (R_*)	$10.53^{+0.13}_{-0.14}$	15.44 ± 0.20

Table 2.10: Results for the planetary parameters from the fit of the TESS light curve. They are broadly consistent with Kepler results.

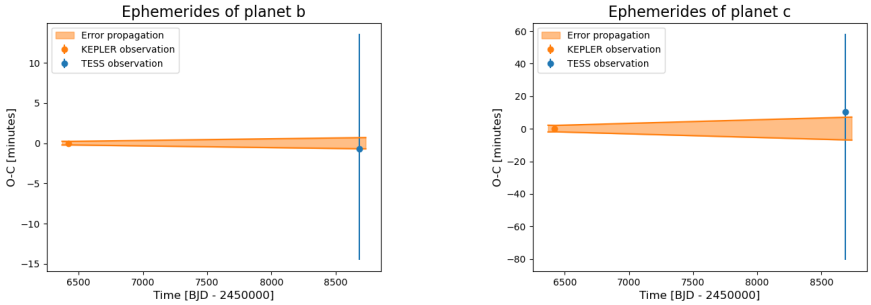

Figure 2.15: Left panel: ephemerides of planet b as described in the text. Right panel: ephemerides of planet c as described in the text. Note the different scale on the y axis of the two plots: for planet c all the errors are larger but the situation is almost identical with the TESS value at the edge of the Kepler prediction.

found with Kepler data. They clearly does not improve the knowledge of the Kepler-68 system that can be obtained from Kepler data. They are important because the data were obtained in a different epoch. So it is possible to compare the results and see if they changed. In particular, it is possible to see if the calculation of the time of transit made with TESS data matches the prediction made with Kepler data. This is matter for the next subsection.

2.8.4 Long-term transit time variations

Summarizing, we have a transit time for planet b derived from the Kepler time series and one derived from the TESS' one; the same holds for planet c. The aim is to calculate the ephemerides for both planets, in other words the prediction of the transit event at the TESS epoch made using Kepler transit time and period. And to see if the predicted transit matches with the TESS observation. Practically, it is enough to add to the transit time an integer number of periods until the absolute value of the difference between Kepler prediction and TESS observation is the least. It is needed to propagate properly the errors on the period too, to obtain a confidence interval at the new epoch. In the figure 2.15 this concept is visualized easily.

It can be seen that the Kepler prediction is abundantly in agreement with the TESS observation. Despite the error propagation, the transit time calculated with Kepler data is still more precise than the one observed by TESS. Thus, in conclusion, it can be stated that there are not visible trends in the transit time over a time span of about 10 years.

Chapter 3

Analysis of spectroscopic data

In this chapter a description of the steps performed to analyse spectroscopic data is presented. To the presentation of the instrument will follow the analysis of the stellar activity. Finally, we will present the fitting procedure and the model selection.

3.1 Introduction to spectroscopy

3.1.1 Spectroscopy

Spectroscopy is the study of electromagnetic radiation as a function of wavelength or frequency. When light interacts with matter by emission, absorption, transmission or else, it happens that different wavelengths act in different ways, leading to the formation of particular features in the so called spectrum, which is the representation of the amount of detected light as function of wavelength. These features can be absorpion lines, when the source at that wavelength looks dimmer than it is at slightly different wavelengths, or emission lines, when the source at that wavelength looks brighter than it is at similar wavelengths. Each spectrum is characterized by a resolution and a precision. The resolution R is given as the ratio $\Delta\lambda/\lambda$, where $\Delta\lambda$ is the separation between two consecutive wavelengths at which the spectrograph is able to measure the amount of light and λ is the wavelength at which R is calculated. The precision is the ability of a detector to measure a flux of photons. Greater is the precision, smaller is the error bar on the measurement. Most common instruments work near the optical range of wavelengths. Nowadays only a handful of instruments are able to overcome the 100'000 barrier of resolution.

Until today, the vast majority of spectrographs operate from ground. The technology needed to ship high resolution spectrographs in space is yet to be reached. The reason is that these instruments are relatively big and fragile, they need maintenance frequently.

3.1.2 HARPS-North spectrograph

HARPS stands for High Accuracy Radial velocity Planet Searcher. HARPS-N has been mounted at the 3.5 m Telescopio Nazionale Galileo (TNG) in 2012 (Cosentino et al., 2012). It is a twin of HARPS, which is located in the southern hemisphere, in La Silla, Chile. It has a wavelength coverage from 3830 Å to 6930 Å and a resolution of 115'000. To guarantee extremely stable efficiency it is sealed in a vacuum vessel with temperature and pressure control. It boasts the presence of octagonal fibers and Th-Ar simultaneous calibration (see next section). The precision of the instrument, in terms of radial velocities, can go below the 1 m/s barrier, as its twin HARPS does (Pepe, Ehrenreich, & Meyer, 2014).

3.1.3 HIRES spectrograph

HIRES, which stands for HIgh Resolution Echelle Spectrometer, is mounted at the Keck Telescopes, Hawaii since 1993 (Vogt et al., 1994). It has a slightly bigger wavelength coverage with respect to HARPS, ranging from about 0.3 μm to about 1 μm. Its maximum achievable resolution is 85'000, that is less than HARPS' but still very high. It is equipped with an iodine cell and it is not sealed in vacuum (see next section). The precision in radial velocity measurements is peaked around 1.5 m/s (Butler et al., 2017).

3.1.4 Kepler-68

Radial velocity data from HIRES spectrograph are publicly available at the Keck Observatory Archive (KOA), which is operated by the W. M. Keck Observatory and the NASA Exoplanet Science Institute (NExScI), under contract with the National Aeronautics and Space Administration. RV data from HARPS-N instead belong to the Consortium that operate the instrument and have kindly granted data access for this study. The data reduction pipeline used is the most updated one, the 2.3.1. Data from HIRES span a huge 8,25 years, from 2010 to 2018. Data from HARPS-N are more recent, spanning about 5.1 years, from 2016 to June 2021. The total time span is 11 years and 2 months; as a comparison, the orbital period of Jupiter is 11 years and 10 months.

3.2 Astrophysical noise

The signal produced with the RV method is a superposition of stellar activity and, eventually, a keplerian signal from an orbiting planet. Signals produced by stellar activity, also called jitters, are carrying a lot of information and there is a whole branch of astrophysics that is devoted to their study. In the case of exoplanetary astrophysics their main effect is to produce noise hiding the keplerian information. If the precision of $1ms^{-1}$ is to be reached, they must be carefully taken into account. In this section the most important sources of

jitter are reported, with a brief explanation (Lovis & Fischer, 2010, for all the subsections).

3.2.1 P-mode oscillations

Turbulent convection on the surface of stars can excite p-mode oscillations. One oscillation may have an amplitude of a few tens of cms^{-1}, but the total RV variation can reach several ms^{-1}, depending on the luminosity-to-mass ratio. Their amplitude also depends on the type of star: low-mass, non-evolved stars have lower p-mode amplitudes. Since the timescale is of the order of a few minutes, usually an exposure time of 15 minutes is enough to average out this noise below the ms^{-1} barrier.

3.2.2 Granulation and super-granulation

For stars having a convective envelope, the large-scale convective motions carry a photospheric signature: granulation. The typical velocity of these up and down movements is around $1 - 2\ kms^{-1}$. The convective motion in each cell in the envelope is uncorrelated with the others, therefore they average out and the overall signature is much smaller, more or less on the ms^{-1} scale. The time scale of these phenomena is about 10 minutes for stars like the sun. As for p-mode oscillations, small and late-type stars have a smaller signature.

On larger time scales, from hours up to one day, super-granulation may occur. Its origin is attributed to large-scale structures in the stellar photosphere, but there is still some debate around them.

3.2.3 Magnetic activity

Magnetic fields on the surface of stars generate inhomogeneities able to cause undesired RV signals. Basically speaking the inhomogeneities are dark spots and bright plages, covering part of the stellar disk and varying with time. They are carried from the blue-shifted part to the red-shifted part of the stellar disk by the stellar rotation. For this reason the time scale of their RV signal is of the order of several days on average. In young stars ($\lesssim 1\ Gyr$) this problem is more important, leading to RV variations greater than $10 - 100\ ms^{-1}$. What is even more significant is that the timescale of a few days is the same as the period of the closest hot Jupiters, making them undetectable.

When we talk about magnetic activity we should cite also the Zeeman effect. It is the effect of splitting of a spectral line into several sub-lines when a static magnetic field is turned on. Normally these lines are superimposed inside the main one, that is then affected by degeneracy, but magnetic field breaks it. This effect is as efficient as temperature changes in producing RV signals (Fischer et al., 2016).

3.2.4 Activity cycles

The Sun has an activity cycle of about 11 years where the number and the intensity of spots and plages oscillate periodically. One way to identify these cycles is to measure *Ca II H and K* lines. There may also be systematic trends that will affect the reliability of RV measurements over the years on the same star.

3.3 Instrumental noise

Real-world instruments are not perfect and it is very important to know how they work, to be able to correct all the bad behaviours. This section contains a description of the main instrumental noises that has to be considered if the ms^{-1} precision is to be achieved. But before going deep into this list, it is worth to introduce the two main options for the wavelength calibration in use until today: Iodine cells and Thorium-Argon lamps. For all the section, see Lovis and Fischer (2010)

3.3.1 Iodine cells

An iodine cell is a transparent box containing iodine gas positioned on the optical path of the light coming from the telescope, before the slit. It is introducing a forest of molecular absorption lines on the top of the stellar spectrum. The idea is that those lines, that are well known and stable, provide a calibration reference for the *science lines* of the star.

To achieve the desired level of precision needed for 1 ms^{-1} RV, particular care must be used in the building process. The cell must be perfectly sealed to prevent loss of I_2 gas and some chemical treatments are needed to ensure the lack of other chemical species inside it, that would add some spurious lines. Apart from the unwanted absorption lines, the presence of water inside may end up in a decreasing transparency of the cell. In a spectrograph it is very important that as more light as possible is reaching the CCD and even the smallest light loss can make the difference on the fainter stars. For this reason also an anti-reflective coating is applied on the flat sides of the cell, the ones perpendicular to the light path. Even with all these attentions, the resulting opacity is responsible of about 15% light loss, in the best cases. Moreover when in use it must be maintained at a constant temperature greater than $37°C$. This is to ensure that all the iodine is in the gaseous state and the column density is constant during the night.

If all the possible cares are applied, the iodine cell can provide a stable and reliable reference for years, just what is needed for high-precision RVs. One other advantage is that the reference is falling in the same region of the CCD of the spectrum of the observed star. This avoids instrumental systematic errors between the reference and the spectrum. The main disadvantage of iodine cells is their limited wavelength coverage: it ranges only from 5000 Å to 6200 Å. In this range, since the absorption features of iodine are superimposed with the

stellar spectrum, a work of careful deconvolution is needed to separate the two spectra.

3.3.2 Thorium-Argon lamps

The Th-Ar lamp is providing a simultaneous reference for the stellar spectrum, consisting in the well known and characterized absorption features of these two chemical elements. In this configuration, there are two channels entering the spectrograph; the science one and the reference one. At the beginning of the night the light from the lamp is injected in both the channels to provide a solid calibration. Then, the lamp is kept in the reference channel to keep track of the instrumental drifts that, in principle, affect in the same way the two channels. In this way a reliable reference is built, solving the problems of air refraction index variations and mechanical flexures affecting both the channels.

Slit illumination is the next problem to solve. In the past decades it became clear that the telescope guidance alone wasn't enough to guarantee the desired precision in the illumination. They opted for the usage of optical fibers to carry the PSF from the focal plane of the telescope to the slit. This method allows to have a steady illumination of the slit, moreover the spectrograph can be moved from the telescope dome to an adjacent building.

For the achievement of the ms^{-1} precision and beyond, however a number of cares must be applied even in the presence of optical fibers. It has been empirically observed that for circular-section fibers the angle with which the light exits is not independent from the angle with which it enters. So if the PSF is not exactly in the center of the fiber the output is influenced and spurious RV signals are introduced. to correct this problem octagonal-section fibers are introduced, because experimentally it has been observed that they allow a better scrambling of the light in the output. Moreover, a mechanism to keep the PSF centered on the fiber is needed. The most in use is a mirror with a tiny hole in the middle: the light that pass through the hole enters the fiber, while the tail of the PSF is reflected back to a camera. Based on what the camera sees, the fiber-mirror system can be adjusted to keep the PSF always centered.

One advantage of the Th-Ar lamp with respect to the iodine cell is the wider wavelength coverage that spans the $3800 - 6900$ Å range (Lovis & Pepe, 2007). One disadvantage on the long period is the natural decline of the cathode due to the normal usage of the lamp. In time the position of the lines may vary and their broadening too. It is therefore important to model these behaviours using theoretical and empirical knowledge.

To practically measure the radial velocity shifts on the calibrated spectrum a variant of the cross-correlation method is used. The basic idea is that the shift is so small that it cannot be measured comparing one line from the reference and one from the science. For this reason binary masks are used, i.e. a transmission function containing only zeros and ones, where the ones are in correspondence with stellar lines in the rest frame. By doing the cross-correlation between the mask and the stellar spectrum it is possible to gather all the Doppler information in one function. From a simple fit of this function, the overall RV value is

extracted, for each observation.

3.3.3 Air refraction index variations

If the temperature and the pressure inside the instrument are not extremely stable, there may be little changes in the refraction index of the air. As a direct consequence the spectrum moves on the detector and this translates into wavelength shifts of up to several hundreds of ms^{-1}. For this reason it is important to maintain the instrument under extremely stable conditions: even a variation of 0.01 K or 0.01 $mbar$ are sufficient to induce a shift of 1 ms^{-1}.

3.3.4 Mechanical flexures

The multitude of mechanical and optical parts that compose a spectrograph can heat up, inducing mechanical flexures and changes in the optical path. This translates directly into wavelength drifts on the detector. Their amplitude can reach the order of hundreds of ms^{-1}.

3.3.5 Slit illumination variations

The functioning of a spectrograph is based on the dispersion of the image of the slit on the detector. The slit is illuminated by the point spread function (PSF), therefore a change in the position and/or the diameter of the PSF is sufficient to cause drifts along the wavelength axis on the detector. These kind of variations could come from change in seeing, airmass, telescope guiding, telescope focus, vignetting, etc. In this way, shifts of tens of ms^{-1} on the measured RV are easily produced.

