

FACULTAD DE CIENCIAS FÍSICAS
MÁSTER EN ASTROFÍSICA



**New habitable worlds:
the KOBE experiment**

Por:

Olga Balsalobre Ruza

Tutor:

Dr. Jorge Lillo-Box

TRABAJO FIN DE MÁSTER

Madrid, Junio de 2021

Agradecimientos

No puedo estar más agradecida a Jorge Lillo-Box. No sólo por haber sido un tutor inmejorable durante este trabajo. También por confiar en mí desde el primer momento, por saber transmitir su pasión por la Astrofísica, y por enseñarme tanto de este campo, que antes tan solo miraba a la distancia, y del que ahora quiero formar parte. Este trabajo ha sido el impulso que necesitaba.

Gracias a mis compis de clase, que a pesar de habernos visto más a través de una pantalla que en persona, me han hecho sentir que no estábamos tan lejos. Especialmente gracias a Iris y Alba, por alegrarnos de los logros de cada una como si fueran el de una misma. A Alex, por habernos entendido tan bien. A Gojo, por sacar tiempo en mitad del caos pre-exámenes y echarme un cable con sus trucos de hacker. A ellos, y a todos los que me dejo por mencionar. Porque este Máster ha estado marcado por el buen rollo que fluye en todas direcciones. Sin olvidarme del profesorado, porque como solemos comentar entre nosotros, el departamento de Astrofísica de la Complutense está lleno de personas encantadoras. Gracias.

Gracias a mi abuela Maruja, por ponerme una velita en cada examen desde hace años. A mi abuela Amalia, que aunque es más mayor de lo que aparenta su piel, no le teme a preguntarme sobre la inmensidad del Universo, aunque siempre acabe con un: «Bueno, a mí esto ya se me escapa». Prometo recuperar el tiempo y los abrazos que no les he dado desde el famoso Marzo del año pasado.

A las tres personas que ven todo lo que estoy luchando por mis sueños y no dejan de animarme y apoyarme para que los alcance. Sin vosotros no podría seguir avanzando como lo estoy haciendo, papá, mamá y Lauri. Mi madre, que por fin deja de ser opositora para ser funcionaria, ella me ha demostrado que con trabajo duro y paciencia se recoge lo que se siembra.

Y sobretodo, a quien consiguió que creyera más en mí y trata de que no me ponga límites. Que siempre está a mi lado, que se ilusiona cuando le hablo de lo increíble que es detectar estos mundos tan lejanos, que me apoya cuando las cosas no van tan bien. Gracias a Antonio, por todo lo que él ya sabe.

Abstract

The search for potentially habitable worlds beyond the Solar System so far, has been focused on those stars similar to the Sun (G-dwarfs), and in low mass stars where Earth-mass planets are easier to detect (M-dwarfs). However, stars between these two types (K-dwarfs), present the most favorable characteristics in terms of habitability-detectability trade-off. The KOBE project, as a legacy program of Calar Alto, has begun to monitor the radial velocities of 50 late K-type stars, expecting to detect between 14 and 40 new exoplanets with CARMENES instrument.

In this work, we make the selection of the 50 objects for the KOBE sample. We do it by searching for archival data in high-resolution spectrographs and by using the Spanish Virtual Observatory (SVO) tools. We select the most favorable stars taking into account brightness, absence of stellar activity, and spectral features that allow an optimal extraction of the radial velocity (low rotational velocity, main sequence stage, solar-like metallicity or metal-rich).

We also develop a Bayesian-based algorithm to improve the efficiency of the observations: KOBESim. We start with time-equispaced observations according to a given cadence to a star. Once there is enough data to distinguish a signal in the periodogram, that period is pursued with the proposed algorithm. The method we follow to speed up the detection of the hypothetical planet is based on selecting the next observing date that maximizes the increase in the Bayesian evidence for such periodicity. Testing it with simulated data, we prove that KOBESim accelerates the detection process by almost a factor 2 for low-mass planets ($M < 10 M_{\oplus}$).

Resumen

Hasta ahora, la búsqueda de mundos potencialmente habitables a las afueras del Sistema Solar ha estado enfocada en aquellas estrellas más similares al Sol (enanas G), y en estrellas de baja masa donde es más sencillo detectar un planeta de masa terrestre (enanas M). Sin embargo, las intermedias entre estos dos tipos (enanas K), presentan unas características más favorables en cuanto a compromiso habitabilidad-detectabilidad. El proyecto KOBE, como programa de legado de Calar Alto, ha comenzado a monitorear las velocidades radiales de 50 estrellas de tipo K tardío, esperando detectar entre 14 y 40 nuevos exoplanetas con el instrumento CARMENES.

En este trabajo, realizamos la selección de los 50 objetos que conforman la muestra de KOBE. Lo hacemos mediante la búsqueda en los archivos de espectrógrafos de alta resolución y del uso de herramientas del Observatorio Virtual Español (SVO). Seleccionamos las estrellas más favorables teniendo en cuenta el brillo, ausencia de actividad estelar, y características espectrales que permitan una extracción óptima de la velocidad radial (velocidad de rotación baja, etapa en la secuencia principal, metalicidad similar a la solar o rica en metales).

También desarrollamos un algoritmo basado en estadística Bayesiana para mejorar la eficiencia de las observaciones: KOBESim. Comenzamos con observaciones temporalmente equiespaciadas según una cadencia dada a una estrella. Una vez que se tienen los suficientes datos para distinguir una señal en el periodograma, se persigue dicho periodo con el algoritmo propuesto. El método que seguimos para acelerar la detección del hipotético planeta, radica en seleccionar la próxima fecha de observación que maximiza el incremento en evidencia Bayesiana para dicha periodicidad. Con su aplicación a datos simulados, probamos que KOBESim consigue acelerar el proceso de detección en casi un factor 2 para planetas poco masivos ($M < 10 M_{\oplus}$).

Contents

1	Introduction	1
1.1	Hunting for exoplanets	2
1.2	Planets in the Habitable Zone	4
1.3	An emptiness around K-dwarfs	5
1.4	The KOBE experiment	7
1.4.1	Blind-search surveys: observational strategies	7
1.5	Objectives of this work	9
2	Selection of the KOBE sample	9
3	KOBESim: a Bayesian strategy	11
3.1	Bayesian inference	11
3.2	KOBESim	13
3.2.1	Goals and architecture	13
3.2.2	Testing KOBESim on simulated datasets	16
3.2.3	Application to real data: the case of KOBE-002	21
4	Conclusions and Future Work	23
A	Radial Velocity	26
B	Additional figures for KOBESim	27
B.1	Simulated data	27
B.2	Real KOBE-002 data	31

1 Introduction

Moving away from the geocentric idea to discovering the other components of our Solar System was a game changer in our comprehension. It sows the seed to whether other planetary systems beyond us could exist. It was not until 1992 that a claim of the discovery of an extrasolar planet was undoubtful. The periodicity of a pulsar showed some irregularities revealing the presence of at least two orbiting planets (Wolszczan & Frail, 1992). However, to discover the first exoplanet around a main sequence star an improvement in the wavelength calibration and stability of high-resolution spectrographs was still needed¹. Mayor and Queloz were the first to achieve this feat with the radial velocity (RV) method (Mayor & Queloz, 1995), a revolution in astrophysics that set precedent to a field that today has resulted in the discovery of 4401 extrasolar planets².

Of special scientific interest are the exoplanets that reside in the habitable zone (hereafter, HZ) of their host star. This is the range of distances from their star that allows a rocky planet to have liquid water on its surface under the assumption of atmospheric pressure (Kasting *et al.*, 1993; Kopparapu *et al.*, 2013). In the search for potentially habitable worlds, the presence of water is considered a key requirement. Water enables the transport of substances, and it is the solvent used by living organisms in chemical reactions, thus it is indispensable at least for terrestrial life (e.g. Cockell *et al.*, 2016). To date, the pursuit of these planets has focused on G- and M-type stars, leaving a barely explored desert around K-type stars despite being ideal hosts for life development, as we will see through this dissertation. The KOBE experiment (**K**-dwarfs **O**rbited **B**y habitable **E**xoplanets)³ arises with the purpose of filling this emptiness. Making use of the fiber-fed echelle spectrograph CARMENES (**C**alar **A**lto high-**R**esolution search for **M** dwarfs with **E**xoearth with **N**ear-infrared and optical **E**chelle **S**pectrographs, Quirrenbach *et al.*, 2016), located at the 3.5 m telescope of Calar Alto observatory (CAHA), KOBE will monitor the RV of 50 K-dwarfs across 5 semesters. The project aims at detecting between 14-40 new worlds, a handful of them being rocky within the HZ. To achieve this ambitious goal, a careful selection of the stars together with the use of strategies that optimize the observations, are needed.

This Master Thesis is framed in the preparation and first steps of the KOBE experiment and has two main goals. Firstly, the selection of the 50 stars to be monitored. Secondly, the development of a Bayesian-based algorithm to optimize the efficiency of the observing strategy. In this Section 1, we describe the scientific context and the KOBE experiment. Here, we present the methods used to detect exoplanets, as well as the HZ definition. Besides, we outline the reasons behind searching for planets around K-dwarfs, the main characteristics of KOBE, and the motivations and objectives of this work. In Section 2 we present the target sample. In Section 3 we describe the development of the algorithm, KOBEsim, and we give the basics of Bayesian inference. Finally, Section 4 is a summary where we discuss the conclusions and future work.

¹Wavelength calibration in high-precision spectrographs was improved by (1) simultaneously observing the star and the calibration lamp and (2) inserting HF or iodine cells in the optical path of the instrument.

²According to NASA Exoplanet Archive (Akeson *et al.*, 2013) on 6 of June 2021

³<https://kobe.caha.es/>

1.1 Hunting for exoplanets

Three decades ago the only information we had on planetary systems was on the Solar System. There was no observational record of the existence of planets around any star other than the Sun. However, the detection of a Jupiter-like planet around a main sequence star, 51 Pegasi b, demonstrated the plausibility of discovering new worlds. It was a milestone awarded in 2019 with the Nobel Prize in Physics (Mayor & Queloz, 1995). The reason why extrasolar planets were not discovered earlier is the great challenge of detecting objects of such small sizes and masses and their intrinsic faintness. This fact creates the need to resort to indirect observational techniques.

The two most effective techniques are:

- **Transit:** infers the presence of a planet due to the brightness dimming of its host star when they are in conjunction, as the star is being eclipsed by the planet from the point of view of the observer.
- **Radial Velocity:** detects the wobble of the star around the center of mass of the system when a planet surrounds it. It is measured by the Doppler effect which shifts spectral lines to the red when moving away in the line of sight and towards the blue when approaching.

We show a diagram for both methods in Figure 1. These two methods are responsible for the detection of 95.1% of exoplanets known to date.

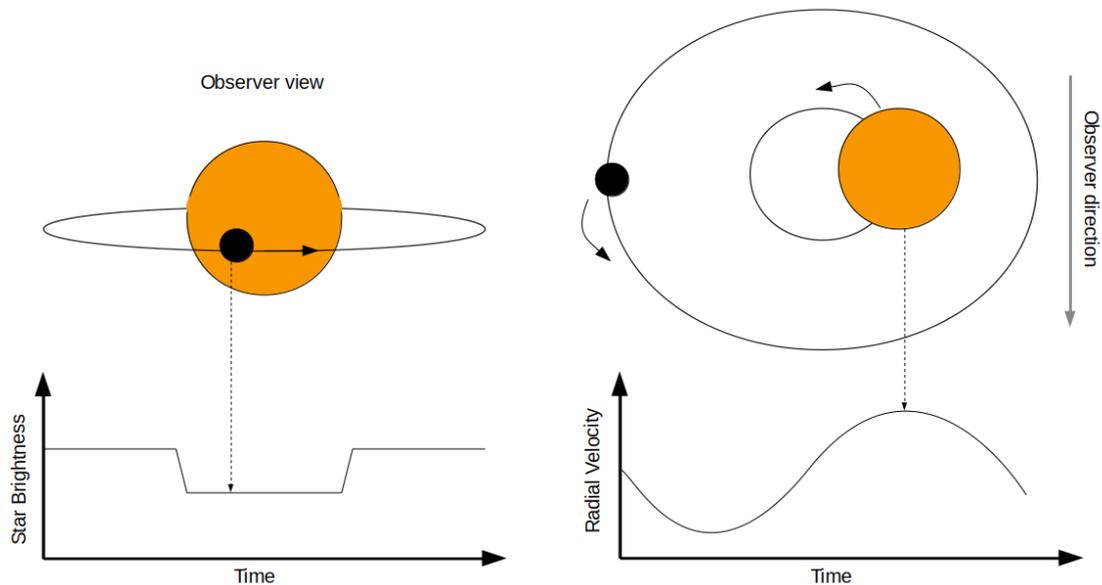


Figure 1: Left: representation of the transit of a planet through its host star and the resulting light curve. Right: wobble diagram of the host star around the center of mass of the system and the resulting RV graph.

These two techniques are complementary. Through transits, it is possible to determine the relative size of the planet and the star, while by the RV, a lower limit of the planetary mass⁴ can

⁴ $M_p \sin i$, where i is the angle between the plane of the orbit and the plane of the sky known as the orbital inclination (See Figure 14).

be inferred. It has to be taken into account when doing statistical studies on the exoplanet population, that the nature of these methods leads to selection bias. From the geometric point of view, for a transit to be observable, the orbit needs to be approximately edge-on ($i \approx 90^\circ$). Also, to detect the Doppler shift, this angle cannot be zero. Apart from that, the transit method is favored by large planets, and RV by massive ones. Being, both of them, favored by planets close to their host star.

Although not as efficient, other methods have provided 4.1% of all detections:

- **Direct imaging:** making use of coronagraphs and complex adaptive optics systems, it is possible to directly observe the reflected light of planets in very wide orbits. The complexity of the direct detection lies in the low contrast of the planets compared to their host star.
- **Microlensing:** as General Relativity dictates, light bends when it travels through deformed space-time by the gravitational effect of bodies. If a planetary system interjects between the observer and another star, the signal received from the distant star increases being able to distinguish the lens effect caused by the host star and the orbiting planet.
- **Transit timing variations (TTV):** by time variations in the transit of an already known planet, the presence of other planets in the system can be inferred.

During the first fifteen years of exoplanet hunt, detections were dominated by the RV method. The precision reached by stabilized spectrographs was improved from 10 m s^{-1} in 1995, to 1 m s^{-1} in 2005 with HARPS (Mayor *et al.*, 2003). In 2018 it became possible to reach the few tens of cm s^{-1} with ESPRESSO (Pepe *et al.*, 2020), being the only spectrograph that allows the detection of Earth-sized planets in Sun-like stars. Nonetheless, it was the transit method that caused the ‘boom’ of the exoplanets. In 2009 the space mission Kepler (Borucki *et al.*, 2009) was launched, which detected about 2400 planets by this method.

