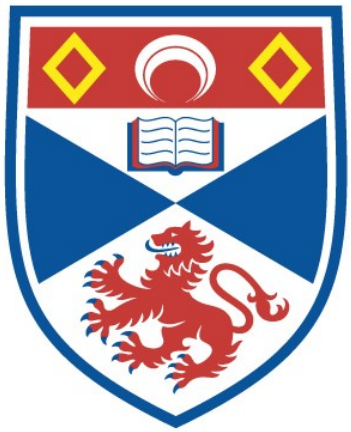
Yin or yang:
confronting stellar activity in radial-velocity searches for exo-Earths

Ancy Anna John

A thesis submitted for the degree of PhD
at the
University of St Andrews



2024

Declaration

Candidate's declaration

I, Ancy Anna John, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 30,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2020.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 21 June 2024 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD at the University of St Andrews and that the candidate is qualified to submit this thesis in the application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 21 June 2024 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Ancy Anna John, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by the candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 21 June 2024 Signature of candidate

Date 21 June 2024 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Ancy Anna John, understand that by declaring that I have original research data or digital outputs, I should make every effort to meet the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 21 June 2024 Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by the candidate and supervisor regarding the publication of underpinning research data or digital outputs: No embargo on underpinning research data or digital outputs.

Date 21 June 2024 Signature of candidate

Date 21 June 2024 Signature of supervisor

Abstract

"*Somewhere something incredible is waiting to be known.*" - Carl Sagan

Stellar activity has a radial velocity (RV) signature which complicates the detection of planetary signals using RV method. Because magnetic activity of the host star occurs on identical timescales to small planets, it can entirely hide or mimic their RV signals. This is stopping us from harnessing the full potential of the state-of-the-art instruments to definitively enter the realm of terrestrial planet discovery. I formulated a workflow called TWEAKS to obtain reliable detections of planets whose RV signals were previously below the detection threshold. This workflow integrates stellar activity mitigation techniques in wavelength and time domains, utilizing SCALPELS for spectral line-shape decorrelation and KIMA Gaussian Process to counteract any shift-like activity signal leakage.

I used this statistically rigorous approach to search for planets around stars with different levels of intrinsic variability. I chose CoRoT-7 to represent the high-activity star case. I applied TWEAKS to the archival HARPS data and refined the mass of the transiting Super-Earth CoRoT-7b to a precision of 10%, while confidently confirming the existence of an additional non-transiting Neptune, CoRoT-7d. My work definitively demonstrated how stellar variability and limited knowledge about additional planets in the system, complicate the detection, mass determination, and characterisation of exoplanetary systems near the detection threshold.

To represent the moderate and low activity cases, I extended the search to two HARPS-N RPS stars, HD144579 and HD166620. Despite finding no planets in the 10 years of data of either system, TWEAKS enabled me to achieve a sub-ms^{-1} detection limit (0.54 ms^{-1}), close to the long-term instrumental calibration precision of the HARPS-N spectrograph (0.50 ms^{-1}). TWEAKS have also facilitated independent robust detections and precise mass and bulk-density determinations for various Super-Earths and mini-Neptunes across multiple *TESS*, *K2*, and RPS planetary systems, such as TOI1778, TOI1730, TOI5398, HD99492, and HD48948. In 78% of these cases, TWEAKS provided mass measurements that are either more precise or comparable to those obtained using conventional time-domain alone approaches.

Acknowledgements

To my Pappa and Amma

To my supervisor, Andrew, my rock throughout this voyage, thank you for your guidance and unwavering support. I appreciate that you have raised the bar so high for supervisors that no one can even attempt to touch it. Thank you for being available always! Thank you for sending me to all those amazing conferences and meetings. Thank you for valuing all my efforts, instincts and decisions. Thank you for encouraging me to keep my head up and making me believe in my worth. I hope to carry a small portion of your contagious enthusiasm for life and research with me. I will always treasure the walk you and Moira took me on Teide, where it felt like we were on Mars, and the sunset we saw in Tenerife celebrating my first observational run, with the incredible Caiprinha!

To my beloved, Rakesh, thank you for being the wind beneath my wings. You left everything behind to support me to achieve my dreams. I am forever indebted for your love and respect towards who I am and what I love. Thank you for keeping me sane through my troughs. Thank you for making all those conference trips unimaginably memorable. I cannot imagine how you kept waiting for me in the icy cold, whether it was outside the office, at a conference, at a public event, at a dinner, or at the British Parliament. You're my home.

To my family, my forever constants, thank you for always holding me in your hearts and hands. Pappa, this is my gift to you, for encouraging me to pursue my dreams despite the difficulties, and criticism you endured. The cost of your sweat is the shade I have today. Amma, this is your dream! You've taught me the meaning of compassion. Thank you for remembering all my deadlines (small and big). Thank you for staying up (despite being in different time zones), until I reach home after office, talks, travel or whatever. Pappa & Amma, without you both, this would mean nothing to me. You have always been a feathered blanket, that kept me warm. Words are not enough to acknowledge the sacrifices you made to provide me with the best even in our nothingness. Thank you Achan & Amma for supporting my dreams. Thank you for standing up for me. I appreciate you, Amma, for lending me your gold bangle so I could pay for my qualifying examinations. Thank you for being the Mom you are. Kichu, I

remember you lying on the floor by the door, waiting for me to finish my work. Thank you Kichu Mon for making my PhD preparation (and rejection) days alive.