3.3.6 Wavelength calibration

It is important to have the wavelength axis very well calibrated. In order to do so a well-characterized reference provided with a dense forest of absorption lines is needed, that will be stable and reliable for years. The most successful instruments used for this purpose until today are the iodine cell and the Thorium-Argon lamp. Both instruments have their own pros and cons, as it has been explained above. In any case, on a single night of observations this is not the main source of error if we aim to the ms^{-1} precision. On the long-term time scale instead they both have demonstrated a stability at the $1-2$ ms^{-1} level.

3.3.7 CCD imperfections

CCD imperfections are not introducing additional RV drifts, they are decreasing the precision of the measurement instead. The main parameter of a pixel is its sensitivity. It can vary according to the pixel size and intrinsic sensitivity (i.e. how many photons are captured on the total hitting the pixel, inside the functioning wavelength range). A possible solution would be to scan

and therefore map the CCD to know exactly the dimensions and sensitivities of all the pixels. An even worse problem arises in presence of thermal changes: The pixels size and properties may vary. But even if the pixels are all perfectly characterized, there can be issues in the charge transfer efficiency. During CCD readout a small fraction of the total collected charge is inevitably lost. This effect is greater for pixels far away from the readout row, so being a differential issue it is introducing an unwanted shift in the detected radial velocity.

3.3.8 Spectrum contamination

The recorded spectrum of the star do not come uncontaminated. Mainly caused by telluric lines from the Earth atmosphere, there can be spurious radial velocity signals. Where it is possible, telluric lines must considered and properly detrended from the spectrum. Where they are too dense and deep, especially in the red and near-IR, it is better to reject completely a wavelength interval around them. Since they are naturally steady in the reference frame of the Earth, they are moving back and forth with a periodicity of one year in the reference of the barycentre of the Solar System. This trend can be removed but still one must be careful with the possible variability of telluric lines from night to night due to atmospheric instabilities, that in the perspective of the ms^{-1} precision, are important as well.

Another possible source of noise in the spectrum is moonlight. It carries the signature of Sun's absorption lines thanks to the reflected light. Finally, in some cases, also the spectrum of background or foreground stars can enter the slit and thus pollute the recorded spectrum on the CCD. This leads to other spurious RV noises.

3.4 Stellar activity

From the spectroscopic point of view the stellar activity is studied through some quantities called activity and asymmetry indices. Activity indices like Calcium H & K, Sodium doublet or Hydrogen α indices are photon fluxes arbitrarily normalized. For what concerns asymmetry indices, they are calculated over the cross-correlation function of the spectrum with a mask of choice, depending on which lines are to be used. The most widespread are bisector span, full width at half maximum and contrast. In this section we analyse in great detail the stellar activity using these indices and indicators.

3.4.1 Activity indicators

Activity can be inferred by measuring the amount of radiation emitted in certain particular spectral lines. Commonly the flux inside the given line is measured, with a narrow filter of predetermined width, then the flux on both sides of the line is measured, defining the continuum region, with other two narrow bandpass. Finally a combination of those measures gives the index. In

particular it is worth to go into details of the indices retrievable from the two Calcium lines H & K. The S index is calculated in the following way:

$$S = a_j \frac{N_H + N_K}{N_V + N_R}$$

where a_j is a normalizing constant depending on the instrument; N_H is the number of counts at the center of the H line using a band 1 Å wide; N_K is the same for line K; N_V is the number of counts in a band 25 Å wide at slightly shorter wavelengths with respect to K line (which is at 3934 Å); N_R is the same but at slightly longer wavelengths with respect to H line (which is at 3969 Å). Moreover, the centers of the two reference bands V and R are 250 Å apart to each other, in a symmetric position with respect to the two H & K Calcium lines. All the counts are corrected for the background (Vaughan, Preston, & Wilson, 1978).

Another index, probably more used nowadays, is the $log(R'_{HK})$ index. It is described in Noyes, Hartmann, Baliunas, Duncan, and Vaughan (1984) as follows.

$$R_{HK} = \frac{N_H + N_K}{N_{TOT}} = \frac{F_{HK}}{\sigma T_{eff}^4}$$

where N_H and N_K are the same as before and N_{TOT} are the total bolometric counts of the star; F_{HK} is the total flux per cm^2 at the stellar surface in the H & K bands; σ is the Stefan-Boltzmann constant and T_{eff} is the effective temperature. Contribution to N_H and N_K comes from both the chromosphere and the photosphere, so the idea of Noyes and collaborators was to estimate the R index of the photosphere alone by measuring the flux in the wings of the H & K lines where the chromospheric contribution is negligible.

$$R_{phot} = \frac{F_{outside}}{\sigma T_{eff}^4}$$

where $F_{outside}$ is the flux registered in the wings of the H & K lines, summed. In this way, they obtained the R index of the chromosphere alone by subtraction:

$$R'_{HK} = R_{HK} - R_{phot}$$

and for the sake of convenience they took the logarithm of it, reaching the final $log(R'_{HK})$.

3.4.2 Asymmetry indices

As it has been anticipated, proxies of the stellar activity can be the shape of the stellar cross-correlation function, changing in time. The full width at half maximum (FWHM), as its name suggests, is the width of a line, in km/s, measured at one half of the total depth of the line itself. The contrast itself, is the depth of a line measured from the continuum to the valley of an absorption line. It is a unit-less number included between zero (no line at all) and one or

Index	Best period (days)	FAP*	FAP**
Calcium H & K	230.5 ± 8.3	0.010	/
Hydrogen α	889 ± 77	0.001	0.0019
Sodium doublet	387 ± 14	0.000	0.0000
Contrast	211.0 ± 3.4	0.000	0.0000
FWHM	$1667 \pm nan$	0.076	/
BIS	22.328 ± 0.058	0.208	/

Table 3.1: Best sine period found by GLS, followed by the corresponding false alarm probability calculated using bootstraps. * = calculated using 1'000 bootstraps. ** = calculated using 20'000 bootstraps.

sometimes one hundred (saturated line). The so called bisector span or bisector inverse span (BIS), often measured in meters per second on the wavelength axis, is the difference between the baricenter of a line measured at two different heights. If a line, or the cross-correlation function of an entire spectrum, is asymmetric, it could be that its baricenter measured near the continuum falls at a different wavelength with respect to the baricenter measured near the bottom of the (absorpion) line. The difference of the two wavelengths, translated in meters per second, is the BIS.

3.4.3 GLS analysis

In order to see if there is some periodicity in the temporal evolution of the indices presented before, it is useful to perform a GLS analysis. GLS is a code that has already been presented: it is able to try the fit of a sinusoidal function and to build a power spectrum from which it is possible to see the best fitting period. It is computationally cheap but it does not allow eccentric curves. In the table 3.1 the best sine period is reported with the relative error. But this is useless without a measure of how significant the peak in the power spectrum is. For this reason we have decided to calculate a false alarm probability (FAP). Through the bootstrapping method it has been possible to generate thousands of random signals and to apply the GLS algorithm to each of them. In this framework the FAP is the number of bootstraps that give a peak higher than the original signal, divided by the total number of bootstraps. Higher is the FAP, higher is the probability that the periodicity found in the original signal is not significant, it is just a fruit of chance instead.

The last column is the result of a further study. Contrast, Hydrogen and Sodium showed up with a very low FAP from the first set of bootstraps. For this reason we decided to calculate much more bootstraps to have a better statistics. They resulted in an equally low FAP, meaning that from this test they resulted as significant. Since from the analysis of photometry it was not possible to see any activity, it was quite peculiar to see it in spectroscopic indices. On top of this, the periods revealed are all different one from the other. To study more into details the phenomena we decided to perform a Bayesian analysis on the

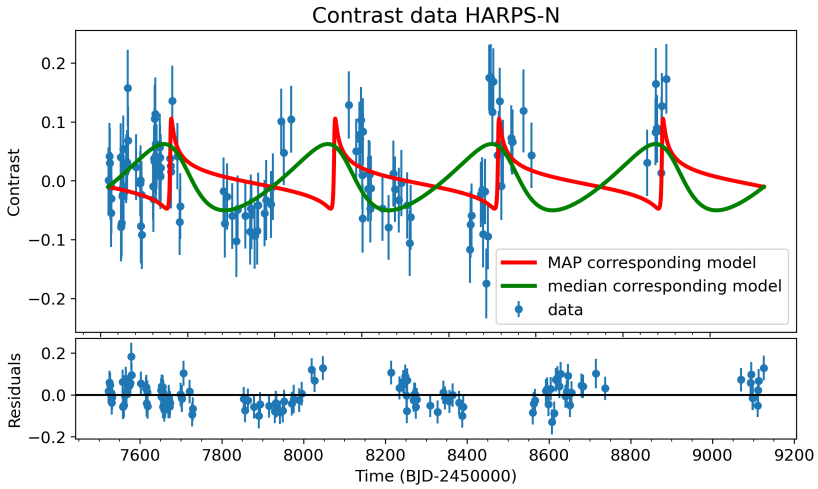

Figure 3.1: Example of fitting model with residuals. The data represent the contrast of the spectra measured by HARPS-N.

previously mentioned three indices.

3.4.4 Bayesian analysis

While preparing the data it was possible to notice some outliers. From a preliminary PyORBIT analysis on all the three indices it has been possible to see they had a huge impact on the search for the best fit: the code was always increasing the eccentricity to try to include them, landing on nonphysical conclusions. For this reason we removed the outliers by hand and performed a longer analysis. We removed three points from the contrast, two from Hydrogen and four from Sodium, for a total of six points, because three were in common between two indices. To perform the fit we used the keplerian model of radial velocities. In this case though, quantities as semi-major axis or mass of the planet have no meaning. The focus stayed on the period and the semi-amplitude, while always keeping an eye on the jitter. In figure 3.1 the fit of the contrast is shown as an example, together with the residuals. As we can see, the fitting model does not match the data and thus the scattering in the residuals is not reduced with respect to the data themselves. In table 3.2 the main quantities are shown for contrast, Hydrogen index and Sodium index.

The majority of the results reported are misleading if not supported by the relative plots. In figure 3.2 is shown a corner plot for the Hydrogen index. Now it is clear that for this index, period and eccentricity posteriors are less strong than they seemed. Moreover, it is worth to notice that the semi-amplitudes found are always below the measured jitter, meaning that the results are not strong. Hence the hypothesis of very low stellar activity is corroborated. For those reasons we decided to ignore stellar activity in the following treatment.

Index	Period (days)	Semi-amplitude	Jitter
Contrast	471 ± 17	0.056 ± 0.011	0.0578 ± 0.0046
Hydrogen α	460^{+600}_{-460}	0.0013	0.00229 ± 0.00023
Sodium doublet	850 ± 440	0.0032 ± 0.0018	0.00333 ± 0.00025

Table 3.2: Results of the PyORBIT fits for the quantities that gave a very low false alarm probability in the GLS test. The units of semi-amplitude and jitter are the same and are arbitrary. For the semi-amplitude of the Hydrogen index an upper limit at 68.27% is reported.

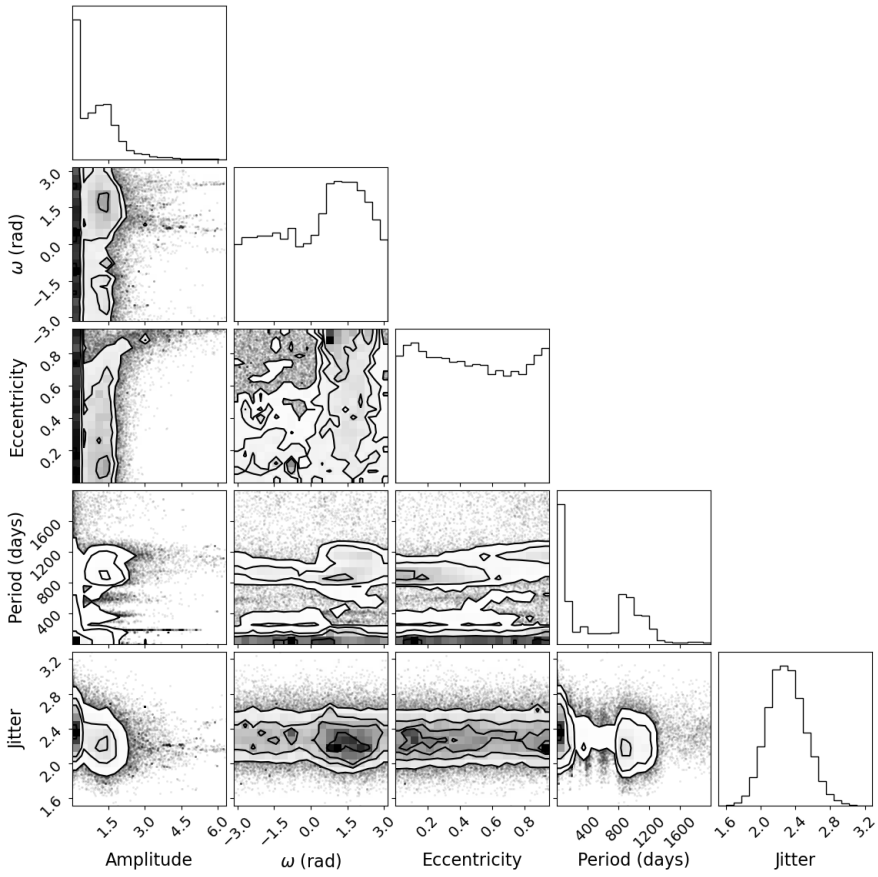

Figure 3.2: Example of corner plot for the Hydrogen index. It is clearly visible that the posteriors are not clean nor even uni-modal in some cases. The units of semi-amplitude and jitter are the same and are arbitrary.

Quantity	Planet b	Planet c	Planet d
Period (days)	$5.398751(96 \pm 79)$*	$9.6050(43 \pm 15)$*	632.1 ± 1.4
Time of transit (BJD-2,450,000)	$5066.244(94 \pm 15)$*	$5065.42(91 \pm 14)$*	8420.9 ± 5.6
Eccentricity	0.093**	0.35**	0.164 ± 0.019
Semi-ampl. (m/s)	2.84 ± 0.31	0.10**	18.12 ± 0.28
Mass (M_{Earth})	8.06 ± 0.87	0.34**	249.0 ± 5.8***

Table 3.3: Results of the PyORBT analysis comprehensive of Kepler-68 b, c and d.
* = This quantity was constrained by a prior coming from photometry.
** = This is an upper bound: 68.27% of the values are lower than this.
*** = This is a minimum mass, since the inclination is unknown.