The next few years promise great strides thanks to the new generation of missions designed for planet-hunting by NASA and ESA teams. The spacecraft operating since 2018, TESS (**T**ransiting **E**xoplanet **S**urvey **S**atellite, Ricker *et al.*, 2015), is mapping the entire celestial sphere, aiming at detecting thousands of new planets in our galactic neighborhood. Some of the transient planets already discovered by TESS in the southern hemisphere have begun to be monitored with ground-based spectrographs like ESPRESSO, CARMENES, or HARPS (to mention only some examples). Launched at the end of 2019, Cheops (**C**haracterising **E**x**O**planet **S**atellite, Benz *et al.*, 2021), is designed to characterize already known planets. It allows determining the density of planets already observed by RV, and infer the presence of an atmosphere. The HST (**H**ubble **S**pace **T**elescope) successor scheduled to be launched in October 2021, is JWST (**J**ames **W**ebb **S**pace **T**elescope, Gardner *et al.*, 2006). It is the largest space telescope to date and with a high fraction of its observing time dedicated to exoplanets⁵. Its instrumentation is designed to work in the near and mid-IR, which will take direct images of the planets as well as characterizing their atmospheres. PLATO (**P**LAnetary **T**ransits and **O**scillations of stars, Rauer & Heras, 2018)

⁵<https://www.stsci.edu/jwst/science-execution/approved-programs/cycle-1-go>

estimated to be launch in 2026, will focus on studying the characteristics of terrestrial planets in the HZ around Sun-like stars. Finally, also focused on the infrared range and planned to 2029, Ariel (**A**tmospheric **R**emote-sensing **I**nfrared **E**xoplanet **L**arge-survey, [Tinetti *et al.*, 2021](#)) will have all its observing time devoted to this purpose.

1.2 Planets in the Habitable Zone

The yearly exponential growth in the number of detected extrasolar planets confirms that the Solar System is not an isolated case, but one more among the many that exist only in our neighborhood. The question of whether life can develop on any of these planets leads to define the concept of HZ of a host star. This region is the range of orbital distances that a rocky world must be at to hold water on its surface in a liquid state under the assumption of atmospheric pressure.

Since the luminosity of the star is a key feature, depending on its spectral type the HZ will be closer (cold stars) or farther (hot stars). In a first approximation, only the irradiance of the host star is considered a source of energy. However, the greenhouse effect caused by the presence of certain chemical components in the atmosphere of the planet, and internal heating in the case of young or massive planets, are also determining factors. The boundaries of the HZ are commonly delimited by: (1) the **optimistic HZ**, from the so-called current Venus to the young Mars; (2) the **conservative HZ**, from “Runaway Greenhouse” to “Maximum Greenhouse” ([Kopparapu *et al.*, 2013](#)).

The reason for establishing the HZ as a pre-requisite in the search for habitable worlds is that water has proven to be a key piece for life as we know it. The solvent capacity of water facilitates the transport of substances and allows the development of multiple chemical processes, such as photosynthesis. This is why all living creatures we know are made up of water. Moreover, the beginning of life on Earth took place inside this solvent. Thus, it seems reasonable to be the starting point in our search for life beyond our planet (e.g. [Cockell *et al.*, 2016](#)). In Figure 2, we show the number of planets in the HZ discovered per year and method, in which the work carried out by the Kepler mission in 2016 is noteworthy. Although the HZ is the main objective of the KOBE experiment, there is evidence of liquid water beneath the surface of moons in the Solar System (such as Europa, Enceladus, or Titan). Searching for bodies where this kind of oceans can take place, is also of great interest in the Astrobiology field (e.g. [Tjoa *et al.*, 2020](#)).

Obtaining a sample of potentially habitable planets is crucial for the subsequent characterization of their atmospheres. That will allow inferring the presence of water vapor, as well as other molecules that trace life habitats, for example, greenhouse gases like CO₂ or CH₄ (e.g. [Noack *et al.*, 2017](#); [Dorn *et al.*, 2018](#)). However, atmospheric characterizations are carried out by means of photometry during transits (e.g. JWST, Ariel), but worlds discovered by the RV method are most likely not transients. Thus, it will be necessary to resort to alternative techniques, such as space-based interferometers (e.g. LIFE, [Quanz *et al.*, 2021](#)).

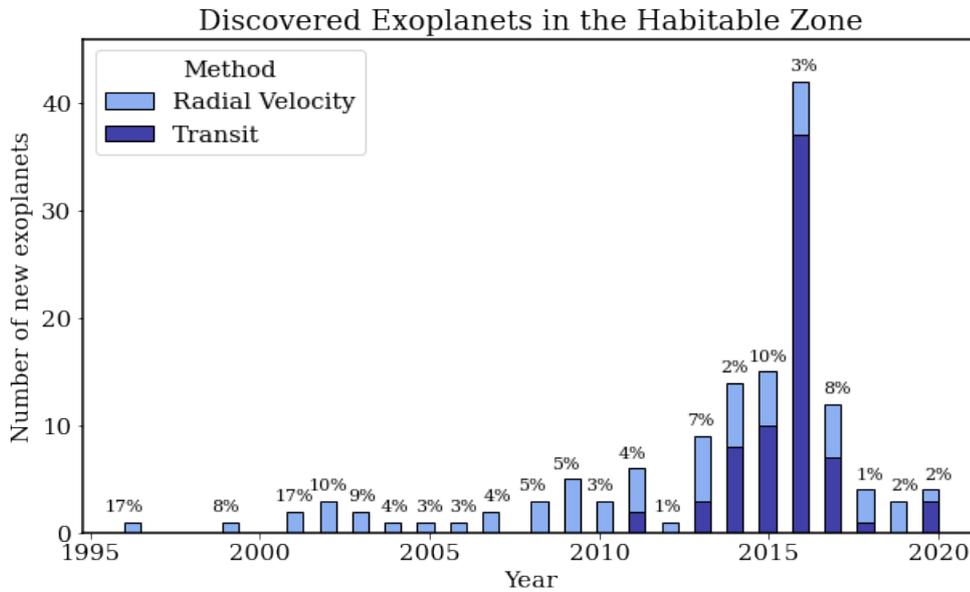


Figure 2: Number of exoplanets discovered per year and method with 100% of its orbit inside the boundaries of the optimistic HZ. Labels above each bar show the percentage of the whole exoplanets detected in the corresponding year. Data used for this graph from *The Habitable Zone Gallery* (Kane & Gelino, 2012) and *The NASA Exoplanet Archive* (Akeso et al., 2013).

1.3 An emptiness around K-dwarfs

The desert around K-dwarfs is an observational bias. Attempts to find rocky planets within the HZ have focused on Solar-type stars (G-dwarfs), and in the smaller and colder M-dwarfs. The former were the main targets for the Kepler mission and the ground-based spectrograph HARPS. The latter takes advantage of the fact that the HZ of less massive stars is closer, thus facilitating the detection of planets around them. This has resulted in a remarkable void of planets detected in K-dwarfs despite being exceptional candidates, as they achieve the balance between both G and M pros and cons.

From the detectability point of view, the planets in the HZ of G-type stars orbit at high distances (optimistic HZ between 0.75-1.77 AU, corresponding to orbital periods of 238-867 days). This represents a challenge for both RV searches (as the RV semi-amplitude decreases with the square root of the orbital distance) and transit techniques (as the probability of transit rapidly decreases as we move far from the star). By contrast, M-dwarfs have their great interest here, as the rocky planets that can retain water on their surface are at very close orbits. The closer the planet is to its star and the more comparable their masses are, the stronger the effect of the gravitational interaction will be inducing larger RV semi-amplitudes.

Another factor to consider is stellar activity as it generates RV signals that can induce false detection of planets, or prevent the detection of a real one. The effects of convective zones on the stellar surface, such as p-mode oscillations and granulation, produce signals between 10-400 cm s^{-1} in time scales around 5 and 25 minutes (Schrijver & Zwaan, 2000). In order to reduce the effects of these phenomena, considerably long exposure times are taken (about 15 min) (e.g.

Oshagh et al., 2017; Dumusque et al., 2011). Apart from that, the presence of active regions such as spots or plages create differences in brightness between one side of the star and the other, resulting in a Doppler shift as the star rotates. These effects are more complicated to deal with as their periods are dozens of days. Dumusque et al. (2011) showed that K-type stars are the ones that present the least variation in RV due to these components, being around 2 m s^{-1} in G2-type stars, and 0.8 m s^{-1} in K2-type stars. Stellar activity dramatically increases for M-types. This fact together with the proximity of its planets supposes a problem for the survivability of life since the flares could cause large variations in the irradiance on the planet.

An additional problem in M-type stars, is that in close orbits the planets are tidally locked such as the Moon, always facing the same side to the body they orbit around. Consequently, there is a large temperature gradient between the day and night sides and this would significantly hinder the development of life (e.g. Auclair-Desrotour & Heng, 2020). Taking all of these factors into account, M-dwarfs although advantageous from a detectability point of view, are not optimal for the search for habitable exoplanets (e.g. Meadows & Barnes, 2018).

K-dwarfs accomplish an excellent trade-off between detectability and habitability conditions. Table 1 provides some of the general characteristics of each spectral type in the main sequence. Stellar parameters from type G2 corresponds to the Sun, K3 to the standard HD 219134, K7 to HD 201092 (61 Cygni B), and M2 to HD 95735 (Lalande 21185). From these data, using the definition of the habitability region given by Kopparapu (Kopparapu et al., 2013) and considering circular orbits in the third Kepler law, we obtain the period that must have a planet residing within the HZ. In the table, we show the RV semi-amplitude that would produce a planet of 5 and $10 M_{\oplus}$ for the range of periods within the optimistic definition of HZ (OHZ). As shown, for late K-dwarfs (K5 to M0), the signal produced for both kinds of planets would be detectable for current spectrographs. Moreover, this kind of star show much less stellar activity, and they are not tidally locked as compared to early M-dwarfs (M2).

	Spectral Type			
	G2	K3	K7	M2
$M_{\star} (M_{\odot})$	1	0.81	0.63	0.46
$L (L_{\odot})$	1	0.26	0.09	0.02
$T_{eff} (K)$	5772	4699	4120	3526
$P_{OHZ} (\text{days})$	238 - 867	103 - 403	54 - 221	21 - 89
$P_{CHZ} (\text{days})$	353 - 817	156 - 380	82 - 208	32 - 84
$K_{5M_{\oplus}} (\text{m/s})$	0.33 - 0.51	0.50 - 0.78	0.72 - 1.15	1.20 - 1.69
$K_{10M_{\oplus}} (\text{m/s})$	0.67 - 1.03	0.99 - 1.57	1.44 - 2.30	2.40 - 3.37

Table 1: General characteristics according to the spectral type of dwarf-stars. From top to bottom: stellar mass, luminosity, effective temperature, the period range for planets in the optimistic (OHZ) and conservative (CHZ) HZ, and the RV semi-amplitude for a planet of 10 and 5 masses of the Earth (M_{\oplus}).

1.4 The KOBE experiment

In response to the need for bridging the gap around K-type stars (see Figure 3), the KOBE experiment started in January 2021 as a legacy survey of Calar Alto (Almería, Spain). Over 5 semesters with a total of 175 nights guaranteed, a minimum of 50 late K-dwarfs (K4-M0) will be monitored using the CARMENES instrument located at the 3.5m telescope. It consists of two spectrographs covering both visible and near-infrared ranges, from 0.52 to 1.71 μm . Its precision is of the order of 1 m s⁻¹ in the visible arm, allowing the detection of planets of 5 M_{\oplus} .

According to [Kunimoto & Bryson \(2020\)](#), based on Kepler survey, each K-type star hosts 2.6 ± 0.3 planets with a period lower than 400 days. Besides, around two thirds of all K-type stars host a planet within the HZ, 17% of those having a rocky composition and being detectable by CARMENES. These occurrence ratios, added to the careful selection of targets made for KOBE, lead us to estimate that the survey will result in the detection of between 14-40 new worlds (1-4 rocky planets in HZ) (Lillo-Box, private communication). Although the main objective is to find potentially habitable terrestrial planets, also gas giants are of scientific relevance: due to the subsequent atmospheric characterization with surveys such as JWST (e.g. [Danielski *et al.*, 2018](#)), for the search of co-orbital worlds (e.g. [Lillo-Box *et al.*, 2018](#)), or even a future hunt for habitable exomoons (e.g. [Tjoa *et al.*, 2020](#)).

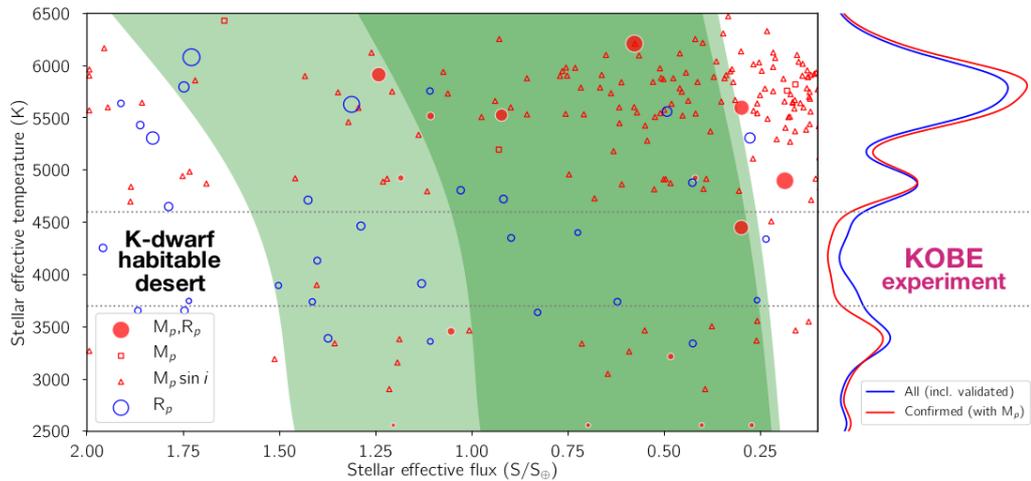


Figure 3: Optimistic and conservative Habitable Zone (light and dark green regions) for different stellar types. On the right, distribution of discovered planets, where the K-dwarf habitable desert can be seen. Source: KOBE experiment proposal Lillo-Box *et al.*, private communication.

1.4.1 Blind-search surveys: observational strategies

For the characteristics of the KOBE experiment, each star requires an average of 90 RV data points distributed over time to detect a rocky planet of 10 M_{\oplus} inside the HZ. One of the great disadvantages of ground-based observatories is the presence of the atmosphere and, therefore, the dependence on meteorological conditions. Due to the preciousness of observing time, it is essential to design observing strategies that allow maximizing the efficiency of the observations to achieve the objectives set by the project.