My brother Robin, sister Shyamily, and Nephew Elias, thank you for all the efforts you made to bring home close to me. My PhD journey would not have been this easy without you. Thank you for making me feel special. Thank you for taking pride in all my small achievements. Thank you for gifting me the amazing telescope. I am grateful to Elias Mon for lending me your fish catcher so that I could use it in my public talks to show how to detect planets using the radial velocity method. You have never failed to light up my face. Neha & Richie, I am so grateful that you sewed my heart together with yours. You are the greatest gift I received from Scotland. Thank you for all the guidance, love, and care you have been pouring upon me. From the warm jacket you gave on my induction day, to the hot dosas you lovingly served, I am grateful for everything. Thank you for being a great part of my journey. You are the ones who taught me how friends can become family. I cannot comprehend how you take wild risks to cheer me on. Why did you fly to London, just to celebrate me winning STEM for Britain?

Rev. Fr. J. Varghese (my achan appachan), I know you are watching me from above. I am grateful that you have inspired me since I could write the word 'dream'. Thank you for connecting me with inspirational human beings, from different arenas, and helping me to find my ikigai. Thank you, Rose (Dr. Rose Waugh), for being there to help me with my stray questions and concerns regarding my PhD and life in general, even before I started my life in St Andrews. Thank you Dr. Yogesh C. Joshi) and ARIES (Aryabhatta Research Institute for observational sciencES, India) for providing me with an excellent project opportunity and supporting me through my PhD application process. Thank you, Priyanka Srivastava and Jayanand Maurya for sitting with me through the late-night de-bugging sessions at ARIES. Thank you Shincy Ma'am (Dr. Shincy John) and Mansi (soon to be Dr. Mansi Padave!) for your invaluable motivation and support. Thank you Shweta (Dr. Shweta Dalal) and Raphie (Dr. Raphaelle Haywood), for being the friendly face within the collaborations and guiding me through the Post-doc applications. Thank you Annelies (Dr. Annelies Mortier), for all your support. Thank you, Tom (Dr. Thomas G. Wilson), for the kindness you showed both personally and professionally. I cannot imagine how I would have survived the initial days of my PhD without your support. I am grateful to all the teachers who have illuminated my path. I want to thank my motherland, India, for moulding me into the person I am today. You have my eternal gratitude for supporting my education, from the eighth grade to Masters. Last, but not least, I dedicate this thesis to Kalpana Chawla and Jocelyn Bell Burnell for kindling the fire within me.

Above all, my dear Lord, thank you for keeping me under your wings during all the storms. I'm grateful for making me feel like your favourite.

Thank you my small, yet strong circle!

Funding

This work was supported by the World leading St. Andrews Doctoral Scholarship.

Research Data/Digital Outputs access statement

Research data underpinning this thesis are available at [DOI]

OR

Digital outputs underpinning this thesis are available at [DOI]

Contents

List of Figures

List of Tables

1

Introduction: Hunt for pale blue dots

"To me, it underscores our responsibility to deal more kindly with one another, and to preserve and cherish the PALE BLUE DOT- the only home we have ever known." - Carl Sagan

‘ Exoplanets are planets orbiting stars other than our Sun.’

1.1 Are we alone?

Owing to the seemingly boundless myriad of stars surrounding us, there has always been an innate human urge to look at the night sky and ask,

"Are we alone?".

Renowned science fiction author Arthur C. Clarke once said, "Two possibilities exist: either we are alone in the Universe or we are not. Both are equally terrifying". However, in my opinion, both are equally enticing!

The path-breaking discovery of the first exoplanets orbiting a pulsar in 1992 confirmed that exoplanets are no longer fictional! As of the cut-off date of this thesis, 15-03-2024, the existence of 5599 exoplanets has been confirmed, with more than 1300 multi-planetary systems. And the number is growing steadily and rapidly.

1.1.1 How do we find exoplanets?

Our understanding of exoplanets stems from diverse detection methods, each offering unique insights into their nature. Among these techniques, radial velocity measurements, transit observations, direct imaging and microlensing have become cornerstones in the identification and characterization of exoplanetary systems. While acknowledging the wider range of methods that contribute to our constantly developing understanding of exoplanetary science, this section dives into the specifics of direct imaging, transit observations, and radial velocity measurements.

Direct imaging - taking photographs

In contrast to other techniques that use indirect observations to infer the existence of exoplanets, direct imaging records reflected light from exoplanets. However, this approach presents a unique set of challenges due to the extreme brightness contrast between the host star and

its orbiting planets. Exoplanets similar to planets in our solar system will be billions of times fainter than their host stars, when illuminated by reflected Sunlight. This makes direct detection very difficult, especially given the planet-star separation of at most one or two arcsecs (eg: Marois et al. (2008); Carrión-González et al. (2020)). The case is further more challenging in the optical wavelengths due to the substantial planet/star intensity ratio. This is because, in the thermal infrared, we can observe the planet's intrinsic radiation at a wavelength where the star's contribution is less pronounced than in the optical, where we can only observe scattering. As a result, the planet/star flux ratio is affected by multiple orders of magnitude in optical wavelengths.

Therefore, direct imaging has only made up 1.3% of the entire discovery budget. While individual pixels of light extracted directly from exoplanets may not suffice to unveil their surface characteristics, directly imaged planets offer a compelling alternative: detailed profiles of exoplanetary atmospheres and potential indicators of life-sustaining gases. However, the forthcoming wave of space telescopes powered by rapidly evolving technologies is poised to search for temperate exoplanets directly in reflected starlight (eg: Large Interferometer For Exoplanets (LIFE) in the mid UV, and Habitable Worlds Observatory (HWO) in IR/Optical/UV wavelengths; (Quanz et al., 2023)).

The bulk of exoplanet discoveries to date have come from two indirect detection methods—the transit method and the radial velocity approach. These state-of-the-art methods are rooted in indirectly inferring any orbiting planet's presence, by observing the parent star for eclipses and Doppler shifts.