3.5 Bayesian analysis of radial velocities

In this section the fitting procedure of the radial velocity data coming from HIRES and HARPS-N is presented. Starting from the knowledge present in the literature, the aim is to take advantage of the newly available data to see if there is more to be discovered.

3.5.1 Fit with three planets

Since 2013 Kepler-68 is known to host a third non-transiting planet: Kepler-68d (Gilliland et al., 2013). To confirm the presence of this planet also in the new set of data a PyORBIT fit is performed. It is well justified to use priors on orbital period and time of transit coming from photometry of the two inner planets. Moreover, it was safe to fix the inclination of the same planets and to give a prior to the stellar mass, since these values are only used to convert the radial velocity semi-amplitudes to physical masses. Finally, we impose some non-informative boundaries on the semi-amplitude and the eccentricity of planets b and c. No priors on planet d are set, the algorithm was free to explore all the physically allowed parameter space. The main results of the analysis are shown in the table 3.3.

To fit properly a radial velocity dataset, planetary parameters are not enough. There is the need to take into account an overall offset and a jitter term, that is the stellar activity-induced signal in radial velocity measurements. It is not enough to average data to zero before doing the fit: the offset can be significantly different from the average of the data. As an example, an unlucky sampling would give back all the points near the minima of the leading oscillating keplerian signal. The average would be smaller than the needed offset to correctly fit the data. Actually, since in this case there are two data sets (HIRES and HARPS-N) the offsets are two and the jitter terms too. HIRES' offset resulted in 1.26 ± 0.44 m/s and the jitter 3.37 ± 0.34 m/s. For HARPS-N, the offset was -20861.13 ± 0.30 m/s and the jitter 2.63 ± 0.22 m/s. Note that HIRES data come with an almost-zero average while HARPS-N's ones come as they exit the data reduction pipeline, with an high absolute value. From the algorithmic point

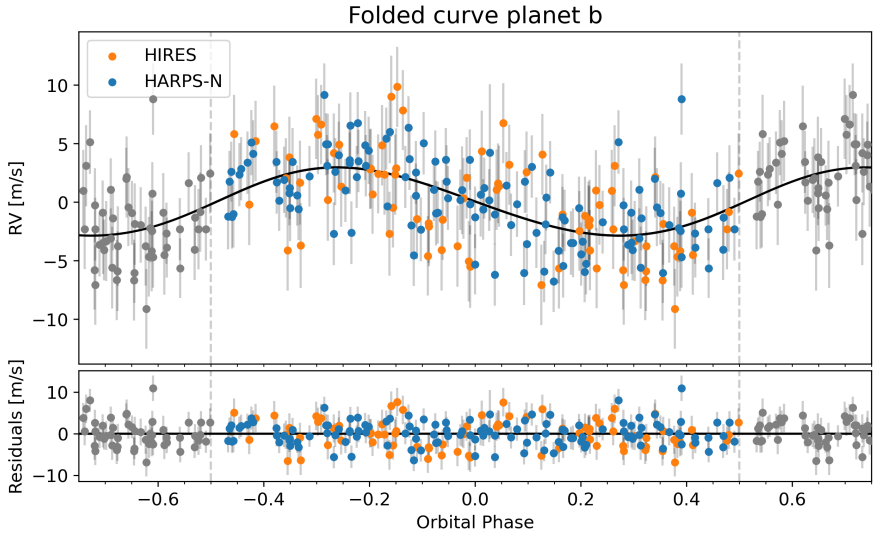

Figure 3.3: Folded radial velocity curve for planet b.

of view this is not a problem at all.

A useful way to visualize the results is to plot the data folded with the resulting periods and to see if the keplerian signal shows up. Of course to see it clearly the signals of the other two planets must be subtracted. In figures 3.3 and 3.4 we show the folded RV curves for planet b and d. Planet c is not shown since its semi-amplitude is consistent with zero thus it can not be appreciated. It is interesting to notice the evident non-sinusoidal shape of the signal of planet d due to its eccentricity.

In order to see if the fit was comprehensive of all the information contained in the data, it is useful to have a look to the residuals, showed in figure 3.5. In the residuals a further periodic signal is probably coming to the surface. In the same plot we superimposed the best-period sinusoidal signal found with a GLS analysis of the residuals themselves. The period is around 1600 days, or 4.4 years, it would be a new planet with an orbit intermediate between the one of Mars and Jupiter. It is clear that there is more in these data than it is known from literature. For this reason, we decided to try to fit different models with different number of planets and to see what is the best in fitting the data. This goal is accomplished using Bayesian evidence. In the next subsection we give an introduction to it and to the code used, then the model selection is presented below.

3.5.2 Dynesty algorithm

Nested Sampling (NS) algorithms are able to generate samples from the posteriors in order to visualize them, just like MCMC-based algorithms, but

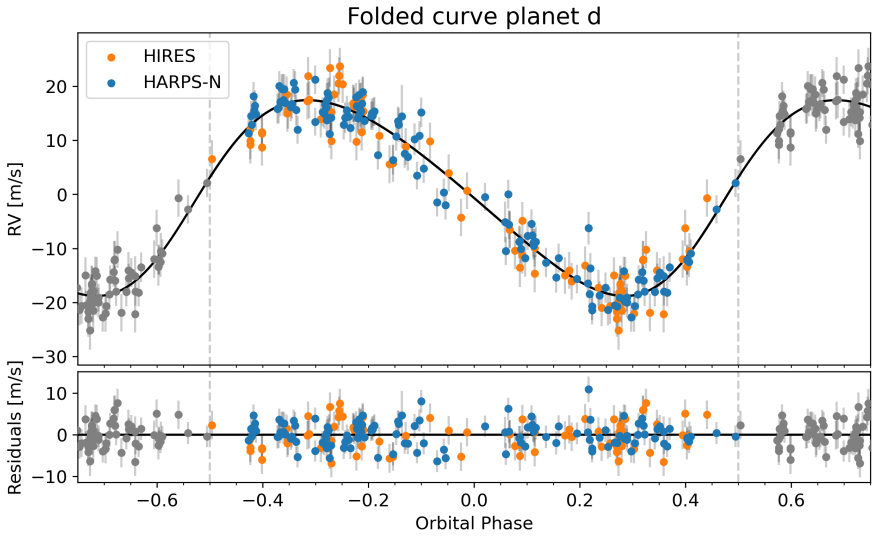

Figure 3.4: Folded radial velocity curve for planet d.

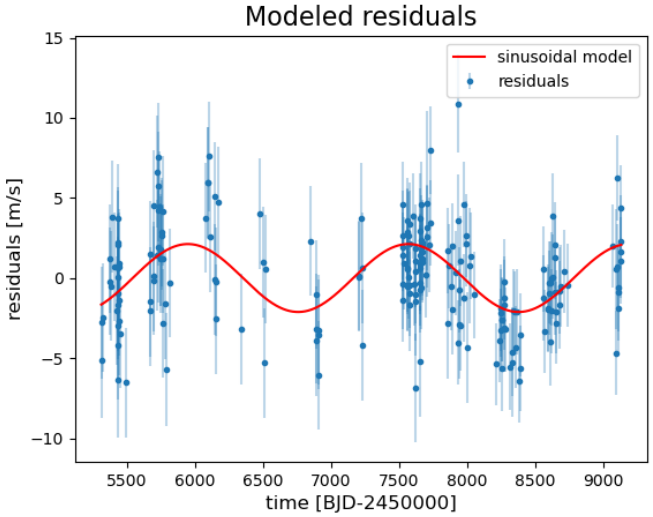

Figure 3.5: Residuals of the described fit with three planets, an offset and a jitter term per instrument.

they are mainly used to compare different models applied to the same dataset. They are based on the application of the Bayes' theorem to two or more different models. Bayes' theorem applied to two models states that

$$P(M_1|D) = \frac{P(D|M_1) \cdot P(M_1)}{P(D)}$$

Where $P(M_1|D)$ is the conditional probability of model M_1 being true given data D and $P(D|M_1)$ is the conditional probability of obtaining data D given that the true model is M_1. $P(M_1)$ is the probability of model M_1 being favoured over M_2 and $P(D)$ is the probability of having data D. The same holds for M_2. The denominator, $P(D)$, can be rewritten as $P(D) = P(D|M_1) \cdot P(M_1) + P(D|M_2) \cdot P(M_2)$ by definition. So the first formula becomes

$$P(M_1|D) = \frac{1}{1 + \frac{P(D|M_2) \cdot P(M_2)}{P(D|M_1) \cdot P(M_1)}}$$

Since a priori in many cases there is no reason to prefer one model on the other, their probabilities $P(M_1)$ and $P(M_2)$ can be replaced with one half each, cancelling out from the last formula. The quantity

$$BF = \frac{P(D|M_2)}{P(D|M_1)}$$

is the so called Bayes factor, used to compare two models on the same dataset. The probabilities in play represents the capacity of model i to explain the data D, so their ratio express the predominance of one on the other. These probabilities can be written as an integral:

$$P(D|M_1) = \int P(D|\theta, M_1) \cdot P(\theta|M_1) d\theta$$

where θ is the set of parameters used in the model. And likewise for M_2. This integral is the fully marginalized likelihood, often called $\mathcal{Z}$, and it is analytically unsolvable in almost all cases. For this reason there is the need of a numerical approach. The first algorithmic implementation able to solve this integral was developed in Skilling (2004). This algorithm has different Python-based implementations, among which PyORBIT supports Dynesty (Speagle, 2020) and UltraNest (Buchner, 2021).

Dynesty stands for dynamic Nested Sampling, it is a more complex version of NS where the number of samples in the parameter space is not constant, it is adjusted step by step instead, to guarantee the best accuracy (Speagle, 2020). It relies on live points, which are allocated in the parameter space in a given amount. This amount can vary from step to step, basing on the need of the code: it is the main implementation of the dynamic code with respect to the basic NS. The integral $P(D|M_1)$ showed above is the Bayesian evidence that we will use in this work for model selection (Nelson et al., 2020).

Planets' number	Trend's order	Bayesian evidence
3	0^{th}	-48.079
3	1^{st}	-36.115
3	2^{nd}	-36.013
4	0^{th}	-7.283
4	1^{st}	0.000
4	2^{nd}	-0.068
5	0^{th}	-10.419
5	1^{st}	-1.395
5	2^{nd}	-1.217

Table 3.4: Bayesian evidence of all the analyses performed, listed in complexity order. Since the absolute value of the evidence is meaningless (it varies changing the dataset), we opted for the subtraction of the biggest one, for an easier comparison. Its value is -491.718

3.5.3 Model Selection

It is known that this system hosts at least three planets, so it would be a waste of computational time to investigate models that contain less than three planets. There are strong evidences that a fourth planet is there, though. A priori the model with four planets is the most promising, while the presence of an hidden fifth planet is not to be excluded. So we decided to test models with three, four or five planets and see which performs better, in a Bayesian sense. Quite often this kind of data sets are better fitted if a trend is added to the keplerian signals of whatever number of planets. So it is useful to test every combination of planets in three different flavours: with the offset only (zeroth order trend), with a linear trend (first order) or a quadratic one (second order). As Bayesian evidence it is used $\mathcal{Z}$, which is the fully marginalized likelihood presented above. Its value for all the analyses is reported in table 3.4.

As it can be seen, the favoured model is the one with four planets plus a first order trend. There are other models that are indistinguishable from this one though, from a Bayesian point of view. If the difference between two Bayesian evidences is less than three or four, the two models are fitting the data with a very similar level of confidence and it is impossible to choose between them using only this information (Feroz, Balan, & Hobson, 2011). In this sense, models with 4 planets and a quadratic term, together with models with 5 planets with a non-zero order trend, are good too. Basing the argument on the Occam's razor principle, it should be fair to prefer models with four planets instead of five. It is quite obvious that a model with more planets will leave less residuals, because of its greater number of parameters. For this reason, a difference greater than four in the Bayesian evidence is needed to actually prefer the more complex model.

Regarding the discrimination between linear and quadratic trends, it is useful to visualise them. The first-order coefficient of the analysis with linear trend is $a_{lin} = -1.26 \pm 0.34\ mm/s/day$. The first-order coefficient of the analysis

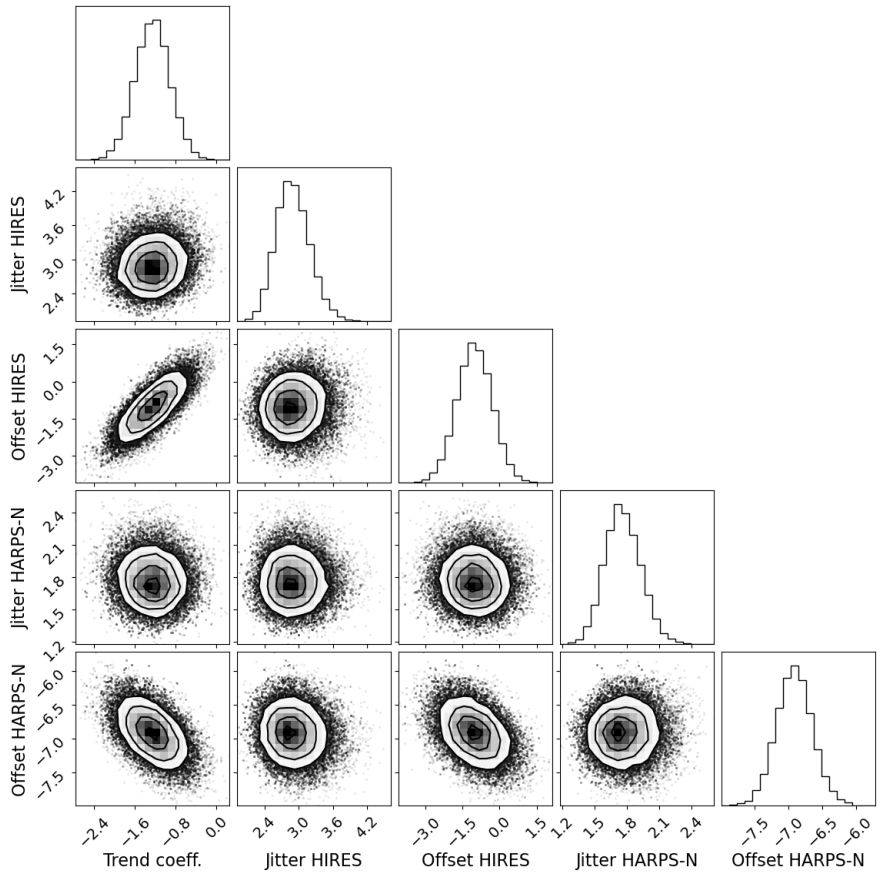

Figure 3.6: Corner plot of the parameters of the selected model. All the units of measure are meters per second (m/s), apart from the trend coefficient that is in millimeters per second per day (mm/s/day). HARPS-N offset has an additional $-20900 m/s$ constant that has been subtracted for displaying purposes.

with quadratic trend is $a_{quad} = -1.26^{+0.38}_{-0.36}\ mm/s/day$, while the second-order coefficient is $b_{quad} = (0.02^{+2.27}_{-2.20})10^{-4}\ mm/s/day^2$. As it is evident, the first-order terms are the same and the second-order one is extensively consistent with zero. We can state that, with this dataset, it is impossible to infer the presence of a quadratic trend. Hence, the selected model is the one with four planets and a linear trend, which comprises an offset of course. In figure 3.6 we present the corner plot with the most important parameters that led the model selection.