Uniform in-phase

With the objective of highlighting how relevant it is to design a proper observational strategy in terms of gathering RV data, we refer to [Cabona *et al.* \(2020\)](#). In that study, a comparison is made between three different scheduling techniques. The most basic one selects the dates randomly. The second provides a more balanced schedule between targets and ensures greater uniformity of RV data throughout the orbital phase. The third compares all possible schedules throughout the survey and chooses the one that gives more uniform in-phase measurements (as the second does). The last one is called a non-myopic strategy by the authors as it can find the best solution and not just one that accomplishes the constraints.

The results of the study through multiple simulations show that the last two schedulers, those that provide more uniform data in-phase, are significantly better. This conclusion is based on the capability to obtain a less biased, more accurate and precise value for the planetary mass than in the case of the random scheduler. However, there is no noticeable difference between these two (myopic vs. non-myopic). The reason might be that the selected temporal extension they choose (3 years) is very large and therefore the strategies end up being compensated. Although, it is expected that at smaller scales the non-myopic strategy will provide better results.

CARMENES scheduler

The CARMENES survey has already implemented an automatic scheduling tool designed for RV experiments that KOBE is making use of, CAST ([García-Piquer *et al.*, 2017](#)). This tool is capable of planning in the most efficient way the different targets of a survey, fulfilling certain constraints and aiming at taking advantage of 99% of the telescope time in good weather conditions, including overheads. It consists of several phases: the long-term, mid-term and short-term strategies. The first creates the observing plan for the entire survey sample. The second specifies the observations to carry out before starting the night finding the schedule with the best trade-off between all the parameters at stake. And the last one is in charge of deciding the very next observation. In this final step, telescope and software are communicated updating in real-time the previous observations and the atmospheric conditions.

The long-term scheduler fulfills the hard constraints, such as the elevation or visibility duration. Then, it makes a plan that optimizes the so-called soft constraints, for instance, the cadence. The cadence is, even a soft constrain, of great relevance in the RV method since the selection of certain strategic times can favor the detectability of a planet. The KOBE experiment selects an initial cadence for each star corresponding to 10% of the period in the middle of the HZ. This would guarantee to sample uniformly the in-phase RV curve for the main purpose of the survey (detecting planets inside the HZ). Once there are a sufficient number of points to see an emerging peak in the periodogram, it will be chased until it is unraveled whether it is a real period produced by a planet or whether it is spurious. This is one of the objectives of this Master Thesis, as we will see below.

1.5 Objectives of this work

Up to this point, we have emphasized the importance of a good observing strategy. The KOBE experiment is an ambitious project, and its success requires the careful selection of targets that meet the requirements that favor detectability and host suitability. Once the observations have started, optimizing the efficiency of the observing time is crucial to detect the planets as soon as possible. This leads us to the motivation to carry out this work, in which the two main objectives are:

1. The selection of the 50 stars to observe during the next 2.5 years with CARMENES.
2. The development of a Bayesian-based algorithm to optimize the observing strategy for more efficient planet detection.

2 Selection of the KOBE sample

Starting from a pre-selection of 200 candidate stars fulfilling the requirements of the scientific project, we choose 50 of them as the final KOBE sample. The criteria established for the long-list accomplish the following criteria: close-by and late K-dwarfs (K4-M0), with stable TESS light curves that do not indicate the presence of binary companions or strong stellar activity. We sort this list attending to criteria such as magnitude, effective temperature or visibility from Calar Alto. To reach the final target list, we study these candidates in more detail. Below, we outline the criteria to discard those candidates that are not optimal. We carry out the procedure using Spanish Virtual Observatory tools (SVO⁶) and by searching in archives of the main High-Resolution spectrographs, such as SOPHIE or HARPS.

First of all, we ensure that the KOBE experiment is focused on the detection of **new habitable worlds**. Thus, stars in which exoplanets have already been discovered or those highly monitored by other experiments are ruled out.

Another reason for discarding targets lies in the nature of the technique used, **avoiding face-on systems**. As previously mentioned, in the RV method the orbital inclination angle of the planet must be different from 0° to infer a detectable signal. Since what is measured is the projection of the velocity in the radial direction, if the orbit is perpendicular to the line of sight there would be no such component. As the planets generally orbit perpendicular to the spin axis of their parent star (e.g. [Albrecht et al., 2021](#)), the strategy is to reject stars with a null projection of their spin ($V \sin i_\star^7 = 0$).

For correct measurement of the RV, the **absorption** lines must be **narrow**. Stars of high rotation speed have very broad spectral lines making them less contrasted and overlapping, making it impossible to measure the Doppler shift produced by orbiting planets (e.g. [Lovis & Debra, 2011](#)). For this reason, we also avoid stars with $V \sin i_\star > 8 \text{ km s}^{-1}$.

⁶<http://svo.cab.inta-csic.es>

⁷Where i_\star is the angle between the spin axis of the star and the plane containing the direction of sight.

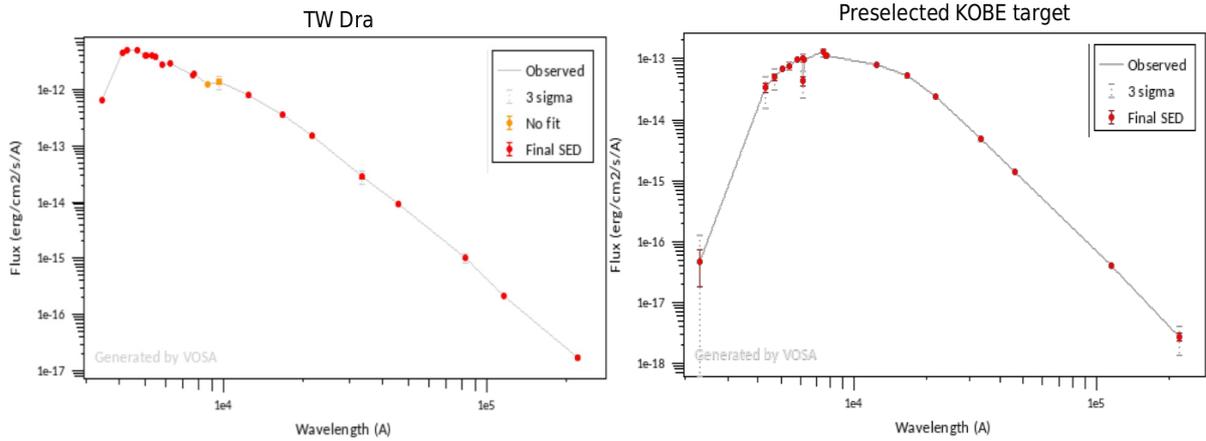


Figure 4: SED comparison of a binary system and an isolated star. **Left:** binary system TW Dra of A-K type stars. **Right:** pre-selected K-dwarf for the KOBE sample. The SVO tool VOSA is used to obtain these graphs.

Another constrain to impose is the selection of **unaccompanied stars**. To verify the absence of companion stars, we study the spectral energy distributions (SEDs) in order to figure out if they are contaminated by additional sources with significantly different spectral types. If we notice an excess of flux at shorter wavelengths (UV) it would indicate the presence of a hot companion, while at the longer (IR) it would unveil a colder one (e.g. [Chulkov *et al.*, 2015](#)). In [Figure 4](#), we show the comparison of two SEDs. One of them corresponds to the binary system TW Dra ([Tkachenko *et al.*, 2010](#)), composed of two stars of types A and K. The other shows one of the KOBE candidate stars. In the former, an excess of flux at high frequencies is evident, explained by the energy emitted by the hot companion of type A, while in the latter there is a clear black body spectrum without contamination.

Finally, we make sure of selecting a **clean field of view**. From the search of the pre-selected list in Gaia EDR3 (Gaia Early Data Release 3, [Gaia Collaboration, 2021](#)) some are found to have sources at close angular distances. For the cases where this distance is less than the size of the CARMENES fiber (1.5 arcsec), this is a compelling reason to discard the stars since the spectra would be contaminated by the objects in the field of view.

After this filtering, the top 50 candidates in the ranking are selected as the final KOBE sample, which has already begun to be monitored since January 1st, 2021. Since the vast majority are of K7 type, considering the mass and effective temperature typical of these stars, we make an estimation of the expected RV semi-amplitude (K) for planets residing within the HZ. In [Figure 5](#), we show the histogram of the minimum and maximum K values for 5 and 10 M_{\oplus} planets, corresponding to the outer and inner boundaries of the optimistic HZ. In view of these values, it is possible for CARMENES⁸ to detect rocky planets of even 5 M_{\oplus} with the possibility of retaining liquid water on their surface.

⁸CARMENES precision in the visible range is around 1 m s^{-1} .

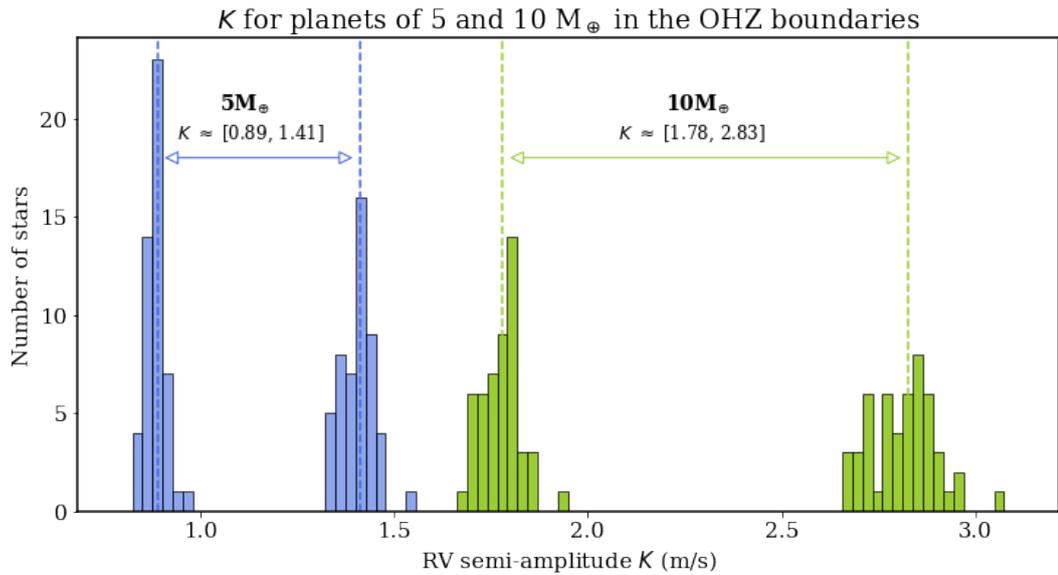


Figure 5: Histograms of the RV semi-amplitude (K) that a planet would produce in the boundary edges of the optimistic HZ for the KOBE sample. In blue the limit values for a planet of 5 and in green of 10 M_{\oplus} . For both cases, the ranges of RV semi-amplitudes in which is expected to detect planets of these masses are indicated.

3 KOBESim: a Bayesian strategy

Parameter inference is essential in Astrophysics and in particular in exoplanet studies since the only way to access exoplanet properties is indirectly through observations. To perform this task, Bayesian statistics combines both inductive and deductive logic to explore all the information available at each state of knowledge. This perspective has demonstrated its power by improving the estimation of the parameters in model inference by several orders of magnitude compared to conventional statistics (also known as frequentist) (Gregory, 2005, Chapter 13). This statistical perspective also offers the possibility to compare competing models by means of the incorporation of the concept of probability of hypothesis (e.g. Loredano, 1992).

KOBESim is the strategy developed for the KOBE experiment to improve the efficiency in the process of gathering RV data in a long-term blind-search survey. It is based on Bayesian statistics and makes use of Markov Chain Monte Carlo (MCMC) simulations. Throughout this section: we show the fundamentals of Bayesian inference (Section 3.1), and we detail the objectives of this strategy, its development, and the results obtained (Section 3.2).

3.1 Bayesian inference

First of all, we have to bear in mind what probability means. Bayesian statistics define it as a measure of the degree to which a proposition is supported by the information in hand. This concept is known as subjective probability (Ramsey, 1931). In contrast to frequentist probability, which only considers the repetition of an experiment, the subjective also assumes the dependence on the uncertainty state. However, being subjective does not mean there is arbitrariness, but

rather the probability distributions are conditional on prior knowledge. Let us consider an experiment to measure the mass of a particle: Bayesian statistics would make use of the prior that it must be positive, which would be ignored by the conventional statistics. Although simple, it is a great example to reflect this concept and understand that it is not a subjectivity against the scientific method, but supported by it. In this way, scientists with the same knowledge must reach the same conclusions (Sivia & Skilling, 2006).

The inference is the process of going from the effects to the cause. This can be achieved by means of the Bayes Theorem, which describes conditional probability:

$$p(\theta|d, M) = \frac{\mathcal{L}(d|\theta, M)\pi(\theta)}{\mathcal{Z}} \quad (3.1)$$

The **posterior** $p(\theta|d, M)$, is the probability of the set of parameters (θ) following a model (M) and given the data obtained in the measurement (d). $\mathcal{L}(d|\theta, M)$ is the **likelihood** function, which describes the distribution that follows the measures. $\pi(\theta)$ is the **prior** probability, expressing the knowledge about the parameters before the measurement. The denominator factor, \mathcal{Z} , is the **evidence** or fully marginalized probability function. The latter is the normalization factor that converts the posterior into a probability.

It is possible to remain only with the parameters of interest (θ_i) through marginalization. This process consists of integrating those nuisance parameters (θ_j):

$$p(\theta_i|d, M) = \int \left(\prod_{j \neq i} d\theta_j \right) p(\theta|d, M) = \frac{\mathcal{L}(d|\theta_i, M)\pi(\theta_i)}{\mathcal{Z}} \quad (3.2)$$

The complexity of the integrals that appear when marginalizing increases exponentially with the number of parameters (curse of dimensionality). In fact, the biggest complication comes when calculating the evidence, as it is the marginalization over all the parameters (equation 3.3). To avoid having to solve them, MCMC simulations are used to sample the marginalized probability function enabling to obtain the parameters of interest (e.g. MacKay, 2003).

$$\mathcal{Z} = \int d\theta \mathcal{L}(d|\theta, M)\pi(\theta) \quad (3.3)$$

Indeed, evidence is a key element in the comparison of models. Although the value of the evidence does not provide information by itself, the ratio between those of competing models informs about how much more evident one model is over another based on available data. This fraction intrinsically penalizes the most complex models (those with more free parameters), which is known as Occam's razor (e.g. MacKay, 2003; Thrane & Talbot, 2020). Mathematically, the ratio between the posterior of each model (the odds) is:

$$\mathcal{O}_{12} = \frac{\pi(\theta_1)\mathcal{Z}_1}{\pi(\theta_2)\mathcal{Z}_2} = \frac{\pi(\theta_1)}{\pi(\theta_2)} BF \quad (3.4)$$

Where the evidence ratio is called the **Bayes Factor** (BF), and subscripts 1 and 2 identify each of the two competing models. In this particular work, we use this factor to claim the detection of an exoplanet. We infer the model from which our data was sampled and compare it with the corresponding one without any planet (null hypothesis). We set the limit to consider a planet detection at $\ln(BF) = \ln(\mathcal{Z}_{planet}) - \ln(\mathcal{Z}_{null}) > 6$. This will be seen in depth in the next subsections.