Transit method - looking for eclipses

The transit method reveals an exoplanet not because it is directly visible to us from tens to hundreds of light-years away, but because it eclipses the parent star and a flux drop can be observed over time. The flux-drop level depends on the transiting planet's size (Figure 1.1). This method is hence used in determining the exoplanet's radius.

If we assume the stellar disc is uniformly bright and ignore any flux from the planet, the fractional loss of light can be expressed as

$$\frac{\Delta F}{F} = \frac{R_p^2}{R_*^2} \tag{1.1}$$

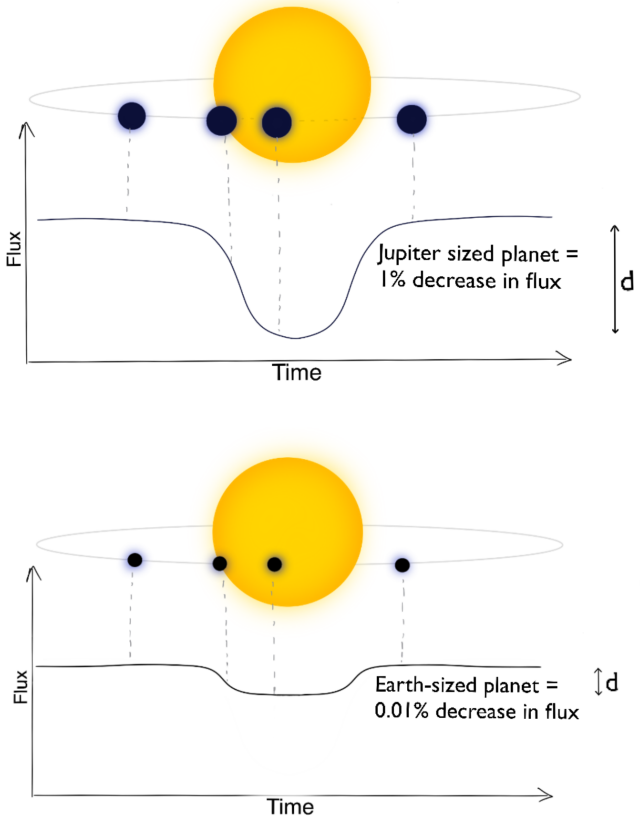


Figure 1.1: Exoplanet detection via transit method: The planet passes in front of the star disc, resulting in a reduction (or dip) in the star's overall brightness (or flux) by casting its shadow upon it.

where R_p and R_* are the radii of the transiting planet and parent star respectively. The depth, epoch, duration, and interval between transits are among the parameters of the transit 'dip' that can be translated into various observables. These include the impact parameter (the proximity of planet transit to the centre of the star), the size of the star, the size of the planet, orbital period (length of planet's year) and eccentricity of the planet's orbit.

Several constraints restrict the transit method's ability to identify exoplanets of all sorts. Only a small percentage of exoplanets eclipse their stars, as the orientation must be edge-on or 'just right' to observe a transit. For this reason, the transit method is biased towards detecting planets with short orbital periods orbiting close to their host star. The likelihood that a planet's orbit will align with the vantage point of an observer is determined by the ratio of the star's radius to the planet's semi-major axis;

$$\text{P(transit)}=\frac{R_*}{a} \tag{1.2}$$

Such an alignment occurs in only about a small fraction of planets with short orbital periods, and the percentage drops even more for planets with longer orbital periods.

Additionally, it has a significant false positive rate (up to 40% in single-planet systems; Santerne et al. 2012). This calls for the need to make additional observations, which frequently involve using a different technique. Therefore, after several vetting and validation procedures for astrophysical false positives, the most intriguing transiting planet candidates are subsequently followed up with the radial velocity method to confirm the detection, supplement the information obtained from transits, get additional observables and search for additional non-transiting planets in the system.

Radial Velocity method - looking for the star dance

Exoplanet detection using the radial-velocity (RV) approach (also known as the "Doppler wobble" method) is an indirect method based on the observation that a star with an orbiting planet is not stationary. The star responds to the planet's gravitational pull by moving slightly around a common centre of mass, which has an impact on the star's observed spectrum. The star's spectrum appears to be shifted towards shorter or bluer wavelengths as it moves towards the observer. In contrast, the spectrum will shift towards longer or redder wavelengths, when the star moves away. Because of the star's large mass when compared to the planet's, this reflex motion is rather tiny, usually on a scale of a few to hundreds of metres per second. High-resolution spectrographs are used to monitor these Doppler shifts to reveal information on the planet's mass, orbital period and eccentricity of the orbit.

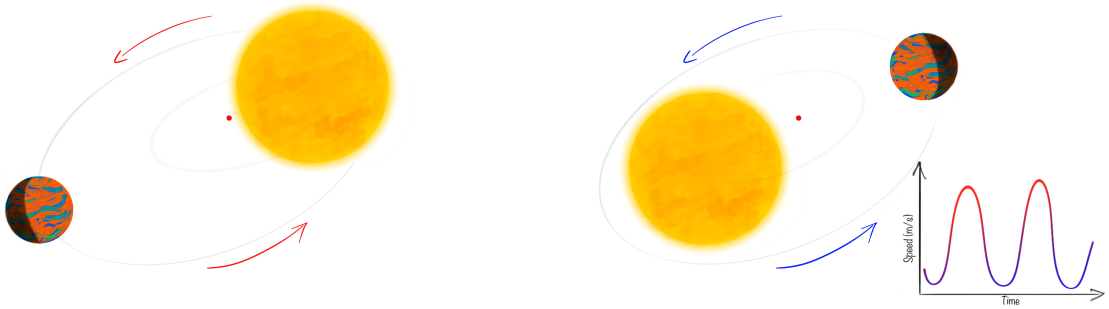


Figure 1.2: Exoplanet detection using radial-velocity method. Stars and orbiting planets exert a gravitational pull on one another. This makes the star wobble, which results in the shift of the stellar light or spectrum towards bluer and redder wavelengths.