The period found for the additional candidate planet, that will be called Kepler-68 e hereafter, is around 1550 ± 40 days, in agreement with the preliminary analysis of the residuals (figure 3.5). The analysis of radial velocity data is concluded. In the next chapter, all the pieces of the puzzle will come together and we will show all the physical parameters of the four planets.

Chapter 4

All-inclusive analysis

In this chapter we will present the fit containing both photometric and spectroscopic data. All the information are combined, to pull out a final estimate of all the physical quantities.

4.1 Preparation of the fit

The configuration file is quite complex since we have to put together photometry and spectroscopy. Here in the following the priors used are presented and explained, in order to justify our choices before presenting the results. See table 4.1 for a schematic summary.

Priors on the period and the central time of transit of planets b and c are chosen based on the TLS analysis. This analysis gave the period, the T_c and the duration of the transit as outputs. Very naively, the true period cannot be greater than the TLS period plus the transit duration and for the same reason it cannot be less than the TLS period minus the transit duration, otherwise the next transit time would fall outside the transit shape. A similar line of reasoning is done for the central time of transit: it can not be greater than the TLS central time plus half of the duration or less than TLS central time minus half the duration, otherwise it would imply that the TLS central time falls outside the transit signature, that is absurd. Uniform priors on eccentricities of all the planets are just meant to restrict the physical space of parameters, knowing that there are not high-eccentricity planets in this system. The semi-amplitude space is explored logarithmically for all the planets. The boundaries are chosen based on the results of the previous model selection, leaving them still quite large. The idea is that they have to be uninformative, just as all the other uniform prior we gave. It means that they must not influence the final result, they are not meant to be constraints. Their goal is to restrict the physical parameter space such that the algorithm takes less time to converge. The half-Gaussian prior on the eccentricity of planet c is taken from a previous study on multiple planetary systems in Kepler dataset (Van Eylen et al., 2019). This study was arguing, using

Quantity	prior type	1st value	2nd value
Lin. coeff. (m/s/day)	$\mathcal{U}$	-1.0	1.0
Planet b			
Period (days)	$\mathcal{U}$	5.3	5.5
T_c (BJD-2,450,000)	$\mathcal{U}$	5066.2	5066.3
Eccentricity	$\mathcal{U}$	0.0	0.5
Semi-amplitude (m/s)	$\mathcal{U}$	0.01	10.0
Planet c			
Period (days)	$\mathcal{U}$	9.53	9.68
T_c (BJD-2,450,000)	$\mathcal{U}$	5065.40	5065.47
Eccentricity	$\mathcal{U}$	0.0	0.5
Eccentricity	$\mathcal{HN}$	0.0	0.083
Semi-amplitude (m/s)	$\mathcal{U}$	0.001	5.0
Planet d			
Period (days)	$\mathcal{U}$	530.	730.
Eccentricity	$\mathcal{U}$	0.0	0.5
Semi-amplitude (m/s)	$\mathcal{U}$	1.0	40.0
Planet e			
Period (days)	$\mathcal{U}$	1000.	2000.
Eccentricity	$\mathcal{U}$	0.0	0.5
Semi-amplitude (m/s)	$\mathcal{U}$	0.1	20.0
Star			
Radius ($R_\odot$)	$\mathcal{N}$	1.2564	0.0084
Mass ($M_\odot$)	$\mathcal{N}$	1.057	0.029
Density ($\rho_\odot$)	$\mathcal{N}$	0.534	0.022

Table 4.1: Priors used in the fit. $\mathcal{U}$ stands for uniform prior, $\mathcal{N}$ stands for Gaussian prior and $\mathcal{HN}$ stands for half-Gaussian prior. In the case of uniform prior, the two values reported are the two extremes of the interval; in the case of Gaussian or half-Gaussian prior, the two values are median and standard deviation. T_c is the central time of transit.

Kepler observations available, that if a planet is in a close multi-planetary system it is likely to have a low eccentricity. We decided to hold it because planet c has a very low RV signal and in doing this we prevented the algorithm to get stuck on very high eccentricity nonphysical orbits driven by some outliers. For other planets this was not necessary since the RV modulation has a signal-to-noise ratio large enough to avoid this problem without further constraints. Finally, for what concerns planets, the uniform priors on the periods of planets d and e are taken from the results of the model selection, just as semi-amplitudes. Their boundaries are very large, more than ten times the error found with the Nested Sampling analysis. This is again to assure that they are completely uninformative.

For what concern the star, we constrained the mass, the radius and the density with Gaussian priors using the results of the isochrones analysis presented in chapter 2. We decided to not constrain the limb darkening coefficients since from the fit of photometry, in the same chapter, we found values that were not completely in agreement with the theoretical ones. We therefore decided to trust more the data and do not use the theoretical results, hence leaving the coefficients free. Finally, the flat prior on the linear coefficient of the trend has been chosen to vary between -1 and 1, being those values much larger, in an absolute sense, than what could be even physically explainable.

An emcee analysis is performed, using PyDE as solver (see chapter 2). 200'000 steps are used, with a burn-in after 50'000 steps. The thinning factor is set to 100. In the next section, the results of this analysis as it has been presented are reported in detail.

4.2 Results

The results are shown divided in tables and then explained hereafter. In table 4.2, the physical quantities directly inferred from the fit are shown, i.e., the free parameters of the fit. While in table 4.3 the derived quantities are reported. Moreover, in table 4.4 the quantities not related to the single planets but to the whole system are presented.

Planet b is more than two times larger than the Earth, allowing its classification to be somewhere between a super-Earth and a small Neptune. What makes the difference here is the mass determination. Planet b has a mass of almost 8 times the Earth's, falling in the super-Earths range of masses. This gives us, always in a Bayesian sense, a median value for its density of $\rho = 3.29 \pm 0.30\ g/cm^3$. This is smaller than the Earth mean density of $\rho_\oplus = 5.514\ g/cm^3$ but it is not as low as the densities of gaseous planets like Neptune ($\rho_{Neptune} = 1.638\ g/cm^3$). This planet is in between the physical characteristics of the Earth and Neptune. Because of its low density, planets like this is called mini-Neptune in order to distinguish them from the much denser super-Earths. About planet c we can infer less information, unfortunately. Its radius is very similar to the Earth's one. But its mass is unknown, since its radial velocity signal is undetected. Its upper limit at two sigma confidence comprehends the mass of the Earth: it means that

Quantity	Planet b	Planet c	Planet d	Planet e
Period (days)	$5.398751(94 \pm 79)$	$9.6050(44 \pm 15)$	$631.98^{+1.07}_{-1.05}$	1550^{+45}_{-42}
T_c (BJD-2,450,000)	$5066.244(95 \pm 14)$	$5065.42(90^{+15}_{-13})$	/	/
Radius (R_{star})	$0.017(191^{+97}_{-118})$	$0.007(15^{+13}_{-14})$	/	/
b (R_{star})	$0.507^{+0.033}_{-0.049}$	$0.790^{+0.025}_{-0.031}$	/	/
ω (rad)	$1.6^{+1.6}_{-1.7}$	$-1.1^{+2.1}_{-2.3}$	$-1.92^{+0.13}_{-0.12}$	$-2.18^{+0.95}_{-1.08}$
Eccentricity	0.041, 0.082	0.087, 0.18	0.138 ± 0.016	$0.19^{+0.15}_{-0.13}$
K (m/s)	2.74 ± 0.23	0.080, 0.37	17.38 ± 0.32	$2.31^{+0.31}_{-0.28}$

Table 4.2: Table of the resulting values for the parameters of the fit. Each number is the median value of the corresponding distribution and the error is the one-sigma threshold. The only exceptions are the eccentricity of planets b and c and the RV semi-amplitude of planet c, where the upper limit at one sigma (68.27% probability that the true value is lower than that) and the one at two sigma (95.45% probability that the true value is lower than that) are given. T_c is the central time of transit; b is the impact parameter; ω is the argument of pericenter; K is the radial velocity semi-amplitude.

Quantity	Planet b	Planet c	Planet d	Planet e
a (R_{star})	10.50 ± 0.14	$15.42^{+0.20}_{-0.21}$	$251.4^{+3.3}_{-3.4}$	$457.3^{+10.4}_{-10.1}$
a (AU)	$0.061(35 \pm 56)$	$0.090(08^{+82}_{-83})$	$1.468^{+0.013}_{-0.014}$	$2.671^{+0.057}_{-0.055}$
inclination (deg)	$87.22^{+0.21}_{-0.17}$	$87.054^{+0.088}_{-0.095}$	/	/
Radius (R_{Earth})	$2.356^{+0.21}_{-0.22}$	$0.981^{+0.019}_{-0.020}$	/	/
Radius ($R_{Jupiter}$)	$0.2102^{+0.0019}_{-0.0020}$	$0.0875^{+0.0017}_{-0.0018}$	/	/
$T_4 1$ (hours)	$3.464^{+0.101}_{-0.059}$	$2.98^{+0.17}_{-0.15}$	/	/
$T_3 2$ (hours)	$3.306^{+0.107}_{-0.061}$	$2.86^{+0.18}_{-0.16}$	/	/
T_c (BJD-2,450,000)	/	/	$7160.5^{+4.6}_{-4.7}$	7990^{+89}_{-74}
Mass ($10^{-4} M_{star}$)	0.235 ± 0.020	$(8.2, 38.3) \cdot 10^{-3}$	$7.20^{+0.19}_{-0.18}$	1.27 ± 0.16
Mass (M_{Earth})	$7.81^{+0.66}_{-0.68}$	0.27, 1.27	$239.7^{+6.2}_{-6.1}$	$42.2^{+5.3}_{-5.2}$
Mass ($M_{Jupiter}$)	0.0246 ± 0.0021	$(8.6, 40) \cdot 10^{-4}$	0.754 ± 0.019	$0.133^{+0.017}_{-0.016}$

Table 4.3: Table of the derived parameters, values that are calculated starting from table 4.2. Each number is the median value of the corresponding distribution and the error is the one-sigma threshold. The only exception is the mass of planet c, where upper limits at one and two sigma are reported, as explained in the caption of table 4.2. a is the semi-major axis of the orbit; T_{41} is the entire duration of the transit; T_{32} is the duration of the transit excluding ingress and egress; T_c is the time of transit that the planets would have had if they had 90^o inclination. In the case of the non-transiting planets d and e the reported value is the minimum mass $M \cdot sin(i)$.

Quantity	Value	Quantity	Value
KEPLER jitter	$(6.724^{+0.043}_{-0.044}) \cdot 10^{-5}$	HARPS-N jitter (m/s)	$1.75^{+0.17}_{-0.16}$
Density ($\rho_\odot$)	0.534 ± 0.021	HIRES offset (m/s)	$0.23^{+0.51}_{-0.49}$
ld c_1	$0.391^{+0.046}_{-0.045}$	HARPS-N offset (m/s)	$-20905.65^{+0.53}_{-0.52}$
ld c_2	$0.246^{+0.061}_{-0.063}$	X_0 (BJD-2,450,000)	7000.0
HIRES jitter (m/s)	$2.88^{+0.30}_{-0.27}$	lin. coeff. (mm/s/day)	$-1.26^{+0.33}_{-0.35}$

Table 4.4: Table of the results concerning the star and the planetary system in general. Each number is the median value of the corresponding distribution and the error is the one-sigma threshold. *ld* c_1 and *ld* c_2 are the two limb darkening coefficients; X_0 is the reference time for the calculation of the slope of the linear trend; *lin. coeff.* is the first-order coefficient of the linear trend in radial velocities.

there is a (small) probability that planet c is an Earth analog. Apart from this, the proximity with planet b is suggesting a similar internal composition instead, thus leading to a smaller mass. The 1-sigma and 2-sigma upper limits of the density of planet c, in a Bayesian frame, are $\rho = 1.60\ g/cm^3$ and $\rho = 7.48\ g/cm^3$ respectively. We postpone the discussion about densities to the next section, where the radius-mass diagram is presented.

Planet b and planet c orbits at a fraction of astronomical unit from their host star, less than 10% of it. While planet d orbits at a Mars-like distance from the star and planet e even farther out in an orbit that would fit in between the one of Mars and Jupiter. Their periods, rounded to a few digits, are 5.4, 9.6, 632 and 1550 days respectively. Doing some calculations with these numbers allow us to determine possible resonances in the orbits. It turned out that the ratio of the periods of the inner planets is almost exactly in between 4/7 and 9/5, that is not a strong result to conclude that those two planets are in resonance. But most interestingly, planets d and e have a period that is in a 5-to-2 resonance, compatible with the error bars. This is the same resonance present between Jupiter and Saturn. This fact must have had an important role in the formation of the system and the migration of the planets. We will go deeply in this discussion in the next section, where some hypothesis are made.

Last but not least, it is worth to mention the eccentricity of the planets. From the photometric and spectroscopic data at our disposal, it was not possible to establish a precise eccentricity for planets b and c. They have a low upper bound, compatible with the one empirically calculated in Van Eylen et al. (2019) for multi-planetary systems. Planet d's eccentricity is set around 0.14 with a sharp distribution that is not consistent with zero. Eccentricity of planet e instead, has a posterior distribution a bit skewed towards zero and, most importantly, it is compatible with it. In the table of the results we still decided to use the "median plus or minus one sigma" approach instead of the "1-sigma and 2-sigma upper limits" approach because the posterior has still a well defined maximum around 0.19 and the circular orbit solution is outside the 1-sigma threshold from the median. It is relatively probable, then, that planets d and e have a very similar

eccentricity. When the orbit is eccentric it is worth to study its orientation in space, that is given by the argument of pericenter ω, the inclination i and the longitude of the ascending node Ω. Knowing that from the data at our disposal we would never be able to calculate i and Ω for those two planets, ω is the only thing we can observe. It turns out that they have compatible values around -2 radiants. It is not important the absolute value, only their reciprocal values matter. It seems that their semi-major axes are more or less aligned. This is another sign of their ongoing resonance.