3.2 KOBESim

Generally, the more RV observations we get for a given target, the easier will be to claim or discard the presence of a hypothetical planet that generates the wobble of its parent star. However, we have already seen that it is possible to optimize this process by reducing the total number of observations through a good scheduling strategy. As [Cabona *et al.* \(2020\)](#) proposed (Section 1.4.1), uniformly distributed data along the RV curve in-phase, favors efficiency. As a consequence, observing in an already explored phase is not going to give much information (i.e., it will not increase or decrease our evidence against the null hypothesis). The algorithm we present here, KOBESim, is not only a strategy to ensure uniformity on the data, but also to select the optimum observing date that is expected to increase the most the Bayesian evidence in favor of the presence of a planet (assuming it exists).

3.2.1 Goals and architecture

The ultimate goal of the KOBESim algorithm is to accelerate the process of detecting (providing statistical significance of a given model) or discarding a hypothetical planet. The procedure we follow to achieve this, is to select the next observing date by maximizing the evidence increase of the presence of a planet over the null hypothesis. KOBESim code is written in Python language, and it is structured as follows.

Input

There are four fields accepted as input. Two of them being mandatory: data file and period; and the other two optional: conjunction time and a boolean to prioritize observing on close dates.

- **file:** fits format containing the data. Its columns must be: observing date in Julian days (JD), RV in meters per second (m s^{-1}) and uncertainty of RV in the same units.
- **P:** period to be pursued in days. Once there are enough data points to start discerning a sinusoidal shape, a period can be estimated by periodograms (e.g. [VanderPlas, 2018](#)).
- **t₀:** In case it has also been feasible to estimate the conjunction date, it is possible to give it as an input parameter in JD units. This influences the priors that are taken to obtain the posterior distributions of the orbital parameters, which we are going to discuss later in this section.
- **beta_bool:** boolean to reduce the time between observing dates. In long-period RV curves, the largest $\ln(BF)$ increase may occur at a very distant date, which is against the

efficiency of the observations (although in a low number of observations, the detection could be greatly lengthened temporarily). To prevent this situation, we select `True` value for this input to weight the differences of the $\ln(BF)$ with a beta function:

$$\Delta \ln(BF)_w = \beta(x = \Delta t, a = 1, b = 5) [\ln(BF_{n+1}) - \ln(BF_n)] \quad (3.5)$$

Where Δt is the difference in days between the observation and the current date normalized on all candidate dates, and n denotes the number of observations at the moment of running KOBESim. We choose the arguments of the beta function, a and b , in such a way that to select a date at 40% of the period, the increase in the $\ln(BF)$ should be 5 times greater than for the immediate next day. The default value of this input is `True`.

Inside the code

We can differentiate 3 steps in this algorithm: inference of the orbital parameters, estimation of Bayesian evidence to obtain the BF, and the selection of the phase that most increases the BF. Each of these parts are detailed below.

1. We make the inference on the free parameters that describe the RV model (see Appendix A to see the definition of the orbital parameters⁹): V_{sys} , K , t_0 , P and *jitter*¹⁰. Meanwhile, in the case of the null hypothesis the only free parameters are V_{sys} and *jitter*. For this purpose, we execute a MCMC simulation using the *emcee* code (Foreman-Mackey, 2013). For its use, it is necessary to define a likelihood function and the priors in order to sample the shape of the posterior (equation 3.1). Considering that data will be normally distributed around the theoretical value of the model, our selection of likelihood function is a Gaussian-noise:

$$\mathcal{L}(d|\theta, M) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2} \frac{d - M(\theta)}{\sigma^2}\right) \quad (3.6)$$

Where the model M is the RV equation (A.3) adding the *jitter*, and θ is a sample of the free parameters. Regarding to the priors, the selection we make for each parameter is shown in Table 2. We select a uniform distribution for the non-informative priors and a narrow normal for the informative ones (like the P and, if given as input, also t_0).

To the performance of *emcee* we use 30 walkers (chains) with 15 000 steps each of them (iterations), of which 10 000 steps are in the warm-up phase. As a result of the sampling, we obtain the posterior marginalized distributions of the parameters, that can be translated to median values and associated uncertainties (68.7% confidence interval).

⁹Note that we have assumed circular orbit ($e = 0$) and edge-on system ($i = 90^\circ$). For a more general case, the parameters i , e and ω have to be added to the model.

¹⁰Variations in the RV signal, including instrumental noise and stellar activity.

Parameter	Prior type	Prior	Units
V_{sys}	Uniform	$[-100, 100]$	m s^{-1}
K	Uniform	$[0, 1000]$	m s^{-1}
P	Normal	$N(P_{input}, 2)$	days
t_0	Normal or Uniform	$N(t_{0,input}, 2)$ or $[t_1, t_1 + P_{input}]$	JD
<i>jitter</i>	Uniform	$[0, 100]$	m s^{-1}

Table 2: Prior selection to perform the MCMC. P_{input} is the period of a hypothetical planet given as an input in the code. In the case of t_0 , it can also be given as an input ($t_{0,input}$), or it can be a non-informative prior using a uniform distribution as indicated, where t_1 is the first day the target was observed.

- MCMC does not calculate the Bayesian evidence but obtains the shape of the posterior without normalizing. For this reason, we employ *bayev* code (Díaz et al., 2018) that uses the estimator defined in Perrakis et al. (2014). Giving as input a fraction of the posterior distributions, and the functions defined for the likelihood and the priors, we obtain the $\ln(\mathcal{Z})$ distribution. Carrying out this procedure for the two competing models (planet and null), we get an estimation of $\ln(BF)$ with its corresponding uncertainty.
- We predict the $\ln(BF)$ that would be obtained if observing in each orbital phase, corresponding to future observing dates, of the hypothetical planet. For this, we divide the period into `N_cand_t` sub-phases. Next, a date is matched to each sub-phase interval from the KOBESim experiment schedule (next observing nights assigned), not exceeding 3 months apart from the current date. These are what we call candidate dates. Finally, we predict the RV data using A.3 equation. Parameters introduced in this equation are those previously estimated with the MCMC inference. We take the uncertainty of this simulated value of RV as the mean of the RV errors of the previous measurements.

We want to note that we are optimistic about the uncertainties of $\ln(BF)$. We do not calculate the RV from all the posterior distribution of the parameters V_{sys} , K , t_0 and P , but from its median value. The reason is, this would be very computationally expensive, and this approximation is acceptable since we are interested in the relative value between the evidence of each hypothesis.

Output

As a result of running KOBESim, a csv file is delivered as output. Each row corresponds to a candidate future observing date ranked by preference: from highest to lowest $\Delta\ln(BF)$, or $\Delta\ln(BF)_w$ if `delta_bool` is `True`. For a more illustrative inspection of the results, KOBESim also returns a plot of $\ln(BF_{n+1})$ versus orbital phase.

The columns given in the csv file are: `Calendar_date` (format year-month-day), `JD`, `phase`,

lBF (expected logarithm of the Bayes Factor after observing at the candidate date), its uncertainty σ_{lBF} , δ_{lBF} ($\ln(BF_{n+1}) - \ln(BF_n)$), and its uncertainty $\sigma_{\delta_{\text{lBF}}}$. We show a couple of examples in Appendix B.

3.2.2 Testing KOBEsim on simulated datasets

In an early stage of the project in which the observations had not yet begun, we developed a dataset simulator to test the algorithm. KOBE data started to be gathered on January 1st 2021, but we did not have enough real data to run the algorithm until April 2021. Hence, we needed to create realistic datasets to test the algorithm. In this section, we describe the steps followed for this purpose.

Time simulation

To generate synthetic data, the first step is to simulate the observing dates. For convenience, we use Julian days as it is a continuous measure. The considerations we take into account are:

- **Twilight:** the observing time must be between the astronomical dusk and dawn from CAHA, when the Sun is at 18 degrees below the horizon.
- **Altitude:** the elevation of the target over the horizon must be greater than 40° during the exposure time to avoid large chromatic distortions and extinction due to the atmosphere.
- **Exposure time:** in order to achieve the signal-to-noise ratio established by the scientific objectives ($S/N \sim 100$), and also to counteract the effects of stellar activity of smallest time scales (e.g. Dumusque et al., 2011), the exposure we consider is 700 s.
- **Good weather conditions:** as bad weather conditions prevent from observe, we assume that 70% of the nights meet the appropriate weather conditions (i.e., the dome will be opened) to perform the observations.
- **Guaranteed time:** at the time of developing the simulator, KOBE schedule is known until December 2021. Thus, we consider a probability of 55% for a night beyond 2021 to be assigned to the project. This high fraction of time is due to the fact that KOBE shares nights with other programs (e.g. CARMENES-legacy).

Radial Velocity simulation

The second step is the simulation of the RV measurements. The equation that describes the wobble of the star (A.3) has the following ingredients: the systemic RV (V_{sys}), the orbital period (P), the conjunction time (t_0), the RV semi-amplitude (K), and the eccentricity (e).

The first three parameters are drawn from uniform distributions. Meanwhile, we calculate K with A.4, which requires three more parameters: the stellar mass (M_\star), the planetary mass (M_p), and the orbital inclination (i). We assume M_\star equal to $0.5 M_\odot$ since it is the mean value for the spectral type of the KOBE sample. For M_p , we give a value manually while exploring different

scientific cases (5, 10, 20 and 60 M_{\oplus}). Finally, to simplify the problem as in the KOBESim code, we assume $i = 90^\circ$, and $e = 0$. Below in Table 3, we show the parameter selection that we just mentioned.

V_{sys}	P	t_0	e	M_{\star}	M_p	i
$[-10, 10] \text{ m s}^{-1}$	$[50, 120] \text{ days}$	$[t_1, t_1 + P] \text{ JD}$	0	$0.5 M_{\odot}$	$(5, 10, 20, 60) M_{\oplus}$	90°

Table 3: Values of the orbital parameters used for the RV simulation. V_{sys} , P and t_0 take a value within the indicated range, while M_p value is selected manually from the four options given.

We decide the boundary values for P to cover the properties of the KOBE sample. As Table 1 shows, those values correspond to the HZ of late-type K-dwarfs. It is also remarkable that the t_0 range is dependent on both, first observing date (t_1) and P .

To the resulting RV we add a white noise of $\pm 3 \text{ m s}^{-1}$, mimicking instrumental noise and stellar activity variations. Finally, we consider a Gaussian uncertainty associated with the simulated RV data, with a mean 3 m s^{-1} and a standard deviation of 0.3 m s^{-1} . In Figure 6 we show an example of a RV curve obtained with this simulation.

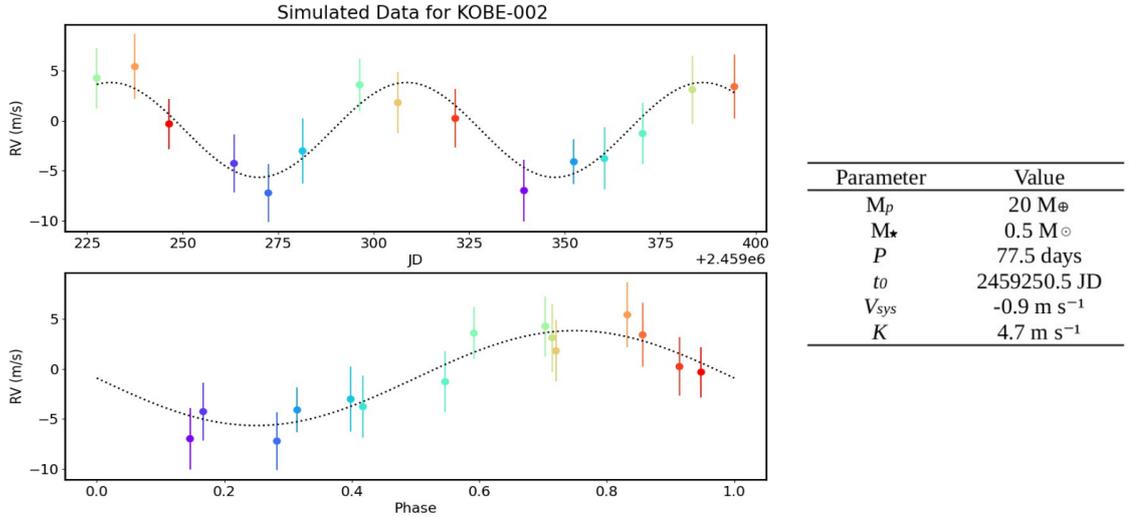


Figure 6: RV curve of a simulated planet-star system over the first semester (from January to June 2021). Data points correspond to observations and the dotted line to the model. The color-code of the symbols indicates the orbital phase from 0 (blue points) to 1 (red points). In the upper panel the x-axis is the Julian Date (JD), whereas the lower panel is phase-folded, where the x-axis corresponds to the phase. At the right, values of the parameters are collected in a table.

Running KOBESim

We use the above described simulation tool to generate simulated observations for one of the KOBE targets. In particular, we use the stellar properties of KOBE-002 because it is a bright target and it would get a relatively large number of observations before the end of this Master

Thesis. We simulate the first 10 observations for this target and assume that a periodicity of $P = 59$ days is appearing in the RV periodogram. This is thus our target periodicity to test with KOBESim. The goal now is to predict the best observing date for the next data point (the 11th in the time series) in order to speed up a possible planet detection at that periodicity. In Figure 7, we show the simulated dataset once the parameter inference is performed for both competing models.