The RV method serves as an independent follow-up method to confirm the planets detected

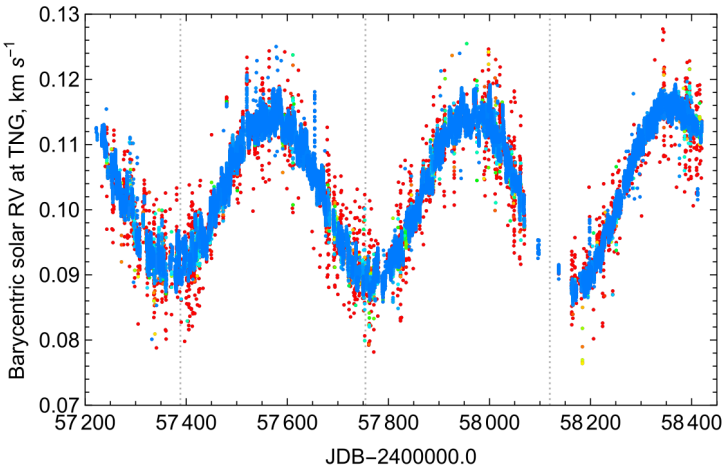


Figure 1.3: The Sun's RV measurements obtained using HARPS-N spectrograph mounted on Telescopio Nazionale Galileo (TNG), as shown in Collier Cameron et al. (2019). JDB stands for the Barycentric Julian Date. The observations are colour-coded based on the data quality (blue-good, red-bad). The sinusoid with a semi-amplitude of 12 m s^{-1} and an apparent period close to 398 days, represents the reflex motion of the Sun in response to the gravitational pull of Jupiter. The 398-day (1.092 years) period is the synodic period of Jupiter observed from Earth.

via the transit method and to reveal the presence of non-transiting planets. The discovery of the first exoplanet orbiting a Sun-like star (51 Pegasi; Mayor & Queloz 1995) was made using this technique. This method has ever since contributed to the discovery of 19.3% of the exoplanets known to date.

1.1.2 How do we weigh exoplanets?

A planet's mass is arguably its most fundamental physical characteristic, which together with its radius determines its bulk density, which in turn provides insight into its interior composition and structure. The RV technique, which uses high-resolution spectra, is currently the most viable tool for determining the masses of HZ Earths orbiting nearby F, G, and K spectral-type stars (Struve, 1952).

The term radial velocity refers to the velocity component of the star in the radial direction (along the observer's line of sight). If λ_{obs} denotes the wavelength of the star's light that underwent a relativistic Doppler shift in comparison to the wavelength if the star was stationary λ_{stat}, then each line in the stellar spectrum will appear to be shifted by an amount equal to:

$$\Delta\lambda = \lambda_{obs} - \lambda_{stat} \tag{1.3}$$

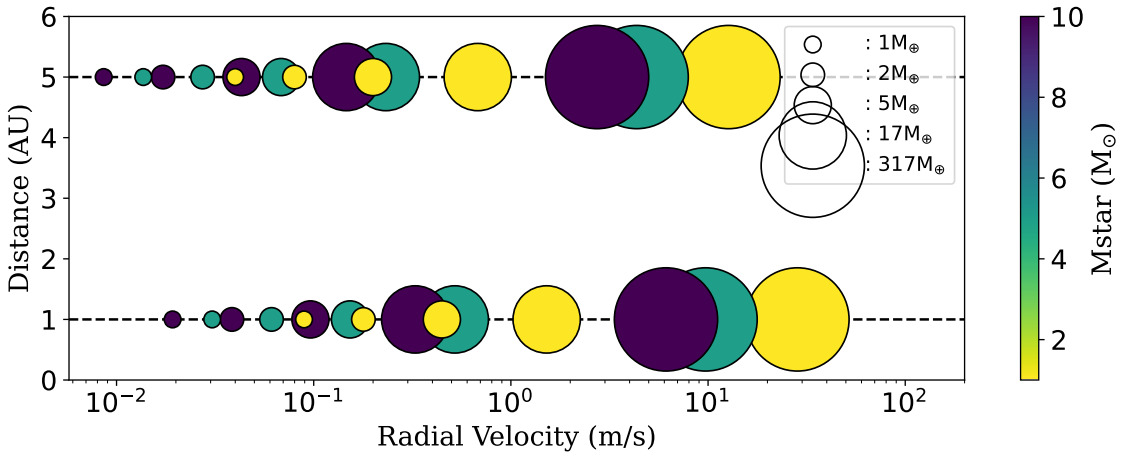


Figure 1.4: The radial velocity semi-amplitudes of some familiar planets, based on their masses (in $M_\oplus$), and distance (in AU) from the parent stars with different masses (in $M_\odot$). To illustrate the resulting RV semi-amplitude in each scenario, all planets are positioned at 1 and 5 AUs around stars with masses of 1, 3, and 10 $M_\odot$. The size of the circles approximately represents the mass of the planets. This figure is inspired by Table 1 from the chapter 'Exoplanets: The Road to Earth Twins' from Lovis & Fischer (2010).

We can then express the radial velocity of the star in terms of this wavelength shift:

$$v_{\text{radial}} = c\frac{\Delta\lambda}{\lambda_{\text{stat}}} \tag{1.4}$$

where c is the velocity of light. By tracking a star's RV over time, we can identify any changes (or shifts) brought about by a planet's orbit. The amplitude of the RV variation is proportional to the mass of the orbiting planet and also depends on the distance from the host star. Figure 1.4 shows a few examples of real and speculative planets orbiting stars with different masses together with their expected RV semi-amplitudes.