In figures 4.1, 4.2, 4.3 and 4.4 the corner plots containing the main free parameters of the fits are shown. Planet b is showing some evident correlations: the one between radius and impact parameter and the one between period and time of transit. If the fit tries to increase the radius, the impact parameter must increase too, and vice versa. This happens because the amount of light blocked by the planet is fixed by the observational data, so enlarging the planet means that the area blocked must be less bright. Thanks to limb darkening effects, the star is less bright in the limbs than in the center. So, in conclusion, if the dimension of the planet is enlarged it must block more peripheral regions of the stellar disk, to still fit the observed data. The correlation between period and central time of transit is even easier to explain. The time of transit is referred to the first transit observed by Kepler at the beginning of its mission. Thus it is straightforward to understand that if the period is slightly increased, there is the necessity to diminish a little bit the time of the first transit to let the last transit to be fitted anyway. The same reason holds diminishing slightly the period and increasing a bit the time of the first transit. It can be shown even with a rough calculation: let's say the period is increased by 0.15 s, as it ca be inferred from the plot. In the lifetime of the Kepler mission, more or less 1350 days, there has been 250 transits of planet b. In such amount of revolutions, that increase in the period would have led to a transit shift of $\Delta Tc = 0.15\ s\ \cdot 250 \simeq 0.6\ min$, that is roughly the one observed in the plot.

Concerning planet c, we can still see the same correlations as planet b, even though the one between radius and impact parameter is less pronounced. Probably this is due to the fact that planet c has an higher impact parameter with respect to planet b. Planet d has one evident correlation between the time of transit and the argument of pericenter and one faint correlation between time of transit and eccentricity. They are both ascribable to the same behaviour. The time of transit of a non-transiting planet is the time of the eventual transit if the planet had 90° inclination. The point is that if the orbit is elliptic, it is not granted that the time of transit corresponds to the moment when the planet is right in the middle of the star's disk. Because the orbital velocity varies along the orbit, the planet might move faster on one half of the transit than in the other. So, for this explanation, rotating a bit the orbit in its plane, that is to say changing a bit its argument of pericenter, may lead to a change in the time of transit. Increasing the eccentricity, this effect can be more pronounced thus leading to a slight variation of the time of transit. Planet e shows the same correlations observed for planet d: time of transit, eccentricity and argument of pericenter are not independent. They are less evident because planet e has some

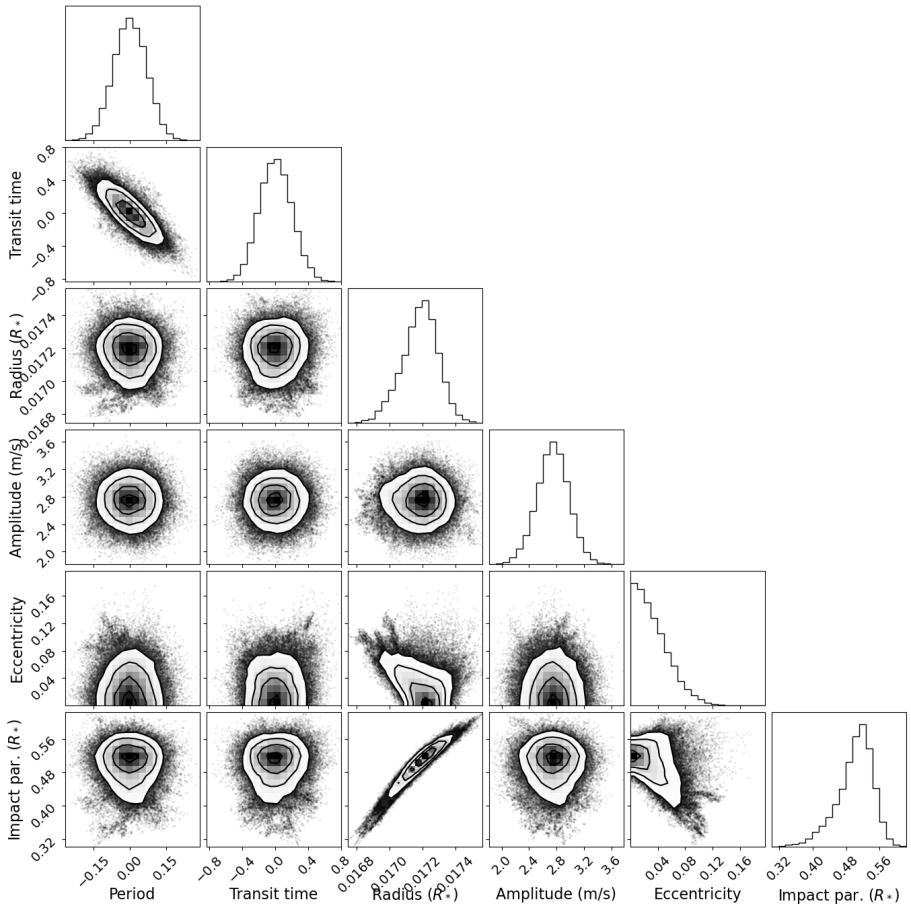

Figure 4.1: Corner plot for planet b. It contains the main results from the fit. Transit time is measured in minutes from the median of the distribution (2, 455, 066.24495 BJD); the period is measured in seconds from the median of the distribution (5.39875194 days).

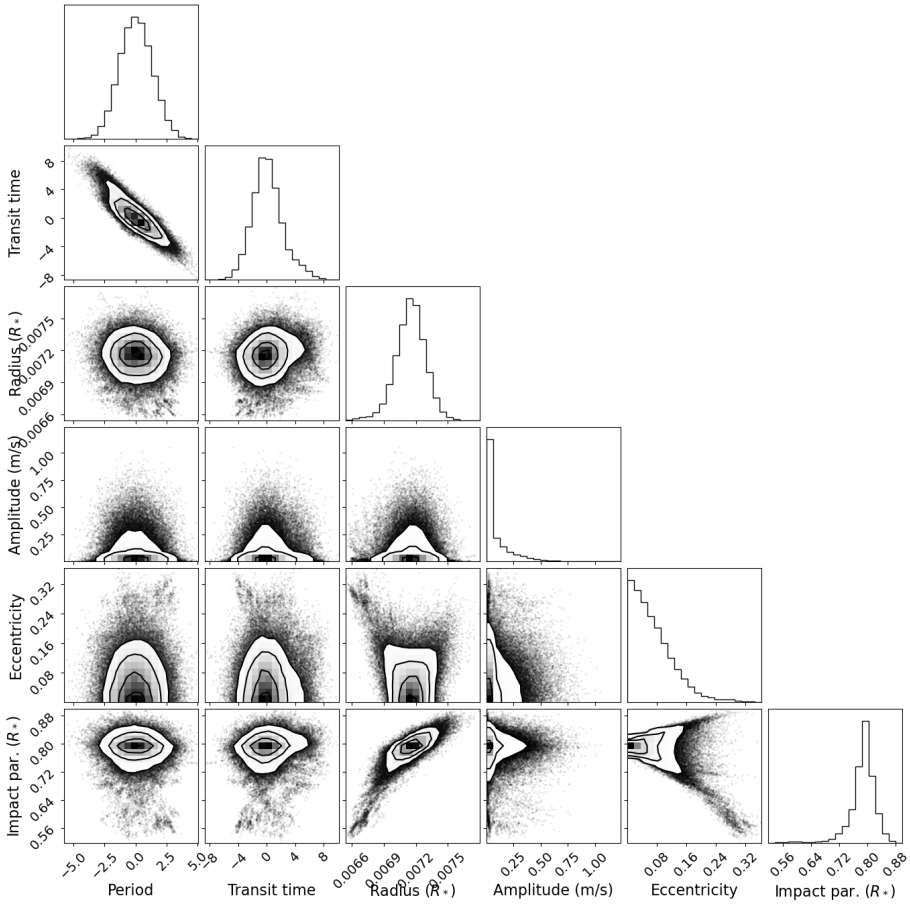

Figure 4.2: Corner plot for planet c. It contains the main results from the fit. Transit time is measured in minutes from the median of the distribution ($2,455,065.4290$ BJD); the period is measured in seconds from the median of the distribution (9.605044 days).

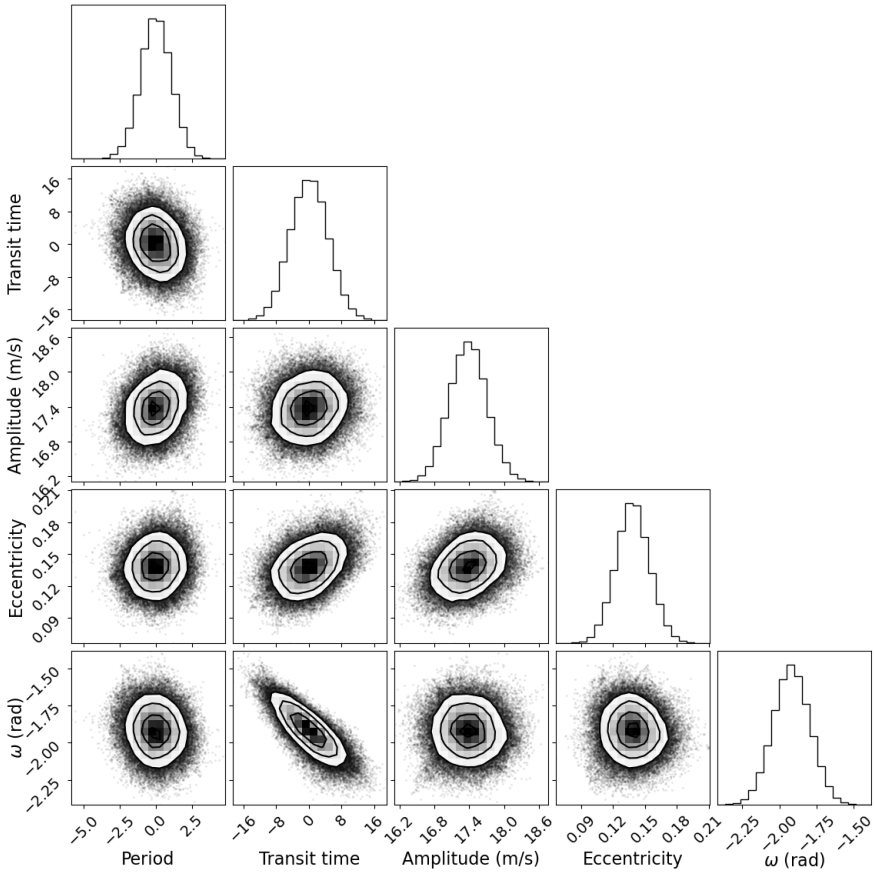

Figure 4.3: Corner plot for planet d. It contains the main results from the fit. Transit time is measured in days from the median of the distribution (2, 457, 160.5 BJD); the period is measured in days from the median of the distribution (631.98 days).

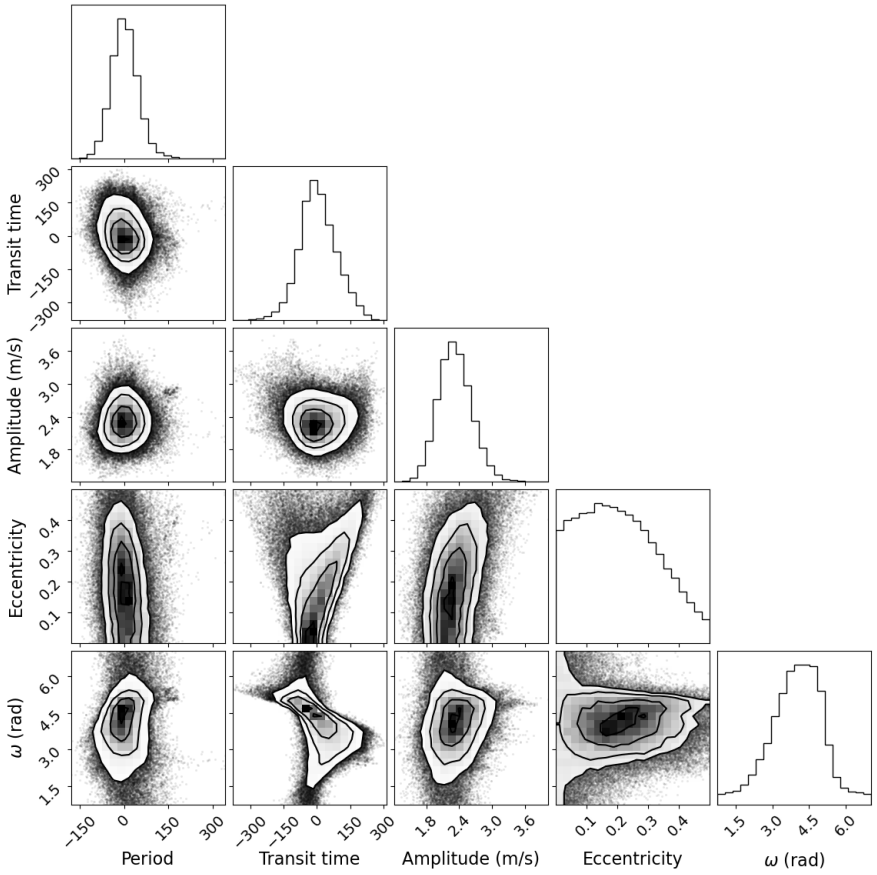

Figure 4.4: Corner plot for planet e. It contains the main results from the fit. Transit time is measured in days from the median of the distribution (2, 457, 990 BJD); the period is measured in days from the median of the distribution (1550 days).