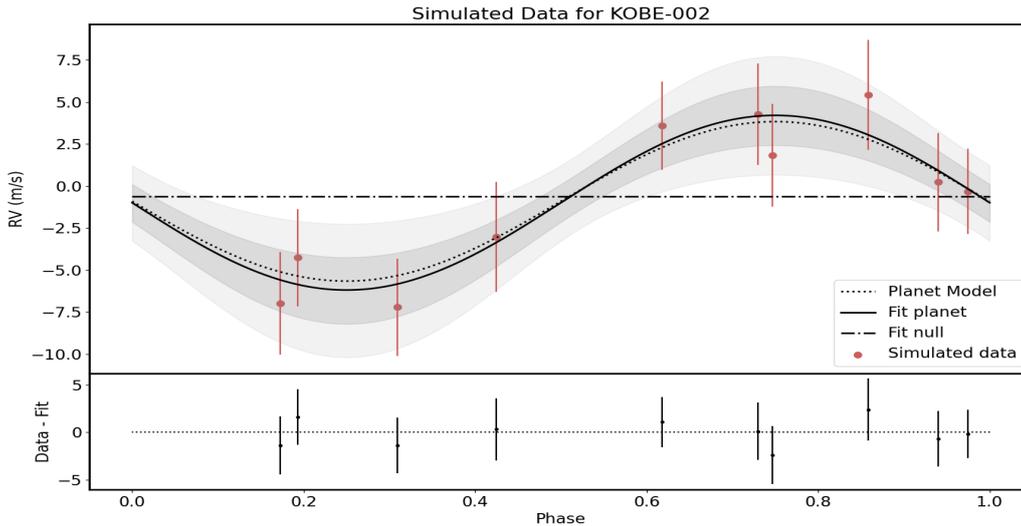


Figure 7: Fit of the simulated RV data for KOBE-002 star. Solid black line is the fit for the model with planet at the given period. Dashdot line is the fit for the null hypothesis. Dashed line is the true model used to simulate the data points (in red). Shaded region shows the confidence interval at 1σ (dark gray) and 2σ (light gray) of the model with planet. Lower panel shows the residuals of planet model fit.

These two models, allow us to quantify how well supported is the planet model against the null hypothesis with the available data. KOBESim, now generates a new synthetic data point and adds it to the previous ones, calculating the increase in BF. This process is repeated independently for different phases. In this case, we compare a total of 20 candidate orbital phases ($N_{\text{cand}_t} = 20$). Once all of them are calculated, KOBESim makes the decision of which, among all the candidate dates, is the best to observe again according to the current state of knowledge.

We show, in Figure 8, the output plot returned by the code. The difference in $\ln(BF)$ of the 1-planet model against the null hypothesis is shown for each phase. It can be noticed some large gaps between phases, and in fact, there are less than the 20 points that we indicate as candidates. This is because in the next 3 months this target cannot be observed in these phases: either because the corresponding dates are not assigned to observe, or because the star will no longer be visible from CAHA. As in this example we are not using the beta function, the criterion for selecting the best observing date is, simply, the one that maximizes the increase in $\ln(BF)$ regardless of how far that date is. This occurs for a phase around $\phi = 0.2$, achieving a $\Delta\ln(BF) = +0.75 \pm 0.09$.

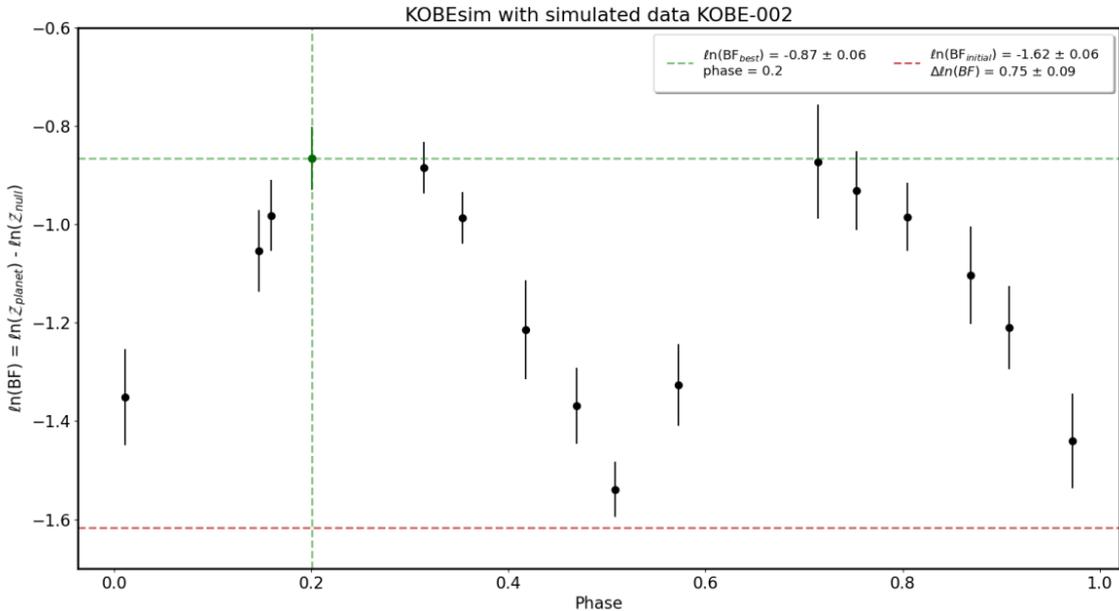


Figure 8: Output graph of KOBESim for the KOBE-002 simulated data. The y-axis shows the logarithm of the expected Bayes Factor and x-axis the associated orbital phase. The legend shows the phase selected as the best option, as well as the corresponding expected value of $\ln(BF)$ (green lines), its increment, and the value before adding that future observation (red line).

In Appendix B we show the analogous case selecting `beta_bool = True`, with both KOBESim outputs (Figure 15, and Table 4 with the csv file). Also, we present in there the MCMC simulation results. The posterior marginal distributions as well as the so-called corner plot showing the parameter dependency (Figure 16 for 1-planet model, and 17 for the null hypothesis).

KOBESim vs Monotonic cadence strategies

To study the efficiency of KOBESim, we estimate how long it would take to detect planets of different masses inside the HZ for the particular case of the KOBE experiment. And we compare this with the time (and number of observations) that a monotonic cadence approach (i.e., obtaining observations every N days) would require. For this purpose, we simulate future observations for KOBE-002, cumulatively, until reaching enough evidence from the planet model over the null hypothesis to claim the detection. We have set this limit at $\ln(BF) > 6$. We do this procedure for both KOBESim modes (with and without beta function), and without using this strategy (spacing the observations with the cadence assigned to this target (6 days), as long as that day meets good weather conditions and is an assigned day, otherwise it is postponed to the next plausible day).

In Figure 9, we show the result for a planet of $20 M_{\oplus}$ and with 10 initial data points. In this figure, we see the gradual increase in $\ln(BF)$ that each strategy follows as a new observations are added following the different approaches. In this specific case, we see that the improvement in terms of the number of observations is not very significant: 20 and 21 observations unweighted

and weighted by the beta function, and 25 with the monotonic cadence approach. However, if we draw our attention to the number of days that are needed (displayed in the legend) since the beginning of gathering data, the improvement is highly noticeable: 136 days if using KOBESim with beta function versus 343 days if KOBESim is not used. This result, although very positive in favor of KOBESim, must be viewed with caution. We know this target sets in July, and it will not be visible from CAHA for 4 months. In the case of using the monotonic cadence approach, the detection would be extended after this period in which we cannot take data, thus around 120 of the days are not useful.

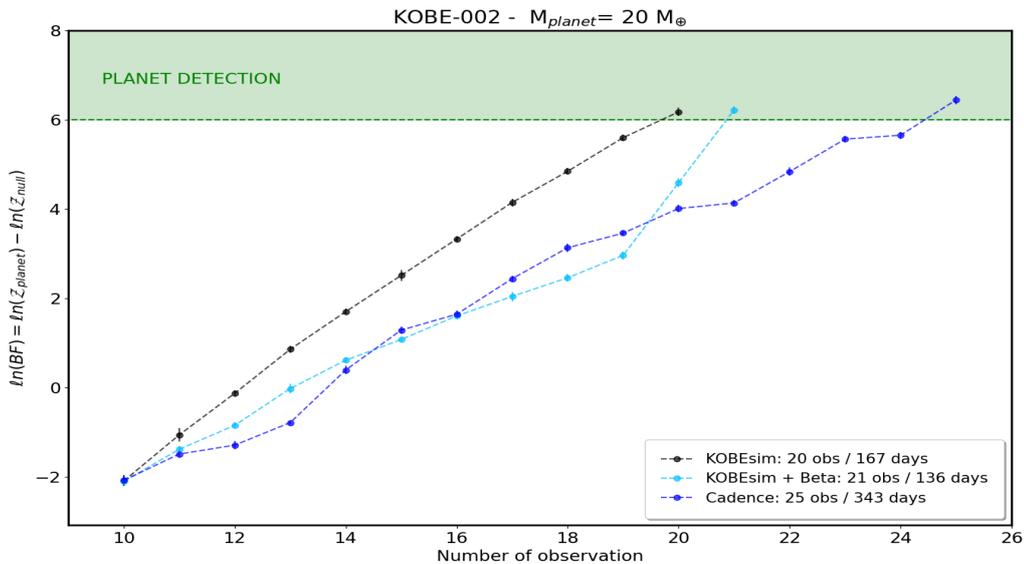


Figure 9: Prediction in the evolution of the logarithm of the Bayes Factor from the simulated data of KOBESim-002, for a $20 M_{\oplus}$ planet with $P = 59$ days. The number of observations and the time it would take to detect the planet, are compared using three different strategies: KOBESim, KOBESim with the beta function, and spacing the observations a fixed cadence (6 days for KOBESim-002).

It seems interesting to ask ourselves what would happen in the detectability limit case for CARMENES. Repeating the same procedure for a simulated planet of $5 M_{\oplus}$ and with $P = 52$ days, we obtain the results of Figure 10. In this case we assume the case of a circumpolar star, never setting at the latitude of CAHA. Although this assumption is not very realistic, it is imposed due to how challenging this objective is. In this case, the usefulness of KOBESim is evident, reducing the number of observations by a factor 1.7. However, the numbers are still too high, especially the required timespan to detect the planet. To be able to reach a goal like this in the KOBESim experiment, we would have to reduce the uncertainties of the RV data. We could increase the exposure time thereby increasing the signal-to-noise ratio. This, would allow going from the 3 m s^{-1} uncertainty, to even a value close to 1 m s^{-1} . This would considerably increase the speed of detection, maybe enabling to detect a planet of these characteristics in the period of time that the experiment lasts. In practice, this would only be feasible with the brightest targets due to the maximum exposure time allowed by the CARMENES instrument is 1800 s. In Appendix B we show the results of this study reducing the mean of the uncertainties to 1.5

m^{-1} to a circumpolar star with a $5 M_{\oplus}$ planet (Figure 18), showing that in just 53 observations and in less than a year, the detection could be possible. For the case of the stellar properties of KOBE-002, we repeat this procedure for planets of $10 M_{\oplus}$ (number of days reduced by a factor 1.9) and $60 M_{\oplus}$ (not improvement is noticed) (Figure 19 and Figure 20).

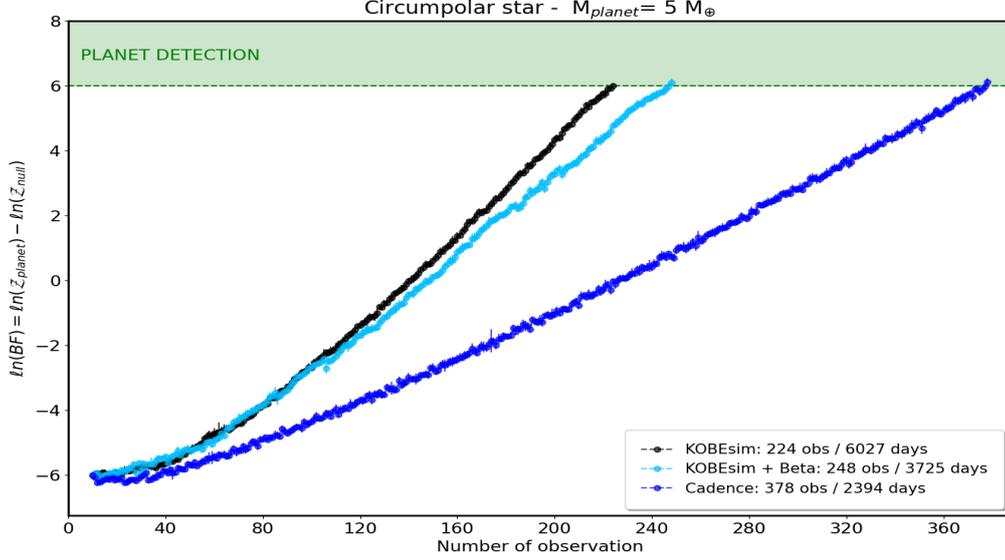


Figure 10: Prediction in the evolution of the logarithm of the Bayes Factor from the simulated data of a circumpolar star, for a $5 M_{\oplus}$ planet with $P = 52$ days. The number of observations and the time it would take to detect the planet, are compared using three different strategies: KOBESim, KOBESim with the beta function, and spacing the observations a fixed cadence (6 days).

3.2.3 Application to real data: the case of KOBE-002

From April 2021 we began to have enough real data points of KOBE-002 to test the algorithm. At the time of writing, there are 10 observations for this star. The study of its periodogram shows an emergent signal at a period inside the HZ for this star, $P = 59.4$ days. For this reason, this target tops the priority in the list of observables.

Applying the code on these data, we obtain the fit shown in Figure 11, where we see that the inferred RV semi-amplitude is about 4 m s^{-1} (considering $M_{\star} = 0.5 M_{\odot}$, and $i = 90^{\circ}$, it could be a Neptune mass planet, $\sim 16 M_{\oplus}$). In case of not using the beta function (unweighted option of KOBESim), we show the result of KOBESim, searching for the best next observing date, in Figure 12. As we can see, the result favors the uniformity of the data in the in-phase curve, as there is a greater increase in evidence of the model with planet over the null hypothesis for the phases that have not yet been observed: $\phi = 0.1 - 0.3$. Since this target will set soon (in July), in this case it is especially interesting to prioritize the closest dates. In Appendix B, we show the case using the beta function, and its table including the prioritization of the observing dates (Figure 21 and Table 5). Posterior distributions for the orbital parameters are shown in Figures

22 (planet hypothesis) and 23 (null hypothesis).