If the K is the RV semi-amplitude of the star with mass M_*, due to the gravitational pull of a planet with mass M_{p} orbiting in a circular orbit in every P days, then the mass function can be expressed as:

$$M' = \frac{M_{\text{p}}^3 sin\ i}{(M_* + M_{\text{p}})^2} = \frac{K^3 P}{2\pi G}(1-e^2)^{\frac{3}{2}} \tag{1.5}$$

Since spectra record the radial component of the 3D motion, the mass will be calculated as $M_{\text{p}}\sin i$, where i is the orbit's inclination with respect to the line-of-sight (i=90°, i.e, sin i=1 for an edge-on orbit), which can be determined via transit measurements, and G represents the gravitational constant (See section 2.1.1 of Perryman (2018) for a detailed derivation).

As we have no access to the i measurement for planets that are not transiting, we can only obtain $M_p \sin i$, which gives the minimum mass for these planets. On the other hand, for transiting planets, we can combine the transit and RV observations to obtain a comprehensive set of orbital and planetary parameters.

We now know that a star wobbles around the centre of mass of the planetary system and the velocity at which it travels along the line of sight of the observer is called the radial velocity. In Section 1.1.2 we saw how we derive the radial velocity from the Doppler shift of stellar spectral lines and how the information translates into the planet's mass. The next question is how precisely we can measure them.

Measuring RVs using CCFs

Acquiring a spectrum with the maximum number of lines is key to measuring the precision of an orbiting planet's RV signal from a star, with high precision. If we have a wavelength-calibrated spectrum with thousands of lines, we can fit simple Gaussian profiles to each of the lines and measure the deviation of each profile's centre from the rest-frame wavelengths. These values can then be used to determine the corresponding RVs for each line using equation 1.4. We can then measure the star's RV by taking a weighted average of these measurements. This is the basic idea. However, the problem is that measuring RVs for individual lines is computationally expensive and also the uncertainty on each measurement can propagate. Nonetheless, some recent investigations demonstrated how to deal with this issue by measuring line-by-line (LBL) RVs (Bouchy et al., 2001) by carefully choosing spectral lines (Dumusque, 2018; Cretignier et al., 2020a) containing maximum shift information. It is also known that shifts are easiest to detect in regions of the spectrum where the gradient, $|\frac{df}{d\lambda}|$ is high.

Let's take the traditional route of measuring the radial velocity of a single mean line that contains 'all' the information in the spectrum.

One technique for calculating the global shift between two comparable signals is cross-correlation (Faria, 2018). To determine the wavelength shift and mean shape of each spectral line, we can cross-correlate the individual spectra using a carefully chosen mask made of box-shaped emission lines (Baranne et al., 1996). This process allows us to generate the cross-correlation function (CCF, See Figure 1.5), which is a mean line centred at λ_{obs}, with a shape that combines all of the spectral lines. By fitting an appropriate model—typically a Gaussian

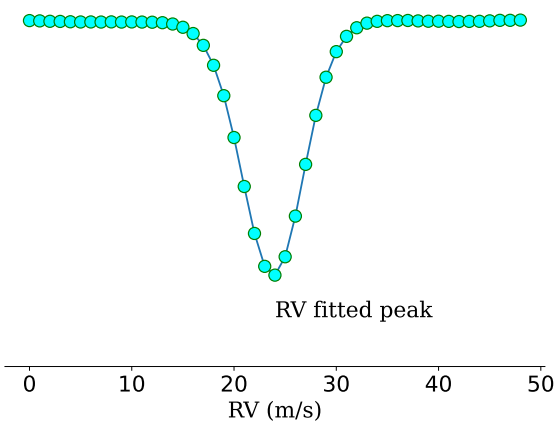


Figure 1.5: The cross-correlation function (CCF) obtained after cross-correlating an instantaneous spectrum of HD 144579 (G8 spectral type) with a digital line mask of G9 spectral type.

function—to the CCF, we can then determine the radial velocity. Deviations from the Gaussian shape are typically much smaller than 1% in observed CCFs (Faria, 2018). This technique has been widely used as a standard procedure to extract RVs from spectra obtained using high precision (eg; HARPS, HARPS-N, CARMENES) and extreme precision spectrographs (eg; EXPRES, ESPRESSO, NEID). However, the cross-correlation method is far from ideal. Because the box-shaped emission lines of the binary mask are not associated with all the real spectral lines (Bouchy et al., 2001). As a result, the entire spectral information is not utilised.

1.1.3 An unfilled lower triangle

When searching for other worlds, we are particularly interested in finding something similar to ours. The wave of groundbreaking surveys that followed the discovery of the 51 Pegasi planetary system has led to 5599 verified detections. Most of these are from transits, and relatively few of these have reliably measured RV amplitudes. Many are therefore 'validated' but not 'confirmed'. The revolutionary ground-based and space-based photometric surveys include HATNet (Bakos et al., 2004), WASP (Pollacco et al., 2006; Collier Cameron et al., 2009), *CoRoT* (Auvergne et al., 2009), *Kepler* (Koch et al., 2010), *K2* (Haas et al., 2014), *TESS* (Ricker et al., 2015) and *CHEOPS* (Isaak & Benz, 2019). High precision RV spectrographs including *HARPS* (Mayor et al., 2003), *HARPS-N* (Cosentino et al., 2012), *CARMENES* (Quirrenbach et al., 2013), and new generation extreme precision spectrographs like *ESPRESSO* (Di Marcantonio, 2019), *NEID* (Halverson et al., 2016) and *EXPRES* (Blackman et al., 2020) have been contributed to characterising these systems and filled up the complementary information space.