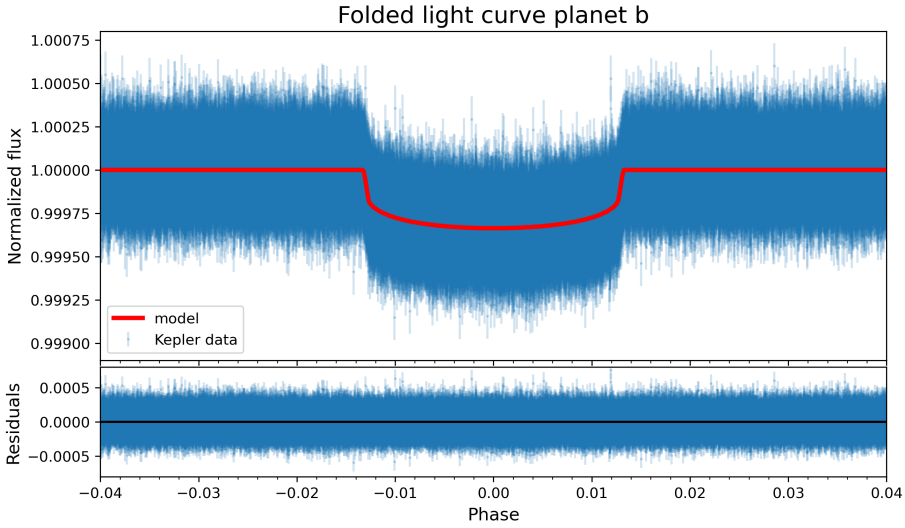

Figure 4.5: Folded light curve for planet b.

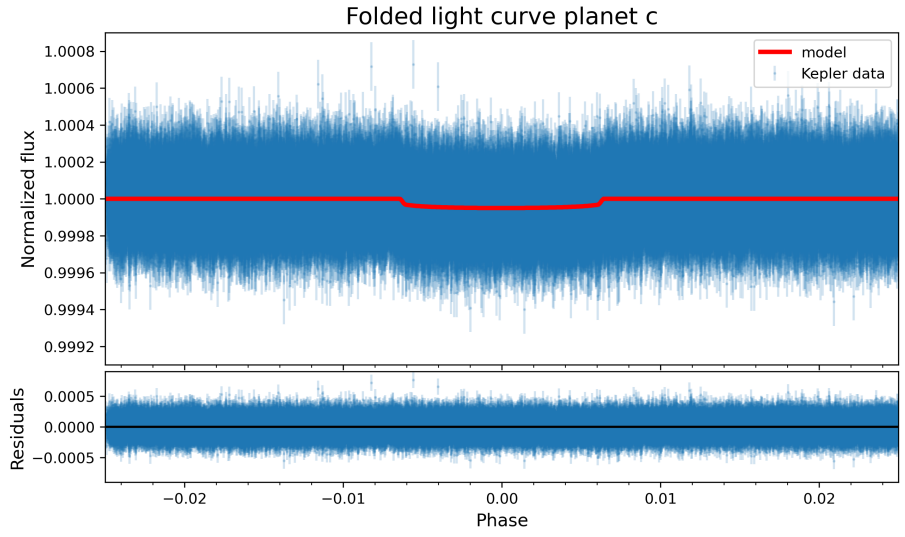

Figure 4.6: Folded light curve for planet c.

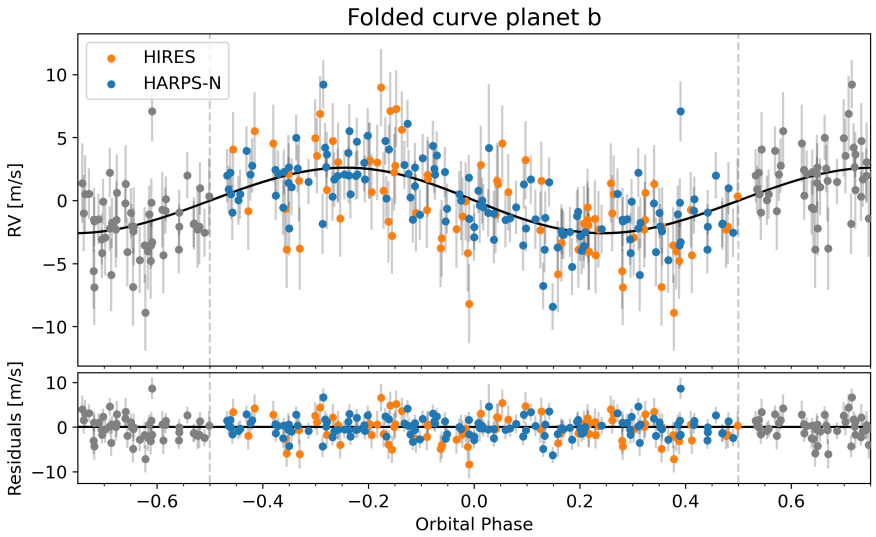

Figure 4.7: Folded RV curve for planet b.

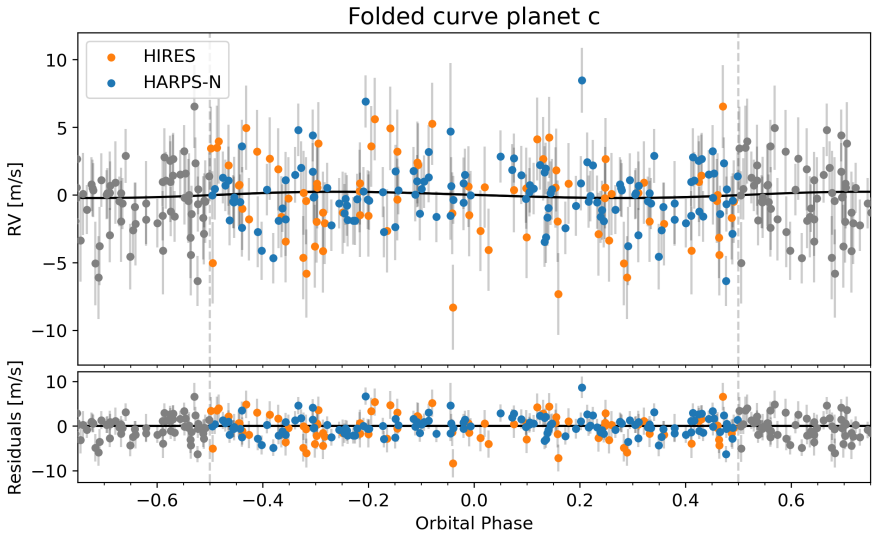

Figure 4.8: Folded RV curve for planet c.

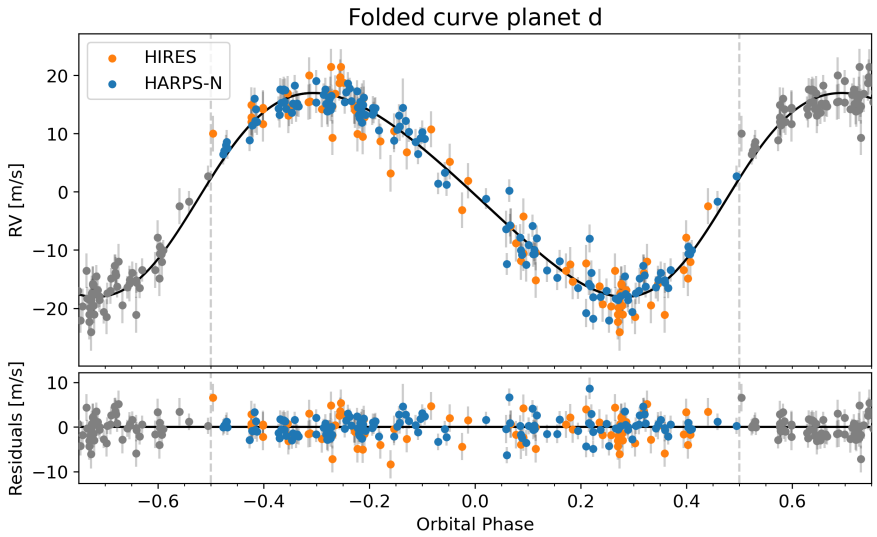

Figure 4.9: Folded RV curve for planet d.

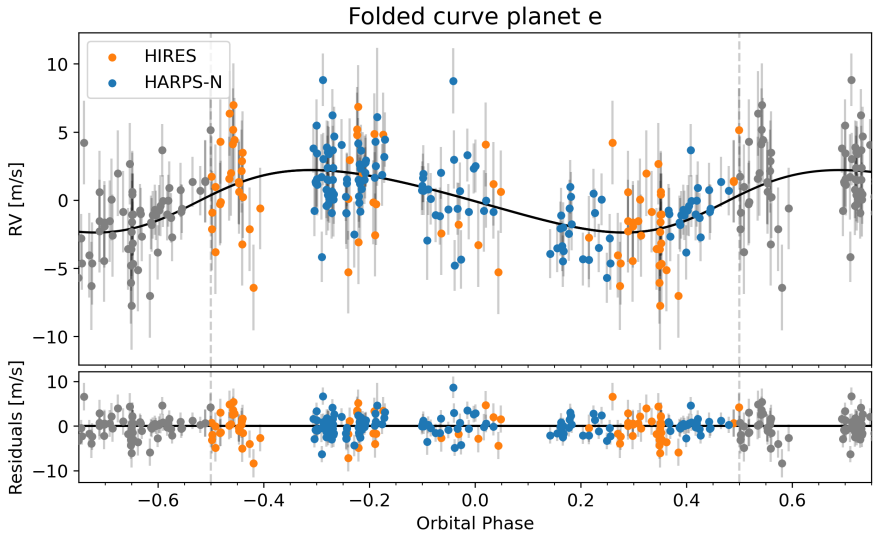

Figure 4.10: Folded RV curve for planet e.

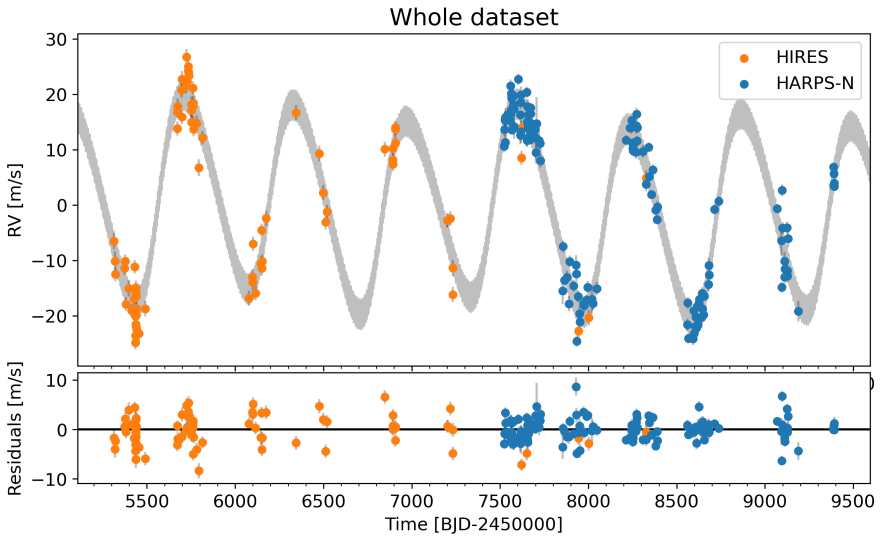

Figure 4.11: Entire RV curve with model superimposed.

solutions with zero eccentricity: this made the correlations fainter.

In figures 4.5 and 4.6 we show the light curve folded on the period of each transiting planet, using Kepler data. In figures 4.7, 4.8, 4.9 and 4.10 we show the radial velocity curve folded around the period of each planet. Each time one planet is plotted, the signal belonging to the other three is subtracted. In figure 4.11 instead, we show the entire radial velocity curve with the overall model and the relative residuals. As we can see, the residuals now are flatter than the ones without including planet e.

4.3 Insights

In this section a couple of discussions about the results of this work are presented: the Radius-Mass diagram and some possible migration paths that the planets which are part of this system may have taken.

4.3.1 Radius-Mass diagram

In figure 4.12 the positions of planets b and c in the Radius-Mass space is shown. The mass axis is logarithmic and the radius axis is in linear scale, for displaying purposes. Other than some already discovered planets, in the background of the plot some iso-density lines corresponding to peculiar internal compositions are drawn. As it can be seen, planet b boasts very small error bars in both axes - the one on the radius is hidden behind the orange dot. Its density of about 3.3 g/cm^3 locates it to the upper left side of the *rocky* dashed line. Kepler-68 b falls in the well-populated region of mini-Neptunes with rocky

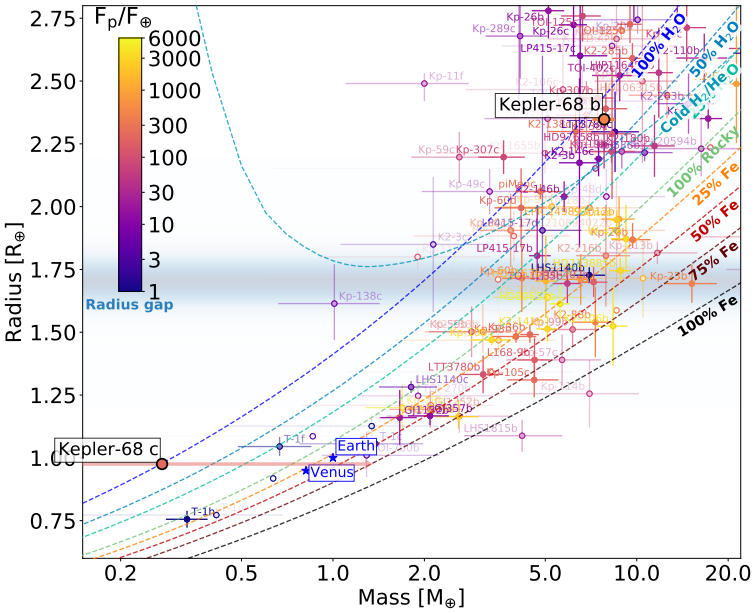

Figure 4.12: Mass-radius plot for planets b and c. Other planets are shown together to put the two planets in a context.

cores and a significant fraction of volatiles or H/He gas, although the observed parameters are compatible with those of a water-world. Notably, this planet is among those with the most precise density measures so far for this kind of planets. Planet c is more difficult to locate on this diagram instead. We only have an upper bound for its mass and thus its density is not univocally defined. In principle, its internal composition could vary from an iron-rich world to a gaseous planet. Since these two planets are very close each other it is plausible that they originated in the same way and migrated together. Therefore it is possible that they have more or less the same internal composition. If that is true, assuming the density of planet b, it is easy to have an estimate of the true mass of planet c: it would be $0.562^{+0.059}_{-0.058}$ $M_\oplus$. Its semi-amplitude in the radial velocity curve would be ~ 16 cm/s, undetectable even with the most modern instruments.