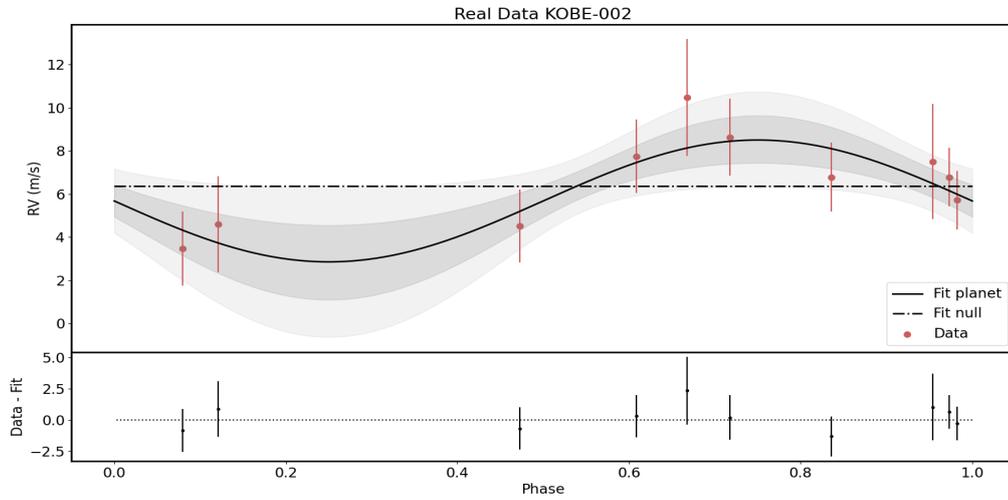


Figure 11: Fit of the real RV data of KOBE-002. Solid black line is the fit for the model with planet at the given period. Dashdot line is the fit for the null hypothesis. Shaded region shows the confidence interval at 1σ (dark gray) and 2σ (light gray) of the model with planet. Lower panel shows the residuals of planet model fit.

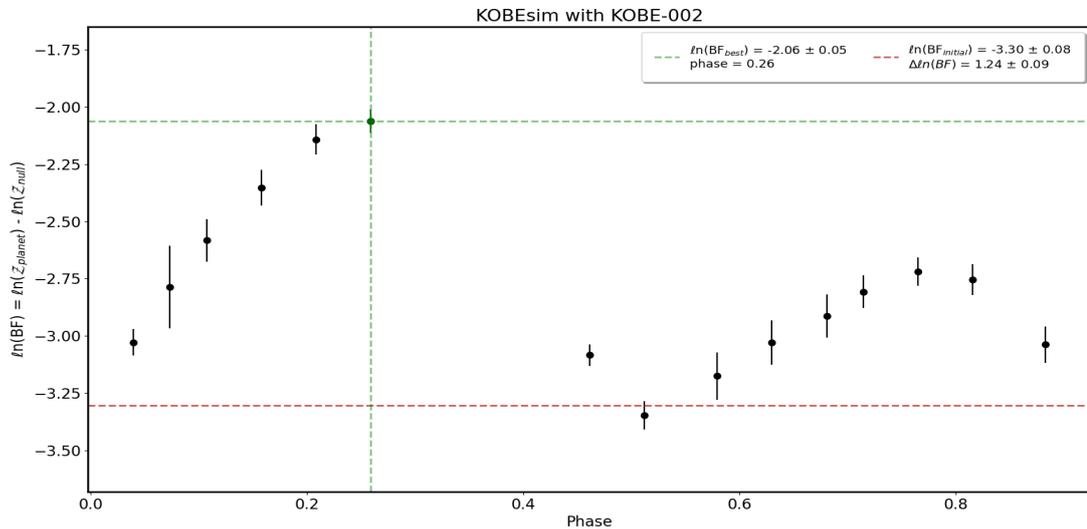


Figure 12: Output graph of KOBESim for real RV data of KOBE-002. The y-axis shows the logarithm of the expected Bayes Factor and x-axis the corresponding orbital phase. The legend shows the phase selected as the best option, as well as the corresponding expected value of $\ln(BF)$ (green lines), its increment, and the value before adding that future observation (red line).

Likewise the previous case with the simulated data, we now estimate how many observations will be necessary to detect the hypothetical planet in case the trend continues. In view of Figure

13, the predictions seem optimistic indicating that the possible detection could occur in twenty more observations. Even before the setting of the target in mid-summer 2021.

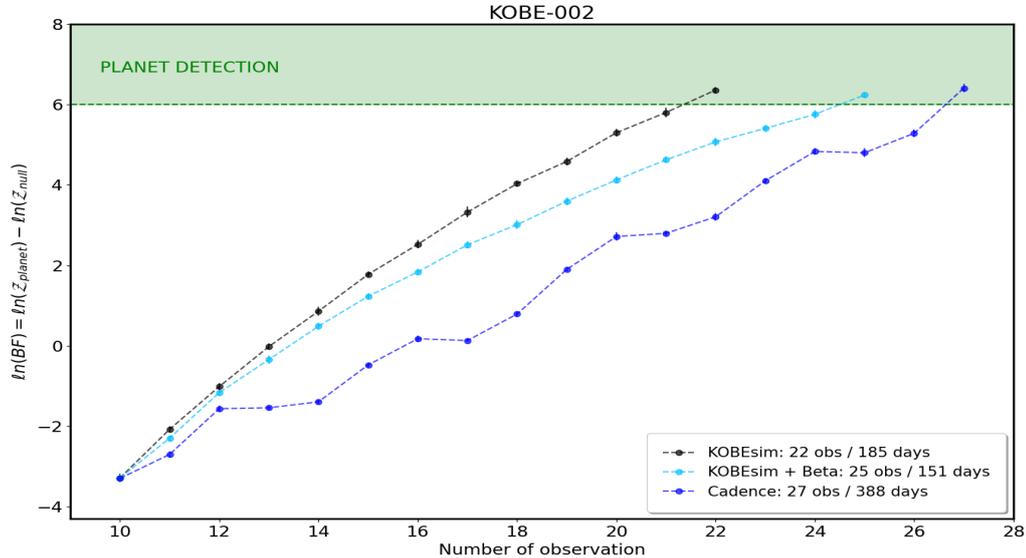


Figure 13: Prediction of the evolution of the logarithm of the Bayes Factor from KOBE-002 real data. The number of observations and the time it would take to detect the planet, are compared using three different strategies: KOBESim, KOBESim with the beta function, and spacing the observations a fixed cadence (6 days for KOBE-002).

Although in this particular case the RV semi-amplitude is big and it seems obvious to observe this target as much as possible to speed up the possible detection, the use of KOBESim could be a key tool to detect less massive planets that take long periods of time (those with M_p between 5-10 M_{\oplus}).

4 Conclusions and Future Work

Throughout this Master Thesis: (1) we have presented the KOBE experiment; (2) we have indicated the selection of the sample to be monitored for the next 2.5 years with the CARMENES instrument; (3) and we have explained in detail the development and application of the KOBESim algorithm, an observational strategy for blind RV surveys developed within the framework of this experiment. We summarize each of these points below:

The KOBE experiment has the aim of discovering around 14-40 new planets hosted by late K-dwarfs. These stars are of exceptional qualities to host habitable planets detectable by current high-resolution spectrographs, but have hardly been observed so far. KOBE is a Calor Alto legacy program with guaranteed observing time for 2.5 years, starting in January 2021. Blind-search surveys, are of great importance at the present time in this field. Since in the next decades the characterization of exoplanet atmospheres will be a priority.

The selection of the KOBE sample has been one of the objectives of this work, guaranteeing the fulfillment of the scientific goals and maximizing the chances of success. To do this, from a list of pre-selected 200 candidates of late K-type stars, we have discarded those that: have already been widely observed, have $V \sin i_{\star} = 0$ or $V \sin i_{\star} > 8 \text{ km s}^{-1}$, show contamination in their SED, and have companion source at an angular distance less than 1.5 arcsec.

As the second of the objectives, the Bayesian-based algorithm KOBESim, has been developed. Through the selection of the observing date that maximizes the increase in BF, we have proven with simulations the optimization in the detection of exoplanets in long-term RV surveys. Specifically, we have seen that this strategy reduces, by almost a factor 2, the number of observations required to achieve the detection of planets between 5 and 10 M_{\oplus} (Figure 10, 18 and 19), being the time periods shortened very significantly with the KOBESim weighted mode described above. An improvement is also made in planets of intermediate mass (around 20 M_{\oplus} , Figure 9), but it is no longer noticeable in planets that generate large RV semi-amplitude (Figure 20). Also, we have estimated the time that it would take us to detect the hypothetical planet that generates the candidate RV signal of KOBE-002. The conclusion is that if the trend continues, it could be detected in a dozen more observations (Figure 13).

We have seen that KOBESim is especially interesting for detecting low massive planets in long-term surveys. To be applicable to the KOBE experiment and thus achieve the detection of such low mass planets, we have to study if this is possible simply by increasing the exposure time, or if it would also be necessary to extend the observations beyond 2.5 years.

As future work, we have to implement the inference of the orbital parameters i , e , and ω to achieve a more realistic model, since until now they have not been considered due to computational time reasons. Also, we have to test if KOBESim is efficient in discarding a hypothetical planet in case of using a spurious period of the periodogram. And finally, in this work we consider 20 candidate phases to select the next observation with KOBESim. However, this will be selected automatically by the code based on the number of data already collected.

References

- [1] Akeson, R.L., Chen, X., Ciardi, D., “The NASA Exoplanet Archive: Data and Tools for Exoplanet Research”, The Astronomical Society of the Pacific, 125, 989, (2013). <https://exoplanetarchive.ipac.caltech.edu/>
- [2] Albrecht, S.H., Marcussen, M.L., Winn, J.N., Dawson, R.I. & Knudstrup, E., “A Preponderance of Perpendicular Planets”, arXiv:2105.09327, (2021).
- [3] Auclair-Desrotour, P. & Heng, K., “Atmospheric stability and collapse on tidally locked rocky planets”, Astronomy & Astrophysics, 638, A77, (2020).
- [4] Benz, W., Broeg, C., Fortier, A. *et al.*, “The CHEOPS mission”, Experimental Astronomy, 51, 109–151, (2021).
- [5] Borucki, W., Koch, D., Batalha, N. *et al.*, “KEPLER: Search for earth-size planets in the habitable zone”, Proceedings of The International Astronomical Union, 253, 289-299, (2009).
- [6] Cabona, L., Viana, P., Landoni, M. & Faria, J., “Scheduling strategies for the ESPRESSO follow-up of TESS targets”, arXiv:2005.14008 [astro-ph.EP], (2020).
- [7] Chulkov, D., Prokhorov, M., Malkov, O. *et al.*, “Detection of unresolved binaries with multicolor photometry”, Baltic Astronomy, 24, 137-143, (2015).
- [8] Cockell, C.S., Bush, T., Bryce, C. *et al.*, “Habitability: A Review”, Astrobiology, 16, 89-117, (2016).

- [9] Danielski, C., Baudino, P., Lagage, P.O. *et al.*, “Atmospheric Characterization of Directly Imaged Exoplanets with JWST/MIRI”, *The Astronomical Journal*, 156:276, 20pp, (2018).
- [10] Díaz, R.F., Ségransan, D., Udry, S. *et al.*, “The HARPS search for southern extra-solar planets. XXXVIII. Bayesian re-analysis of three systems. New super-Earths, unconfirmed signals, and magnetic cycles”, *Astronomy & Astrophysics*, 585, A134, (2016). <https://github.com/exord/bayev>
- [11] Dorn, C., Bower, D.J. & Rozel, A.B., “Assessing the Interior Structure of Terrestrial Exoplanets with Implications for Habitability”, *Handbook of Exoplanets*, Springer, (2018).
- [12] Dumusque, X., Udry, S., Lovis, C., “Planetary detection limits taking into account stellar noise: I. Observational strategies to reduce stellar oscillation and granulation effects”, *Astronomy & Astrophysics*, 525, A140, (2011).
- [13] Foreman-Mackey, D., Hogg, D.W., Lang, D. & Goodman, J., “emcee: The MCMC Hammer”, *The Astronomical Society of the Pacific*, 125, 925, (2013). <https://emcee.readthedocs.io/en/stable/>
- [14] Gaia Collaboration: Brown, A.G.A., Vallenari, A., Prusti, T. *et al.*, “Gaia Early Data Release 3: Summary of the contents and survey properties”, *Astronomy & Astrophysics*, 649, A1, (2021).
- [15] García-Piquer, A., Morales, J.C., Ribas, I. *et al.*, “Efficient scheduling of astronomical observations. Application to the CARMENES RV survey”, *Astronomy & Astrophysics*, 604, A87, (2017).
- [16] Gardner, J.P., Mather, J.C., Clampin, M. *et al.*, “The James Webb Space Telescope”, *Space Science Reviews*, 123, 485-606, (2006).
- [17] Gregory, P., “Bayesian Logical Data Analysis for the Physical Sciences”, Cambridge University Press, (2005).
- [18] Kane, S.R. & Gelino, D.M., “The Habitable Zone Gallery”, *The Astronomical Society of the Pacific*, 124:323–328, (2012). <http://www.hzgallery.org/>
- [19] Kasting, J.F., Whitmire, D. & Reynolds, R.T., “Habitable zones around main sequence stars”, *Icarus*, 101, 108-128, (1993).
- [20] Kopparapu, R. K., Ramirez, R., Kasting, J. *et al.*, “Habitable Zones around Main-Sequence Stars: new estimates”, *The Astrophysical Journal*, 765, 131, (2013).
- [21] Kunimoto, M. & Bryson, S., “Comparing Approximate Bayesian Computation with the Poisson-Likelihood Method for Exoplanet Occurrence Rates”, *Research Notes of the AAS*, 4, (2020).
- [22] Lillo-Box, J., Leleu, A., Parviainen, H. *et al.*, “The TROY project II. Multi-technique constraints on exo-trojans in nine planetary systems”, *Astronomy & Astrophysics*, 618, A42, (2018).
- [23] Loredo, T.J., “The promise of Bayesian Inference for Astrophysics”, Springer-Verlag, pp. 275–297, (1992).
- [24] Lovis, C. & Debra, A.F., “RV Techniques for Exoplanets”, *Exoplanets*, edited by S. Seager. Tucson, AZ: University of Arizona Press, 526, 27-53, (2011).
- [25] MacKay, D.J.C., “Information Theory, Inference, and Learning Algorithms”, Cambridge University Press, (2003).
- [26] Mayor, M., & Queloz, D., “A Jupiter-mass companion to a solar-type star”, *Nature*, 378, 355–359, (1995).
- [27] Mayor, M., Pepe, F., Queloz, D. *et al.*, “Setting New Standards with HARPS”, *The Messenger*, 114, 20-24, (2003).
- [28] Meadows, V. & Barnes, R.K., “Factors Affecting Exoplanet Habitability”, *Handbook of Exoplanets*, (2018).
- [29] Noack, L., Snellen, I.A.G., Rauer, H., “Water in Extrasolar Planets and Implications for Habitability”, *Space Science Reviews*, 212, 877-898, (2017).
- [30] Oshagh, M., Santos, N.C., Figueira, P. *et al.*, “Understanding stellar activity-induced RV jitter using simultaneous K2 photometry and HARPS RV measurements”, *Astronomy & Astrophysics*, 606, A107, (2017).
- [31] Pepe, F., Cristiani, S., Rebolo, R. *et al.*, “ESPRESSO at VLT. On-sky performance and first results”, *Astronomy & Astrophysics*, 645, A96, 26, (2020).
- [32] Perrakis, K., Ntzoufras, I. Tsonas, E.G., “On the use of marginal posteriors in marginal likelihood estimation via importance sampling”, *Computational Statistics Data Analysis*, 77, 54-69, (2014).
- [33] Quanz, S.P., Ottiger, M., Fontanet, E. *et al.*, “Large Interferometer For Exoplanets (LIFE): I. Improved exoplanet detection yield estimates for a large mid-infrared space-interferometer mission”, *arXiv:2101.07500*, (2021).
- [34] Quirrenbach, A., Amado, P.J., Caballero, J.A. *et al.*, “CARMENES: An Overview Six Months After First Light”, *Proceedings of the SPIE*, Volume 9908, id. 990812 14 pp., (2016).
- [35] Ramsey, F.P., “Truth and Probability”, *The Foundations of Mathematics and other Logical Essays*, Ch. VII, p.156-198, (1931).
- [36] Rauer, H. & Heras, A.M., “Space Missions for Exoplanet Science: PLATO”, *Handbook of Exoplanets*, (2018).
- [37] Ricker, G.R., Joshua, N., Vanderspek, R. *et al.*, “Transiting Exoplanet Survey Satellite (TESS)”, *Journal of Astronomical Telescopes, Instruments, and Systems*, Volume 1, id. 014003, (2015).
- [38] Schrijver, C.J. & Zwaan, C., “Solar and Stellar Magnetic Activity”, Cambridge University Press, (2000).
- [39] Sivia, D.S. & Skilling, J., “Data analysis: A Bayesian Tutorial”, Oxford, (2006).
- [40] Thrane, E. & Talbot, C., “An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models”, *Astronomical Society of Australia*, (2020).
- [41] Tkachenko, A., Lehmann, H. & Mkrtichian, D., “Spectroscopic modeling of the Algol-type Star TW Draconis”, *The Astronomical Journal*, 139:1327–1337, (2010).
- [42] Tinetti, G., Eccleston, P., Haswell, C., “Ariel: Enabling planetary science across light-years”, (2020). <https://arxiv.org/abs/2104.04824>
- [43] Tjoa, J.N.K.Y., Mueller, M. & van der Tak, F.F.S., “The subsurface habitability of small, icy exomoons”, *Astronomy & Astrophysics*, 636, A50, (2020).
- [44] VanderPlas, J.T., “Understanding the Lomb–Scargle Periodogram”, *The Astrophysical Journal Supplement Series*, 236:16, 28, (2018).
- [45] Wolszczan, A., & Frail, D.A., “A planetary system around the millisecond pulsar PSR1257 + 12”, *Nature*, 355, 145-147, (1992).