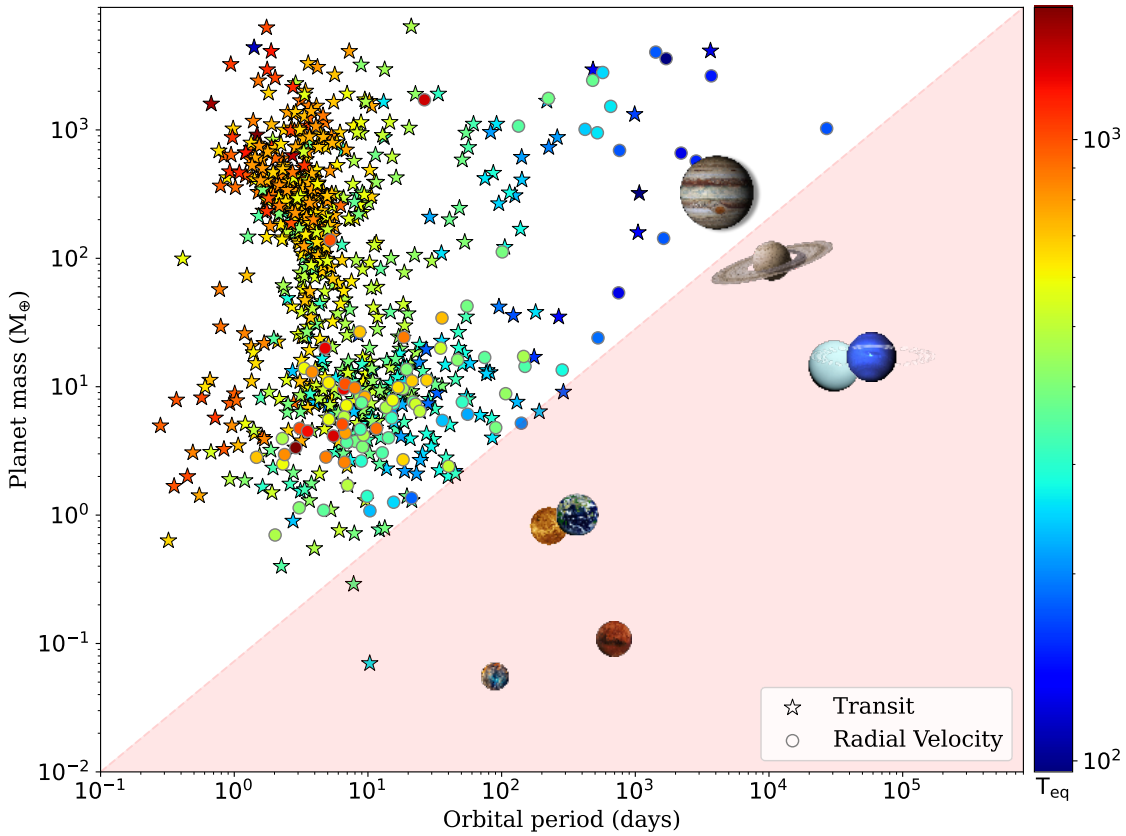


Figure 1.6: All confirmed exoplanets. The planets displayed here have both mass and orbital period estimates or measurements, and their colours correspond to their estimated equilibrium temperature. The planets in the Solar System are also shown for reference. These planets' sizes are not proportionate.

Figure 1.6 shows these confirmed transiting and RV detections (which constitute 94% of all confirmed exoplanets) with their measured orbital period and planetary masses. The term 'confirmed' in this context pertains to transiting planets whose planet-candidature has been verified through the RV method and whose masses have been measured. While appreciating yesterday's discoveries, Figure 1.6 also points towards one of today's challenges; the poorly explored 'lower triangle' in the period-mass space. For RV-only discoveries, the boundary follows approximately a contour of constant RV amplitude in mass-period space, defined by the sensitivity of current instruments. At high planet masses, the boundary turns vertical at the point (currently about 30 years) where the orbital period becomes equal to the observational duration.

As of yet, no planets with masses and orbits comparable to our Solar System planets have been discovered. Except Jupiter, all other Solar System planets fall on the fairly 'empty' side of the big picture. This is because Jupiter has a high enough RV amplitude for detection and a short enough period to have completed $\sim$ 3 orbits since the big RV surveys began in the mid-

1990s. Our Solar System is devoid of super-Earths (planets with masses several times that of the Earth), and close-in planets (planets with orbital periods ranging from a few to tens of days), that are frequently found in other planetary systems. There ought to be two reasons for this bottom triangular vacuum. A selection effect, which can be astrophysical or instrumental, maybe the cause, or our Solar System may be remarkably unusual (Mishra et al., 2023). To determine whether the latter scenario is true, we must address the selection effect.

In the next section, I will talk about this problem in the context of RV detections.

1.2 Why is it difficult to find Earths around other Suns?

For the detection of Earth-sized low-mass exoplanets orbiting in the habitable zones (HZ) of Sun-like stars, an RV precision of 0.09 m s^{-1} is essential (See Figure 1.7). Figure 1.6 showed that, if we look into our Solar System from outside and search for planet signatures in the Solar spectra using the existing instrumentation and astrophysical understandings, we could only find Jupiter with ease. Jupiter is the low-hanging fruit in terms of detectability, owing to its large size, mass and the 12 m s^{-1} level RV fluctuation (Figure 1.3) it induces on the Sun's spectra.

In this section, I'll go over, 1) where we are with the instrumentation, 2) why we haven't been able to detect Earths yet, and 3) what are attempts underway to overcome the obstacles.