4.3.2 Migration paths

From the knowledge that we have nowadays it is clear that planets b and c can not have formed where we observe them today. They must have undergone some migration processes in the early days of the formation of the planetary system. It is a very difficult task to try to address their evolution because it is a very complex and chaotic process and we have very few and imprecise information about the system. Moreover, we can not treat the evolution of the two inner

planets separated from the one of the two outer planets. Of course, we do not have enough information to infer the evolution that these planets underwent. As a comparison, we could argue that there are still some big uncertainties on the past evolution of the Solar System and some of the most important discoveries came out in the last years, even if we have a lot more data compared to any other planetary system (e.g. Morbidelli et al., 2000; Morbidelli, Levison, Tsiganis, & Gomes, 2005). What we can do is to list the main hypothesis of migration that are compatible with the observed distributions of orbits in this system.

From recent studies, it has been found that the location of the snow line is greatly influenced from the initial conditions of the protoplanetary disk, like its surface density and total mass. For Sun-like stars it can be located between 2 and 5 AUs (Mulders, Ciesla, Min, & Pascucci, 2015). The position of the water snow line, and the position of the *snow lines* of all the other important chemical compounds, is one of the most important aspects to consider when studying the chemical composition of the planets. Even if there are still some theories that allows the formation of giant planets around one or two AUs, it is probable that they formed farther out and migrated inwards. The two main paths of migration that could be realistically occurred in this system are the disk-driven migration and the high-eccentricity migration. In both cases planets d and e have formed beyond the snow line (Petrovich & Tremaine, 2016).

In the first case the planet transfers angular momentum to the disk, migrating inwards. In our Solar System, we have the example of Jupiter and Saturn. They are thought to be born around 20 AUs and to have migrated inwards during their various phases of formation and accretion (Pirani, Johansen, & Mustill, 2019). Their migration is halted when they cross the 2:3 mean motion resonance, after which they migrate outwards a little bit and stabilise in a 2:5 resonance at 5.2 and 9.5 AUs. It is interesting to notice that planets d and e have the same resonance ratio, and maybe they even underwent a similar migration. But another unsolved question regards the halting processes: what are the main reasons that force planets to stop migrating? It could be in this case that a particular resonance between the two planets halted the migration just like in the history of Jupiter and Saturn. Or, in a slightly different scenario, that the disk had again a role in stopping the inward motion.

The second scenario is called Lidov-Kozai effect (Kozai, 1962; Lidov, 1962) and it is based on an oscillation between high eccentricities and high inclinations. During the periods of time characterized by high eccentricity, the planet can be a target for tidal forces and have its orbit circularized to an orbit smaller than the original one. This is a very powerful explanation for those warm-Jupiters with a moderate eccentricity ($0.2 \lesssim e \lesssim 0.7$), even if in this case we are at the lower limit of this empirical interval. This kind of migrations can be driven by secular interactions with an external stellar or planetary companion (Anderson & Lai, 2017). In the Kepler-68 system there is a star that is thought to be a companion of the one hosting the planets. But even if it is gravitationally bound to the system, its projected distance from Kepler-68 would be around 1600 AUs, too far to have a real effect on the orbits of these planets. It has not to be excluded the presence of another undetected planet. The data at our disposal underlined

a faint presence of a long term linear trend in radial velocity. There is no reason to exclude an additional long-period planet that could have had a game-changing impact on the lives of the other planets.

Coming back to the two inner planets, this time there is no doubt that they migrated inwards. Population synthesis models of planet formation by core accretion are failing to explain the observed population of small period planets. These models are predicting a "planet desert" in the parameter space below $\sim 20\ M_{\oplus}$ and below 1 AU (Udry & Santos, 2007). This is in contrast with the observations. What these models cannot take into account, by definition, is what happens after the dissipation of the disk. The observed population could be the result of processes that took place when the planets were already formed and shaped (Howard et al., 2012). We are talking about planet-planet scattering (e.g. Weidenschilling & Marzari, 1996), secular and resonant migration (e.g. Wu & Lithwick, 2011) and Lidov-Kozai effect. It is impossible to prefer one out of these possible evolutionary paths, given the small amount of information that we have. What is clear is that planets b and c do populate the planet desert predicted by the population synthesis models and it will be a goal for upcoming theoretical studies on planetary formation and migration to explain it.

4.4 Conclusions

Kepler-68 planetary system is one of the thousands of systems discovered by Kepler and it is of particular interest thanks to its diversity and complexity. In this work it has been possible to study it deeper than it has ever been done thanks to many new data and approaches. We performed a new analysis of the stellar physical parameters, using the new Gaia Early Data Release 3 and several advanced numerical tools to extract temperature, surface gravity and metallicity from spectra. Putting together these information in a Dartmouth isochrones fitting it was possible to determine stellar mass, radius, density and age to an unprecedented level of precision. To complete the study of the star we searched for stellar activity both in the photometric and spectroscopic datasets. Even if the analysis of the activity indices derived from spectra deserved some particular attention, in the end no sign of activity is found in this star.

The analysis of photometry in search of planetary transits went smoothly thanks to the quietness of the star and the usage of very powerful tools like Wōtan and Transit Least Squares. No additional transits are found on top of the ones belonging to the two known planets. Even the search for single-transit planetary signals did not produce positive results. From the Bayesian analysis of photometry the uncertainty on the radii and other significant parameters is improved with respect to the literature. A photometric estimate of the quadratic limb darkening coefficients is found to be slightly inconsistent with the theoretical prediction based on the physical parameters of the star. Hence for the final all-inclusive analysis we decided not to impose priors on them. Moreover, for the first time the analysis of the TESS light curves is related to the Kepler observations. We searched for long term transit time variations but from the

data at our disposal they were clearly absent.

From the spectroscopic point of view the dataset did not need particular preparation. PyORBIT is programmed to handle different datasets in one single fit, so there was no need to match manually data from HIRES and data from HARPS-N. The Bayesian analysis including the three known planets produced residuals that were showing a modulation. We thus performed a model selection using the most modern and updated statistical tools based on Bayesian evidence: the Nested Sampling technique. The favoured model resulted to be the one with 4 planets and a long term linear trend.

The final PyORBIT fit including both photometric and RV data is performed. No big differences are found in the results with respect to the ones of the separated fits. The error bars are greatly improved both in radii and in masses of the planets, and the fourth planet is confirmed and it is given the name Kepler-68 e. The location of the two transiting planets on the Radius-Mass diagram does not allow us to guess their internal composition because their density is intermediate between the one of the Earth and the one of a gaseous planet like Neptune. Finally, a brief description of the possible evolutionary paths followed by the planets during their formation and migration is depicted. It is very hard to address with no doubts the history of these planets, what we can do are hypothesis made on the small amount of data that we have. This planetary system is far from a complete characterization, but it is a very amenable system for future follow-ups both from the photometric and spectroscopic point of view. Especially in the RV field, it would be of fundamental importance continuing to observe Kepler-68 to better constrain the masses of all the planets and the periods of planets d and e. Studying planetary systems like this, the aim is to better understand planetary populations and their diversity, which is of fundamental importance to understand the Solar System, the place where we live.

References

Agol, E., Steffen, J., Sari, R., & Clarkson, W. (2005, May). On detecting terrestrial planets with timing of giant planet transits. *Monthly Notices of the RAS*, *359*(2), 567-579. doi: 10.1111/j.1365-2966.2005.08922.x

Anderson, K. R., & Lai, D. (2017, December). Moderately eccentric warm Jupiters from secular interactions with exterior companions. *Monthly Notices of the RAS*, *472*(3), 3692-3705. doi: 10.1093/mnras/stx2250

Barclay, T., Burke, C. J., Howell, S. B., Rowe, J. F., Huber, D., Isaacson, H., ... Thompson, S. E. (2013, May). A Super-Earth-sized Planet Orbiting in or Near the Habitable Zone around a Sun-like Star. *Astrophysical Journal*, *768*(2), 101. doi: 10.1088/0004-637X/768/2/101

Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., ... Prsa, A. (2010, February). Kepler Planet-Detection Mission: Introduction and First Results. *Science*, *327*(5968), 977. doi: 10.1126/science.1185402

Borucki, W. J., Koch, D. G., Basri, G., Batalha, N., Brown, T. M., Bryson, S. T., ... Still, M. (2011, July). Characteristics of Planetary Candidates Observed by Kepler. II. Analysis of the First Four Months of Data. *Astrophysical Journal*, *736*(1), 19. doi: 10.1088/0004-637X/736/1/19

Buchhave, L. A., Bizzarro, M., Latham, D. W., Sasselov, D., Cochran, W. D., Endl, M., ... Marcy, G. W. (2014, May). Three regimes of extrasolar planet radius inferred from host star metallicities. *Nature*, *509*(7502), 593-595. doi: 10.1038/nature13254

Buchhave, L. A., Latham, D. W., Johansen, A., Bizzarro, M., Torres, G., Rowe, J. F., ... Quinn, S. N. (2012, June). An abundance of small exoplanets around stars with a wide range of metallicities. *Nature*, *486*(7403), 375-377. doi: 10.1038/nature11121

Buchner, J. (2021, April). UltraNest - a robust, general purpose Bayesian inference engine. *The Journal of Open Source Software*, *6*(60), 3001. doi: 10.21105/joss.03001

Butler, R. P., Vogt, S. S., Laughlin, G., Burt, J. A., Rivera, E. J., Tuomi, M., ... Keiser, S. (2017, May). The LCES HIRES/Keck Precision Radial Velocity Exoplanet Survey. *Astronomical Journal*, *153*(5), 208. doi: 10.3847/1538-3881/aa66ca

Chabrier, G., & Baraffe, I. (2000, January). Theory of Low-Mass Stars and Substellar Objects. *Annual Review of Astron and Astrophys*, *38*, 337-377. doi: 10.1146/annurev.astro.38.1.337

Chabrier, G., Johansen, A., Janson, M., & Rafikov, R. (2014, January). Giant Planet and Brown Dwarf Formation. In H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning (Eds.), *Protostars and planets vi* (p. 619). doi: 10.2458/azu-uapress-9780816531240-ch027

Choi, J., Dotter, A., Conroy, C., Cantiello, M., Paxton, B., & Johnson, B. D. (2016, June). Mesa Isochrones and Stellar Tracks (MIST). I. Solar-scaled Models. *Astrophysical Journal*, *823*(2), 102. doi: 10.3847/0004-637X/823/2/102

Christiansen, J., Jenkins, J. M., Caldwell, D. A., Bryson, S. T., Van Cleve, J., Kolodziejczak, J., ... Borucki, W. J. (2010, October). Short Cadence versus Long Cadence: Affects of Cadence Choice on Kepler Light Curve Analyses. In *Aas/division for planetary sciences meeting abstracts #42* (Vol. 42, p. 59.03).

Claret, A. (2018, October). A new method to compute limb-darkening coefficients for stellar atmosphere models with spherical symmetry: the space missions TESS, Kepler, CoRoT, and MOST. *Astronomy and Astrophysics*, *618*, A20. doi: 10.1051/0004-6361/201833060

Claret, A., Hauschildt, P. H., & Witte, S. (2013, April). New limb-darkening coefficients for Phoenix/1d model atmospheres. II. Calculations for 5000 K $\leqslant$ T_{eff} $\leqslant$ 10 000 K Kepler, CoRot, Spitzer, uvby, UBVRIJHK, Sloan, and 2MASS photometric systems. *Astronomy and Astrophysics*, *552*, A16. doi: 10.1051/0004-6361/201220942

Cosentino, R., Lovis, C., Pepe, F., Collier Cameron, A., Latham, D. W., Molinari, E., ... Weber, L. (2012, September). Harps-N: the new planet hunter at TNG. In I. S. McLean, S. K. Ramsay, & H. Takami (Eds.), *Ground-based and airborne instrumentation for astronomy iv* (Vol. 8446, p. 84461V). doi: 10.1117/12.925738

Cutri, R. M., Skrutskie, M. F., van Dyk, S., Beichman, C. A., Carpenter, J. M., Chester, T., ... Zacarias, N. (2003, June). VizieR Online Data Catalog: 2MASS All-Sky Catalog of Point Sources (Cutri+ 2003). *VizieR Online Data Catalog*, II/246.

Cutri, R. M. e. (2012, April). VizieR Online Data Catalog: WISE All-Sky Data Release (Cutri+ 2012). *VizieR Online Data Catalog*, II/311.

Dotter, A. (2016, January). MESA Isochrones and Stellar Tracks (MIST) 0: Methods for the Construction of Stellar Isochrones. *Astrophysical Journal, Supplement*, *222*(1), 8. doi: 10.3847/0067-0049/222/1/8

Dotter, A., Chaboyer, B., Jevremović, D., Baron, E., Ferguson, J. W., Sarajedini, A., & Anderson, J. (2007, July). The ACS Survey of Galactic Globular Clusters. II. Stellar Evolution Tracks, Isochrones, Luminosity Functions, and Synthetic Horizontal-Branch Models. *Astronomical Journal*, *134*(1), 376-390. doi: 10.1086/517915

Dotter, A., Chaboyer, B., Jevremović, D., Kostov, V., Baron, E., & Ferguson, J. W. (2008, September). The Dartmouth Stellar Evolution Database. *Astrophysical Journal, Supplement*, *178*(1), 89-101. doi: 10.1086/589654

Evans, T. M., & Sackett, P. D. (2010, March). An a Priori Investigation of Astrophysical False Positives in Ground-Based Transiting Planet Surveys. *Astrophysical Journal*, *712*(1), 38-51. doi: 10.1088/0004-637X/712/1/38

Fernández Moroni, G., Estrada, J., Cancelo, G., Holland, S. E., Paolini, E. E., & Diehl, H. T. (2012, July). Sub-electron readout noise in a Skipper CCD fabricated on high resistivity silicon. *Experimental Astronomy*, *34*(1), 43-64. doi: 10.1007/s10686-012-9298-x

Feroz, F., Balan, S. T., & Hobson, M. P. (2011, August). Detecting extrasolar planets from stellar radial velocities using Bayesian evidence. *Monthly Notices of the RAS*, *415*(4), 3462-3472. doi: 10.1111/j.1365-

2966.2011.18962.x

Fischer, D. A., Anglada-Escude, G., Arriagada, P., Baluev, R. V., Bean, J. L., Bouchy, F., ... Wright, J. T. (2016, June). State of the Field: Extreme Precision Radial Velocities. *Publications of the ASP*, *128*(964), 066001. doi: 10.1088/1538-3873/128/964/066001

Foreman-Mackey, D., Farr, W., Sinha, M., Archibald, A., Hogg, D., Sanders, J., ... Pla, O. (2019, November). emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC. *The Journal of Open Source Software*, *4*(43), 1864. doi: 10.21105/joss.01864

Frenkel, A., Sartor, M. A., & Wlodawski, M. S. (1997, August). Photon-noise-limited operation of intensified CCD cameras. *Applied Optics*, *36*(22), 5288-5297. doi: 10.1364/AO.36.005288

Fressin, F., Torres, G., Charbonneau, D., Bryson, S. T., Christiansen, J., Dressing, C. D., ... Batalha, N. M. (2013, April). The False Positive Rate of Kepler and the Occurrence of Planets. *Astrophysical Journal*, *766*(2), 81. doi: 10.1088/0004-637X/766/2/81

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., ... Zwitter, T. (2021, May). Gaia Early Data Release 3. Summary of the contents and survey properties. *Astronomy and Astrophysics*, *649*, A1. doi: 10.1051/0004-6361/202039657

Gaudi, B. S. (2010, February). Exoplanetary Microlensing. *arXiv e-prints*, arXiv:1002.0332.