A Radial Velocity

The velocity of a star projected in the line of sight while revolving around the planet-star center of mass, is given by:

$$v_r(t) = V_{sys} + K [\cos(\nu(t) + \omega) + e \cos \omega] \quad (\text{A.1})$$

where V_{sys} is the systemic RV of the whole system, K is the semi-amplitude of the signal, $\nu(t)$ is the true anomaly (angle between the periastron and the position of the planet in the elliptical orbit measured from the central star), ω is the argument of the periastron (angular distance between the line of nodes and the periastron) and e is the eccentricity (degree of deviation of the elliptical orbit from a circle). See Figure 14 to visualize the definition of these orbital parameters.

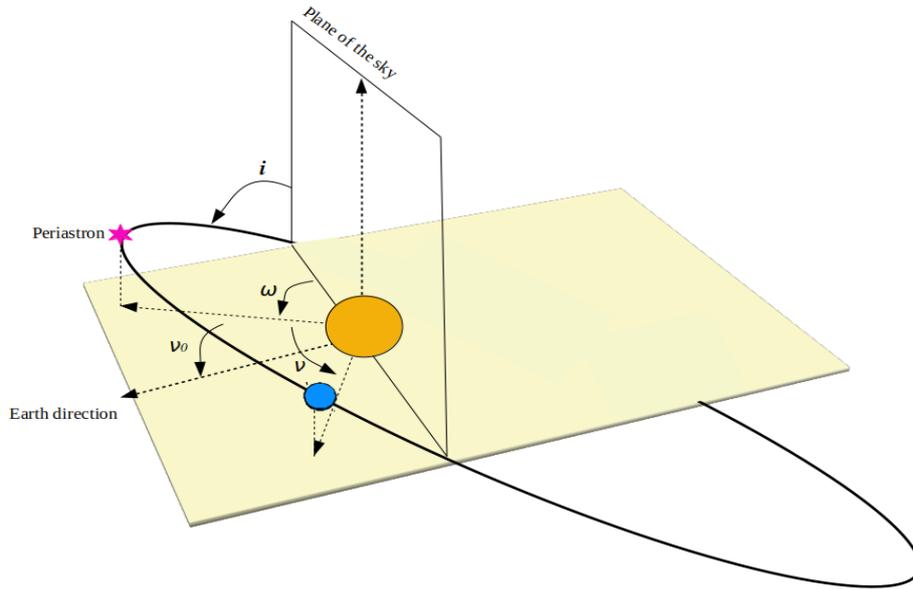


Figure 14: Diagram of the orbital elements. The black solid line is the elliptical orbit of the planet (blue sphere) around the central star (orange sphere), and its periastron is symbolically represented with a magenta star. In pale yellow is shown the plane that contains the line of sight (Earth direction) and that is perpendicular to the plane of the sky. The parameters shown are: orbital inclination (i), argument of the periastron (ω), true anomaly of the conjunction location (ν_0), and true anomaly at a specific time (ν).

Taking as a point of reference the time in which the conjunction is observed, t_0 , we can define the true anomaly at conjunction, ν_0 . To know the position of the planet in the orbit at any time referred to the conjunction, the geometric phase is defined:

$$\phi(t) = \nu(t) + \omega - \frac{\pi}{2} = 2\pi \frac{t - t_0}{P} \quad (\text{A.2})$$

Using this definition in the RV equation A.1, we have:

$$v_r(t) = V_{sys} + K \left[\cos \left(2\pi \frac{t - t_0}{P} + \frac{\pi}{2} \right) + e \cos \omega \right] \quad (\text{A.3})$$

Where the semi-amplitude, by means of the third Kepler law and assuming the mass of the planet much smaller than that of the star, is described by the expression:

$$K = \frac{28.4 \text{ m/s } M_p \sin i}{\sqrt{1 - e^2} M_J} \left(\frac{M_\star}{M_\odot} \right)^{-2/3} \left(\frac{P}{\text{yr}} \right)^{-1/3} \quad (\text{A.4})$$

Where M_p is the mass of the planet, i is the orbital inclination, M_J is the mass of Jupiter, M_\star is the mass of the central star, P is the orbital period of the planet, and M_\odot is the mass of the Sun.

B Additional figures for KOBESim

B.1 Simulated data

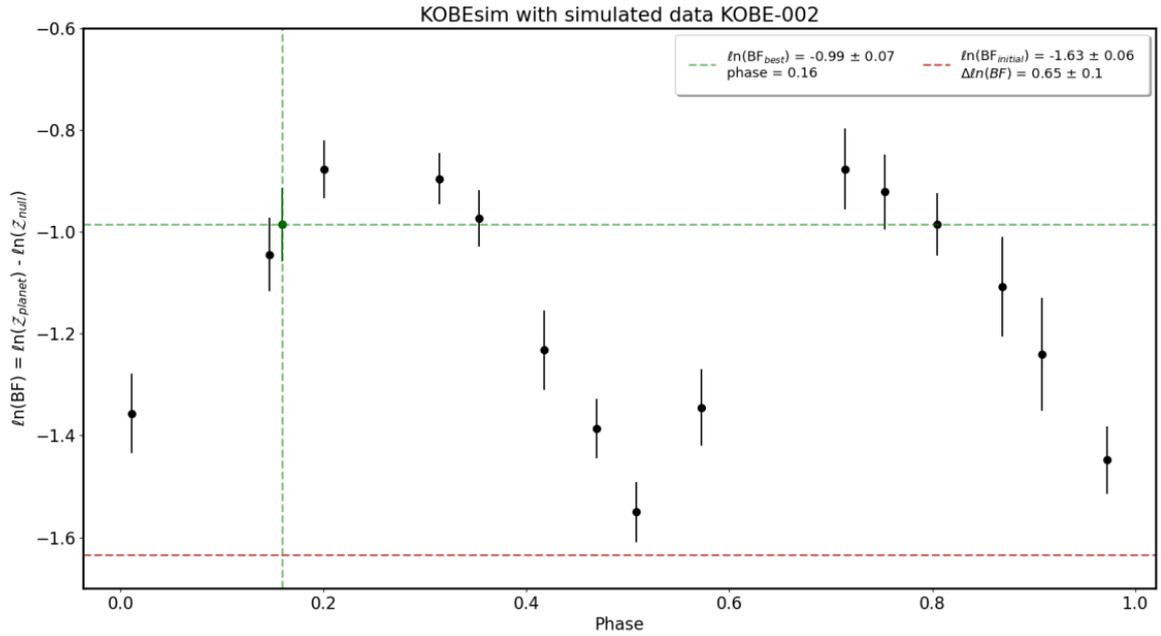


Figure 15: Using beta function mode. Output graph of KOBESim for the KOBE-002 simulated data. The y-axis shows the logarithm of the expected Bayes Factor and x-axis the associated orbital phase. The legend shows the phase selected as the best option, as well as the corresponding expected value of $\ln(BF)$ (green lines), its increment, and the value before adding that future observation (red line).

Calendar_day	JD	phase	IBF	sigma_IBF	delta_IBF	sigma_delta_IBF
2021-05-05	2459340.365	0.159	-0.986	0.072	0.648	0.096
2021-05-04	2459339.363	0.146	-1.045	0.072	0.589	0.095
2021-05-08	2459343.533	0.200	-0.878	0.057	0.757	0.085
2021-05-17	2459352.375	0.314	-0.896	0.050	0.738	0.080
2021-05-20	2459355.377	0.353	-0.974	0.055	0.661	0.083
2021-05-25	2459360.381	0.417	-1.232	0.078	0.403	0.100
2021-05-29	2459364.385	0.469	-1.386	0.058	0.248	0.085
2021-06-06	2459372.390	0.572	-1.345	0.075	0.290	0.098
2021-06-17	2459383.394	0.714	-0.877	0.079	0.758	0.100
2021-06-01	2459367.388	0.508	-1.550	0.059	0.084	0.086
2021-06-20	2459386.396	0.753	-0.922	0.073	0.712	0.096
2021-06-24	2459390.396	0.805	-0.985	0.061	0.650	0.087
2021-06-29	2459395.396	0.869	-1.108	0.098	0.527	0.116
2021-07-02	2459398.394	0.908	-1.241	0.111	0.393	0.128
2021-07-07	2459403.394	0.972	-1.448	0.066	0.186	0.091
2021-07-10	2459406.392	0.011	-1.357	0.078	0.278	0.100

Table 4: Using beta function mode. Output csv file of KOBESim for the KOBE-002 simulated data. Each row corresponds to a candidate observing date, and they are ranked in order of preference, being the first the best option according to KOBESim. From left to right the columns are: date in format year-month-day, the Julian day, the orbital phase, the expected logarithm of the Bayes Factor, its uncertainty, the increase of the logarithm of the Bayes Factor, and its uncertainty.

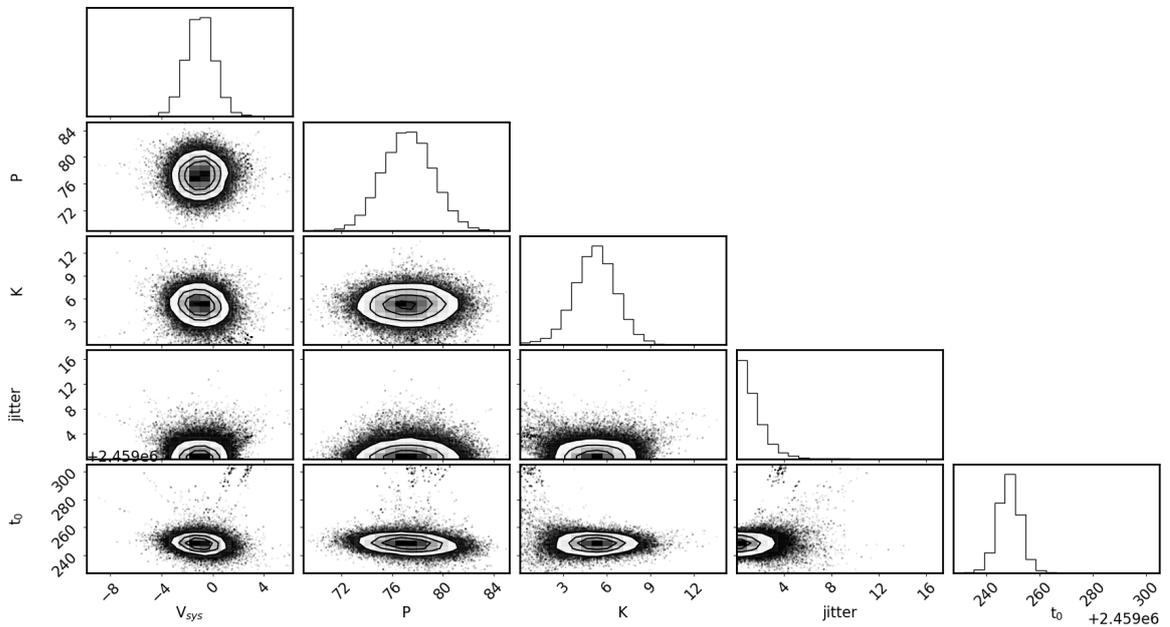


Figure 16: Corner diagram of the parameters for the planet model resulting from applying *emcee*. Case for simulated RV data for KOBE-002, planet of $20M_{\oplus}$ with $P = 77.5$ days.

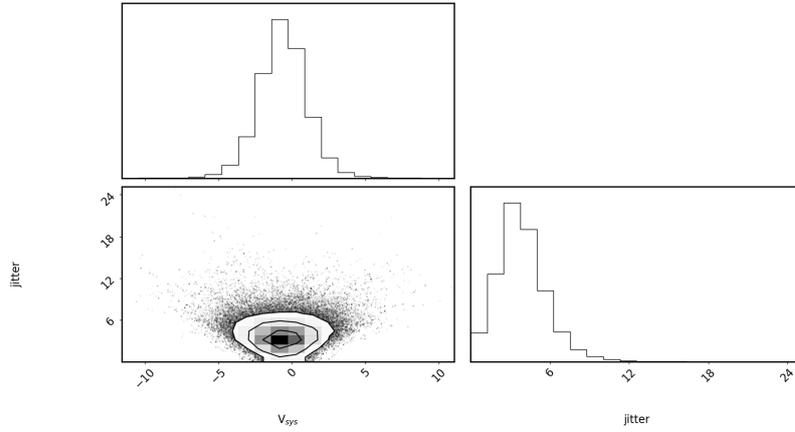


Figure 17: Corner diagram of the parameters for the null hypothesis resulting from applying *emcee*. Case for simulated RV data for KOBE-002.