1.2.1 The state-of-the-art instrumentation

In our relentless pursuit of habitable worlds, we are at the turn of the tide in the realm of ground-based RV, as the cutting-edge extreme precision planet-hunting spectrographs are locked onto a remarkable goal of achieving sub m s^{-1} level RV precision in the optical domain. The groundbreaking discovery of 51 Pegasi b (Mayor & Queloz, 1995) in 1995 was made using the Doppler spectroscopy from the *ELODIE*, a spectrograph with a measurement precision of 13 m s^{-1} . By early 2000s, we have advanced the long-term precision of spectrographs down to 1 m s^{-1} (eg: HARPS Mayor et al. (2003)) and further down to 0.5 m s^{-1} by HARPS-N. Driven by the quest for Earth-like planets, newly deployed instruments now aim for an extreme 10 cm s^{-1} precision.

As this thesis revolves around the high-precision RV observations obtained with the HARPS-

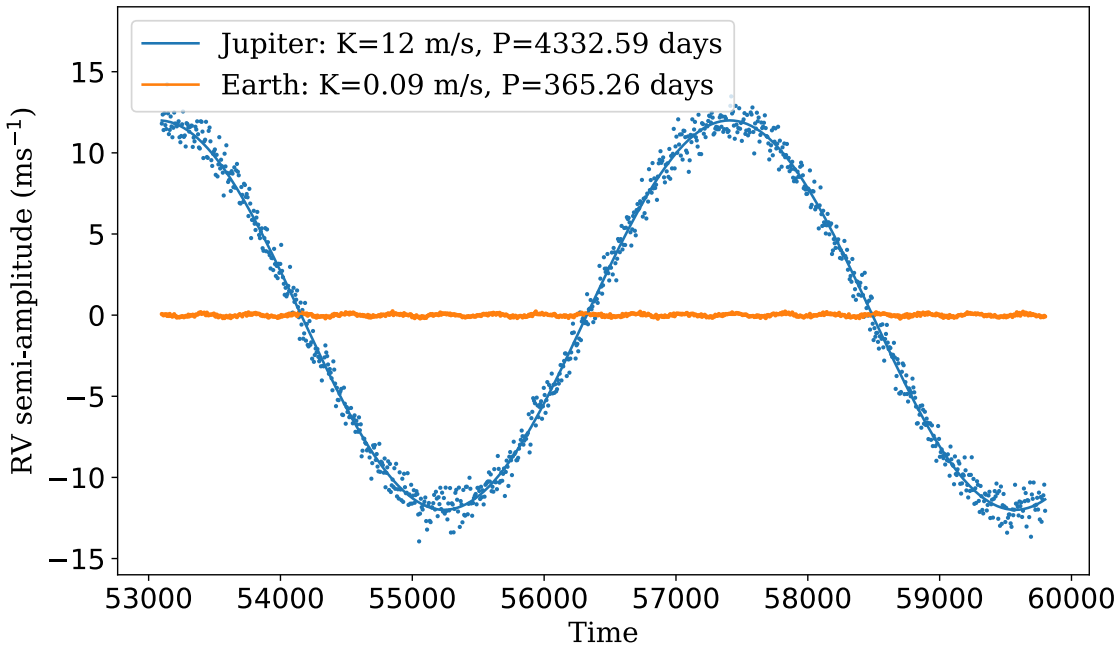


Figure 1.7: Simulated RV sinusoids for Jupiter and Earth. The semi-amplitudes refer to the amplitude of the Sun's reflex motion in response to the gravitational tugs of respective planets. The more massive the planet, the greater the semi-amplitude. The periods given here are the sidereal periods, which is what a distant observer would see.

N instrument, let's start there. HARPS-N is the Northern Hemisphere twin of HARPS, mounted at the 3.57-meter Telescopio Nazionale Galileo (TNG) in La Palma, Spain (Cosentino et al., 2012). It offers a long-term calibration precision better than ∼0.5 m s^{-1} (Dumusque et al., 2021) and has been in operation for 12 years now. Over these years HARPS-N has been characterising numerous transiting Kepler, K2, *TESS* and *CHEOPS* systems while also finding new non-transiting planets with commendable precision (Motalebi et al. (2015); Anna John et al. (2022); Stalport et al. (2023), to name but a few).

To push towards the targeted 10 cm s^{-1} precision, NEID was deployed in 2019 on the 3.5-meter WIYN Telescope at Kitt Peak National Observatory in Arizona. Concurrently, the EXtreme PREcision Spectrometer (EXPRES) also became operational at the Lowell Observatory, offering a comparable sub m s^{-1} level RV precision. The Echelle SPectrograph for Rocky Exoplanets and Stable Spectroscopic Observations or ESPRESSO is another new-generation extreme-precision instrument mounted on the 8-m VLT at Paranal Observatory, Chile. ESPRESSO also has an ultra-stable design capable of achieving a long-term calibration precision of ∼ 0.1 m s^{-1} , aiming to detect Earth twins orbiting in the habitable zone of solar-like stars (Pepe et al., 2021).