Gilliland, R. L., Jenkins, J. M., Borucki, W. J., Bryson, S. T., Caldwell, D. A., Clarke, B. D., ... van Cleve, J. E. (2010, April). Initial Characteristics of Kepler Short Cadence Data. *Astrophysical Journal, Letters*, *713*(2), L160-L163. doi: 10.1088/2041-8205/713/2/L160

Gilliland, R. L., Marcy, G. W., Rowe, J. F., Rogers, L., Torres, G., Fressin, F., ... Thompson, S. E. (2013, March). Kepler-68: Three Planets, One with a Density between that of Earth and Ice Giants. *Astrophysical Journal*, *766*(1), 40. doi: 10.1088/0004-637X/766/1/40

Ginski, C., Mugrauer, M., Seeliger, M., Buder, S., Errmann, R., Avenhaus, H., ... Raetz, S. (2016, April). A lucky imaging multiplicity study of exoplanet host stars - II. *Monthly Notices of the RAS*, *457*(2), 2173-2191. doi: 10.1093/mnras/stw049

Grieves, N., Ge, J., Thomas, N., Willis, K., Ma, B., Lorenzo-Oliveira, D., ... Pan, K. (2018, December). Chemo-kinematics of the Milky Way from the SDSS-III MARVELS survey. *Monthly Notices of the RAS*, *481*(3), 3244-3265. doi: 10.1093/mnras/sty2431

Hermes, J. J. (2018). Timing by Stellar Pulsations as an Exoplanet Discovery Method. In H. J. Deeg & J. A. Belmonte (Eds.), *Handbook of exoplanets* (p. 6). doi: 10.1007/978-3-319-55333-7-6

Hippke, M., David, T. J., Mulders, G. D., & Heller, R. (2019, October). Wōtan: Comprehensive Time-series Detrending in Python. *Astronomical Journal*, *158*(4), 143. doi: 10.3847/1538-3881/ab3984

Hippke, M., & Heller, R. (2019, Mar). Optimized transit detection algorithm to search for periodic transits of small planets. *Astronomy and Astrophysics*,

623, A39. doi: 10.1051/0004-6361/201834672

Høg, E., Fabricius, C., Makarov, V. V., Urban, S., Corbin, T., Wycoff, G., ... Wicenec, A. (2000, March). The Tycho-2 catalogue of the 2.5 million brightest stars. *Astronomy and Astrophysics*, *355*, L27-L30.

Howard, A. W., Marcy, G. W., Bryson, S. T., Jenkins, J. M., Rowe, J. F., Batalha, N. M., ... MacQueen, P. J. (2012, August). Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler. *Astrophysical Journal, Supplement*, *201*(2), 15. doi: 10.1088/0067-0049/201/2/15

Kipping, D. M. (2009, January). Transit timing effects due to an exomoon. *Monthly Notices of the RAS*, *392*(1), 181-189. doi: 10.1111/j.1365-2966.2008.13999.x

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

Kozai, Y. (1962, November). Secular perturbations of asteroids with high inclination and eccentricity. *Astronomical Journal*, *67*, 591-598. doi: 10.1086/108790

Kramer, M. (2018). Pulsar Timing as an Exoplanet Discovery Method. In H. J. Deeg & J. A. Belmonte (Eds.), *Handbook of exoplanets* (p. 5). doi: 10.1007/978-3-319-55333-7-5

Lagrange, A. M., Bonnefoy, M., Chauvin, G., Apai, D., Ehrenreich, D., Boccaletti, A., ... Kasper, M. (2010, July). A Giant Planet Imaged in the Disk of the Young Star β Pictoris. *Science*, *329*(5987), 57. doi: 10.1126/science.1187187

Lang, K. R. (2013). *Essential Astrophysics*. Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-35963-7

Lidov, M. L. (1962, October). The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planetary Space Science*, *9*(10), 719-759. doi: 10.1016/0032-0633(62)90129-0

Lovis, C., & Fischer, D. (2010). Radial Velocity Techniques for Exoplanets. In S. Seager (Ed.), *Exoplanets* (p. 27-53).

Lovis, C., & Pepe, F. (2007, June). A new list of thorium and argon spectral lines in the visible. *Astronomy and Astrophysics*, *468*(3), 1115-1121. doi: 10.1051/0004-6361:20077249

Ma, B., Shang, Z., Hu, Y., Liu, Q., Wang, L., & Wei, P. (2014, July). A new method of CCD dark current correction via extracting the dark Information from scientific images. In A. D. Holland & J. Beletic (Eds.), *High energy, optical, and infrared detectors for astronomy vi* (Vol. 9154, p. 91541T). doi: 10.1117/12.2055416

Malavolta, L. (2016, December). *PyORBIT: Exoplanet orbital parameters and stellar activity.*

Malavolta, L., Lovis, C., Pepe, F., Sneden, C., & Udry, S. (2017, July). *CCFpams: Atmospheric stellar parameters from cross-correlation functions.*

Malbet, F., & Sozzetti, A. (2018). Astrometry as an Exoplanet Discovery Method. In H. J. Deeg & J. A. Belmonte (Eds.), *Handbook of exoplanets*

(p. 196). doi: 10.1007/978-3-319-55333-7-196

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

McQuillan, A., Aigrain, S., & Mazeh, T. (2013, June). Measuring the rotation period distribution of field M dwarfs with Kepler. *Monthly Notices of the RAS*, *432*(2), 1203-1216. doi: 10.1093/mnras/stt536

Morbidelli, A., Chambers, J., Lunine, J. I., Petit, J. M., Robert, F., Valsecchi, G. B., & Cyr, K. E. (2000, November). Source regions and time scales for the delivery of water to Earth. *Meteoritics and Planetary Science*, *35*(6), 1309-1320. doi: 10.1111/j.1945-5100.2000.tb01518.x

Morbidelli, A., Levison, H. F., Tsiganis, K., & Gomes, R. (2005, May). Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature*, *435*(7041), 462-465. doi: 10.1038/nature03540

Morton, T. D. (2015, March). *isochrones: Stellar model grid package.*

Mulders, G. D., Ciesla, F. J., Min, M., & Pascucci, I. (2015, July). The Snow Line in Viscous Disks around Low-mass Stars: Implications for Water Delivery to Terrestrial Planets in the Habitable Zone. *Astrophysical Journal*, *807*(1), 9. doi: 10.1088/0004-637X/807/1/9

Nelson, B. E., Ford, E. B., Buchner, J., Cloutier, R., Díaz, R. F., Faria, J. P., ... Rukdee, S. (2020, February). Quantifying the Bayesian Evidence for a Planet in Radial Velocity Data. *Astronomical Journal*, *159*(2), 73. doi: 10.3847/1538-3881/ab5190

Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. (1984, April). Rotation, convection, and magnetic activity in lower main-sequence stars. *Astrophysical Journal*, *279*, 763-777. doi: 10.1086/161945

Paxton, B., Bildsten, L., Dotter, A., Herwig, F., Lesaffre, P., & Timmes, F. (2011, January). Modules for Experiments in Stellar Astrophysics (MESA). *Astrophysical Journal, Supplement*, *192*(1), 3. doi: 10.1088/0067-0049/192/1/3

Paxton, B., Cantiello, M., Arras, P., Bildsten, L., Brown, E. F., Dotter, A., ... Townsend, R. (2013, September). Modules for Experiments in Stellar Astrophysics (MESA): Planets, Oscillations, Rotation, and Massive Stars. *Astrophysical Journal, Supplement*, *208*(1), 4. doi: 10.1088/0067-0049/208/1/4

Paxton, B., Marchant, P., Schwab, J., Bauer, E. B., Bildsten, L., Cantiello, M., ... Timmes, F. X. (2015, September). Modules for Experiments in Stellar Astrophysics (MESA): Binaries, Pulsations, and Explosions. *Astrophysical Journal, Supplement*, *220*(1), 15. doi: 10.1088/0067-0049/220/1/15

Pepe, F., Ehrenreich, D., & Meyer, M. R. (2014, September). Instrumentation for the detection and characterization of exoplanets. *Nature*, *513*(7518), 358-366. doi: 10.1038/nature13784

Perryman, M. (2018). *The Exoplanet Handbook* (Second Edition ed.). Cambridge University Press.

Petigura, E. A., Howard, A. W., Marcy, G. W., Johnson, J. A., Isaacson, H., Cargile, P. A., ... Crossfield, I. J. M. (2017, September). The California-Kepler Survey. I. High-resolution Spectroscopy of 1305 Stars Hosting Kepler

Transiting Planets. *Astronomical Journal*, *154*(3), 107. doi: 10.3847/1538-3881/aa80de

Petrovich, C., & Tremaine, S. (2016, October). Warm Jupiters from Secular Planet-Planet Interactions. *Astrophysical Journal*, *829*(2), 132. doi: 10.3847/0004-637X/829/2/132

Pirani, S., Johansen, A., & Mustill, A. J. (2019, November). On the inclinations of the Jupiter Trojans. *Astronomy and Astrophysics*, *631*, A89. doi: 10.1051/0004-6361/201936600

Ranalli, P., Hobbs, D., & Lindegren, L. (2018, June). Astrometry and exoplanets in the Gaia era: a Bayesian approach to detection and parameter recovery. *Astronomy and Astrophysics*, *614*, A30. doi: 10.1051/0004-6361/201730921

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., ... Villasenor, J. (2015, January). Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, *1*, 014003. doi: 10.1117/1.JATIS.1.1.014003

Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R., & Zolotukhin, I. (2011, August). Defining and cataloging exoplanets: the exoplanet.eu database. *Astronomy and Astrophysics*, *532*, A79. doi: 10.1051/0004-6361/201116713

Skilling, J. (2004, November). Nested Sampling. In R. Fischer, R. Preuss, & U. V. Toussaint (Eds.), *Bayesian inference and maximum entropy methods in science and engineering: 24th international workshop on bayesian inference and maximum entropy methods in science and engineering* (Vol. 735, p. 395-405). doi: 10.1063/1.1835238

Soter, S. (2006, December). What Is a Planet? *Astronomical Journal*, *132*(6), 2513-2519. doi: 10.1086/508861

Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., & Monteiro, M. J. P. F. G. (2007, July). A new code for automatic determination of equivalent widths: Automatic Routine for line Equivalent widths in stellar Spectra (ARES). *Astronomy and Astrophysics*, *469*(2), 783-791. doi: 10.1051/0004-6361:20077288

Speagle, J. S. (2020, April). DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the RAS*, *493*(3), 3132-3158. doi: 10.1093/mnras/staa278

Stassun, K. G., Oelkers, R. J., Paegert, M., Torres, G., Pepper, J., De Lee, N., ... Winn, J. N. (2019, October). The Revised TESS Input Catalog and Candidate Target List. *Astronomical Journal*, *158*(4), 138. doi: 10.3847/1538-3881/ab3467

Stevens, D. J., & Gaudi, B. S. (2013, August). A Posteriori Transit Probabilities. *Publications of the ASP*, *125*(930), 933. doi: 10.1086/672572

Storn, R., & Price, K. (1997, December). Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, *11*, 341–359. doi: 10.1023/A:1008202821328

Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., Charbonneau, D., Deming, D., Dressing, C. D., ... Woods, D. (2015, August). The Transiting Exoplanet Survey Satellite: Simulations of Planet Detections and Astrophysical

False Positives. *Astrophysical Journal*, *809*(1), 77. doi: 10.1088/0004-637X/809/1/77

Tukey, J. W. (1977). *Exploratory data analysis*.

Udry, S., & Santos, N. C. (2007, September). Statistical Properties of Exoplanets. *Annual Review of Astron and Astrophys*, *45*(1), 397-439. doi: 10.1146/annurev.astro.45.051806.110529

Valenti, J. A., & Piskunov, N. (1996, September). Spectroscopy made easy: A new tool for fitting observations with synthetic spectra. *Astronomy and Astrophysics, Supplement*, *118*, 595-603.

Van Eylen, V., Albrecht, S., Huang, X., MacDonald, M. G., Dawson, R. I., Cai, M. X., ... Winn, J. N. (2019, February). The Orbital Eccentricity of Small Planet Systems. *Astronomical Journal*, *157*(2), 61. doi: 10.3847/1538-3881/aaf22f

Vaughan, A. H., Preston, G. W., & Wilson, O. C. (1978, June). Flux measurements of Ca II and K emission. *Publications of the ASP*, *90*, 267-274. doi: 10.1086/130324

Vogt, S. S., Allen, S. L., Bigelow, B. C., Bresee, L., Brown, B., Cantrall, T., ... Wei, M. Z. (1994, June). HIRES: the high-resolution echelle spectrometer on the Keck 10-m Telescope. In D. L. Crawford & E. R. Craine (Eds.), *Instrumentation in astronomy viii* (Vol. 2198, p. 362). doi: 10.1117/12.176725

Wehrhahn, A. (2021, March). PySME - Spectroscopy Made Easier. In *Cambridge workshop on cool stars, stellar systems, and the sun* (p. 1). doi: 10.5281/zenodo.4537913

Weidenschilling, S. J., & Marzari, F. (1996, December). Gravitational scattering as a possible origin for giant planets at small stellar distances. *Nature*, *384*(6610), 619-621. doi: 10.1038/384619a0

Winn, J. N. (2010, January). Transits and Occultations. *arXiv e-prints*, arXiv:1001.2010.

Wu, Y., & Lithwick, Y. (2011, July). Secular Chaos and the Production of Hot Jupiters. *Astrophysical Journal*, *735*(2), 109. doi: 10.1088/0004-637X/735/2/109

Zechmeister, M., & Kürster, M. (2009, March). The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *Astronomy and Astrophysics*, *496*(2), 577-584. doi: 10.1051/0004-6361:200811296