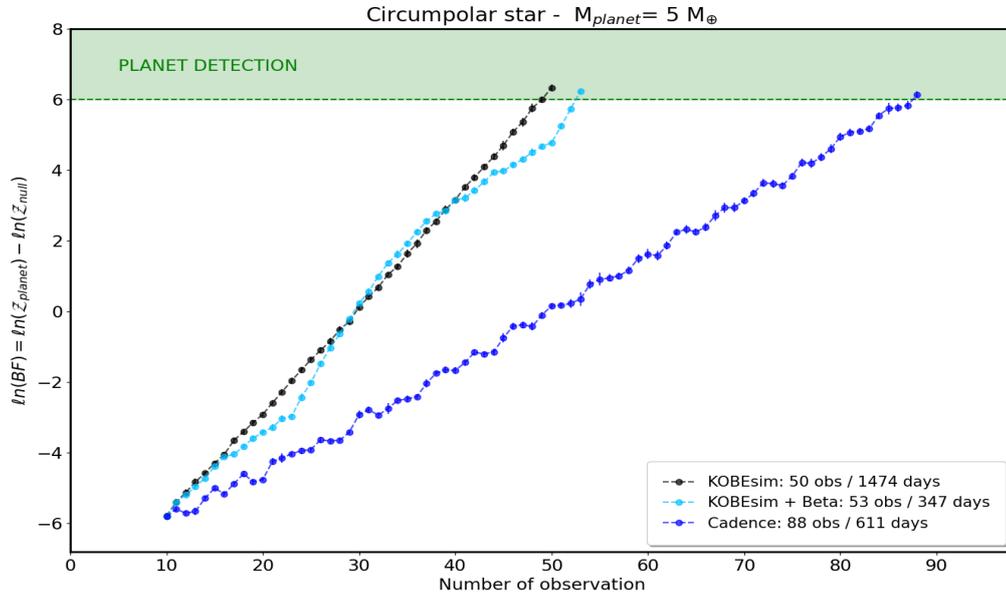


Figure 18: Prediction in the evolution of the logarithm of the Bayes Factor from the simulated data of a circumpolar star, for a $5 M_{\oplus}$ planet with $P = 52$ days and **reducing the uncertainties to a mean value of 1.5 m s^{-1}** . The number of observations and the time it would take to detect the planet, are compared using three different strategies: KOBESim, KOBESim with the beta function, and spacing the observations a fixed cadence (6 days).

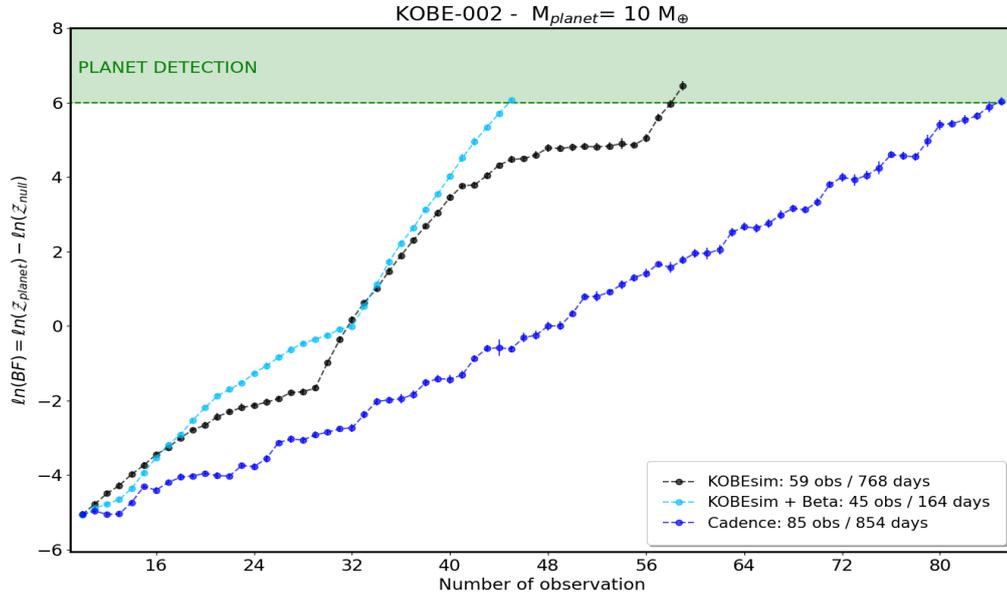


Figure 19: Prediction in the evolution of the logarithm of the Bayes Factor from the simulated data of KOB-002, for a $10 M_{\oplus}$ planet with $P = 52$ days. The number of observations and the time it would take to detect the planet, are compared using three different strategies: KOBEsim, KOBEsim with the beta function, and spacing the observations a fixed cadence (6 days for KOB-002)

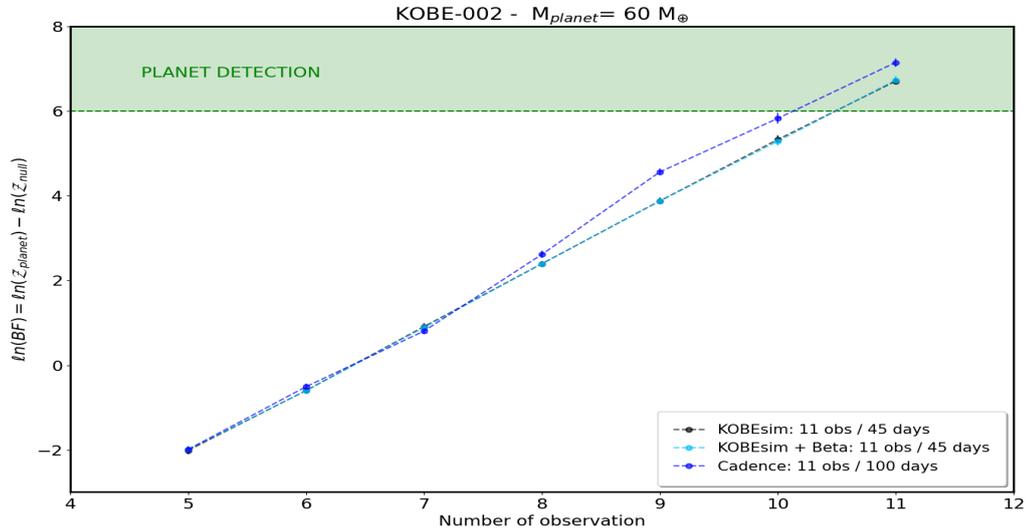


Figure 20: Prediction in the evolution of the logarithm of the Bayes Factor from the simulated data of KOB-002, for a $60 M_{\oplus}$ planet with $P = 52$ days. The number of observations and the time it would take to detect the planet, are compared using three different strategies: KOBEsim, KOBEsim with the beta function, and spacing the observations a fixed cadence (6 days for KOB-002).

B.2 Real KOBE-002 data

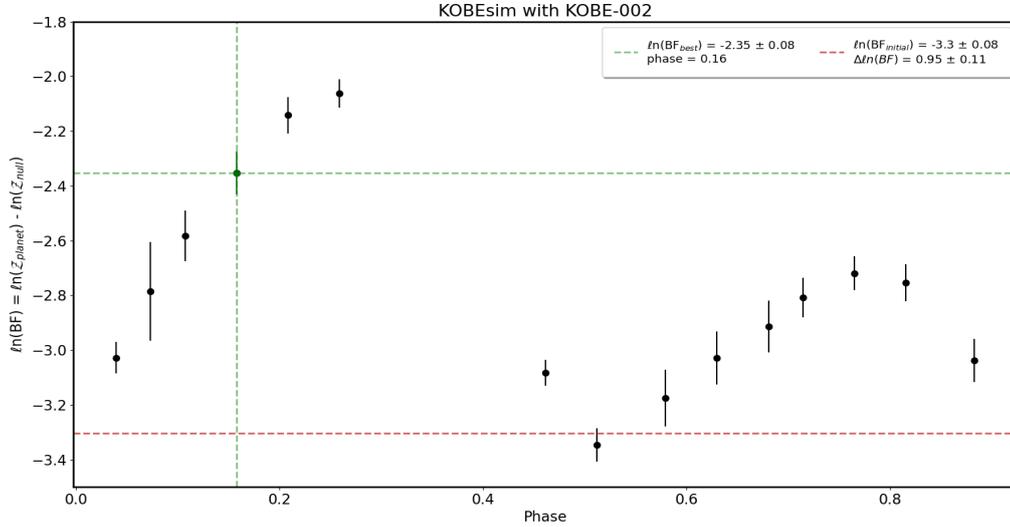


Figure 21: Using beta function mode. Output graph of KOBESim for real RV data of KOBE-002. The y-axis shows the logarithm of the expected Bayes Factor and x-axis the corresponding orbital phase. The legend shows the phase selected as the best option, as well as the corresponding expected value of $\ln(BF)$ (green lines), its increment, and the value before adding that future observation (red line).

Calendar_day	JD	phase	lBF	sigma_lBF	delta_lBF	sigma_delta_lBF
2021-05-30	2459365.385	0.158	-2.353	0.079	0.950	0.109
2021-05-27	2459362.383	0.107	-2.583	0.092	0.720	0.119
2021-06-02	2459368.388	0.208	-2.142	0.066	1.161	0.100
2021-05-25	2459360.381	0.073	-2.786	0.180	0.518	0.195
2021-06-05	2459371.390	0.259	-2.062	0.052	1.242	0.091
2021-05-23	2459358.379	0.040	-3.028	0.058	0.276	0.095
2021-06-17	2459383.394	0.461	-3.083	0.047	0.220	0.089
2021-06-27	2459393.396	0.630	-3.028	0.097	0.276	0.123
2021-06-24	2459390.396	0.579	-3.175	0.104	0.128	0.129
2021-06-30	2459396.396	0.681	-2.914	0.094	0.389	0.120
2021-07-02	2459398.394	0.714	-2.807	0.072	0.497	0.104
2021-07-05	2459401.394	0.765	-2.719	0.063	0.585	0.098
2021-07-08	2459404.392	0.815	-2.754	0.067	0.550	0.101
2021-07-12	2459408.390	0.883	-3.038	0.079	0.265	0.109
2021-06-20	2459386.396	0.512	-3.36	0.061	-0.042	0.096

Table 5: Using beta function mode. Output csv file of KOBESim for real RV data of KOBE-002. Each row corresponds to a candidate observing date, and they are ranked in order of preference, being the first the best option according to KOBESim. From left to right the columns are: date in format year-month-day, the Julian day, the orbital phase, the expected logarithm of the Bayes Factor, its uncertainty, the increase of the logarithm of the Bayes Factor, and its uncertainty.

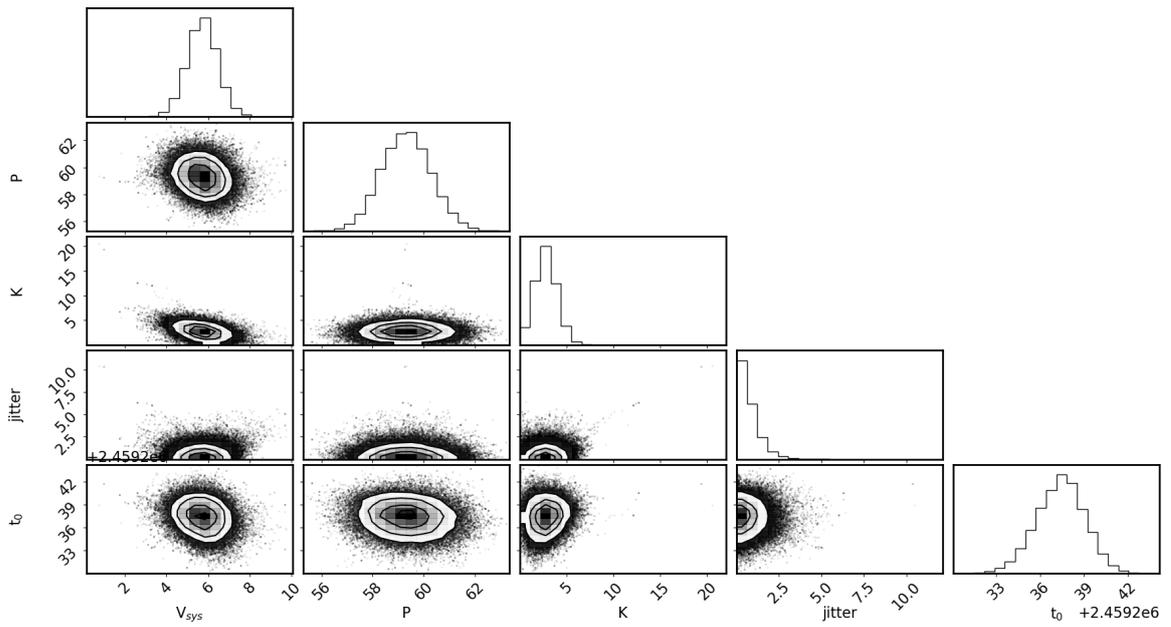


Figure 22: Corner diagram of the parameters for the planet model resulting from applying *emcee*. Case for real RV data of KOBE-002, hypothetical planet with $P = 59.4$ days

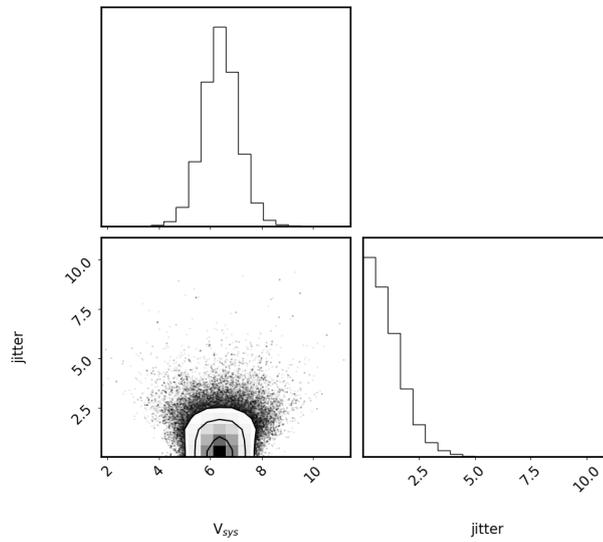


Figure 23: Corner diagram of the parameters for the null hypothesis resulting from applying *emcee*. Case for real RV data of KOBE-002.