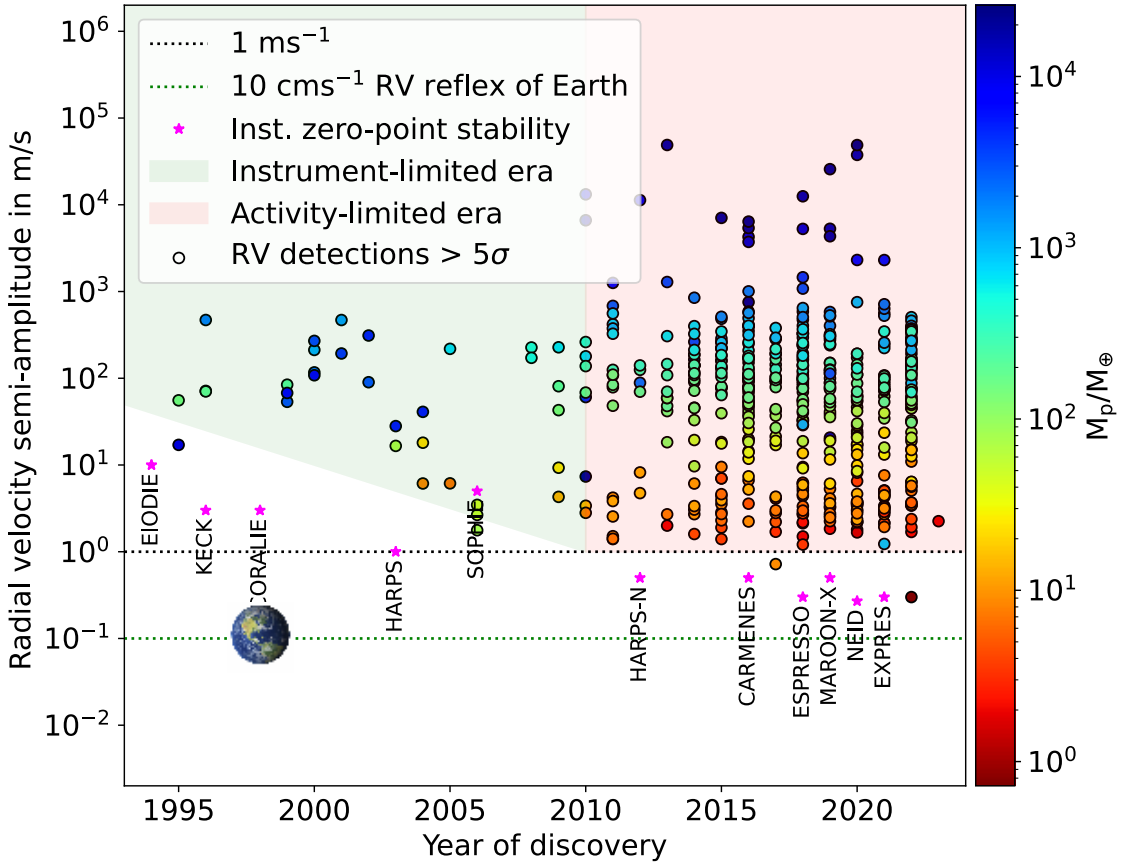


Figure 1.8: RV semi-amplitudes v/s discovery year (Figure: Anna John et al. (2023), Data: NASA's Exoplanet Archive). The exoplanetary reflex motion detection limit plummeted to 1 m s^{-1} from 1995 to 2010, driven by improved wavelength calibration stability in ground-based RV spectrometers. However, despite the impressive advances in instrument precision of new-generation spectrographs, this trend plateaus post-2010 at ∼1 m s^{-1} : an RV amplitude comparable to stellar variability.

1.2.2 The 1 m/s noise floor

Between 1995 and around 2010, the RV detection limit for exoplanetary reflex motion was driven towards 1 m s^{-1} by incremental improvements in the wavelength-calibration stability of ground-based RV spectrometers (Figure 1.8). However, despite further instrumental advancements, the declining trend in the detectable RV semi-amplitude has plateaued after 2010 at ∼ 1 m s^{-1} , an amplitude comparable to RV variations produced by the intrinsic variability of the host star. An Earth analogue orbiting a solar-type star with an orbital period of 1 year has a radial velocity amplitude of 9 cm s^{-1} . Compared to the best activity-mitigation techniques currently in use, this is about an order of magnitude lower (Zhao et al., 2022). Therefore, to determine the masses and orbits of HZ Earths orbiting nearby F, G, and K spectral-type stars, we need to drill through this 1 m s^{-1} noise floor induced by stellar variability and obtain RV measurements with a precision better than ∼10 cm s^{-1} (with a few cm s^{-1} stability over many

years (Crass et al., 2021).

1.3 How stellar activity affects RVs

We would have undoubtedly detected a lot more exoplanets, particularly smaller ones, if the signals in the stellar RV time series were exclusively caused by exoplanets. But, the stars are not uniform fireballs! Various physical processes transpiring across the stellar photospheres over varying lifespans induce RV variations, which complicates the detection of planetary signals.

Stellar variability is an umbrella term for a variety of astrophysical processes (Figure 1.10) that have an impact on the RV of stars on different timescales. Stellar brightness fluctuations, resulting from the bright and dark features traversing the visible stellar disc as the star rotates, are obvious manifestations of these surface inhomogeneities. Thanks to the *Kepler* (Koch et al., 2010; Haas et al., 2014) and *TESS* (Ricker et al., 2015) space telescopes, these fluctuations have now been observed for hundreds of thousands of stars with high precision through extensive stellar rotation surveys (eg: McQuillan et al. (2014); Basri et al. (2011)). Another interesting (yet annoying!) manifestation is that these variability features can entirely mask or mimic exoplanet signals because of their similar timescales. For a $1M_{\oplus}$ planet circling a $1M_{\odot}$ star, the RV variation is approximately 0.1 to 0.3 m s^{-1} for orbital semi-major axes of 1 and 0.1 AU, respectively (Pagano, 2013). By suppressing the flux emitted by parts of the rotating stellar disc (Aigrain et al., 2012), these variability features introduce RV jitter greater than the amplitude of RV variations attributed to the planets. In this section, we will see different processes that come under the stellar variability umbrella, and how they affect the efficiency of RV detections.

1.3.1 Variability timescales and RV signatures

As the Sun is the only star whose magnetic activity phenomena can be observed with incredible spatial detail (down to the sub-arcsecond level), the interpretation of stellar activity generally relies on the paradigm known as the 'solar-stellar connection', which states that the majority of the variability phenomena observed in late-type stars can be decoded by using solar analogous phenomena and their interpretation as pointers (Pagano, 2013). This strategy has also been effective in explaining the variations from the solar case.

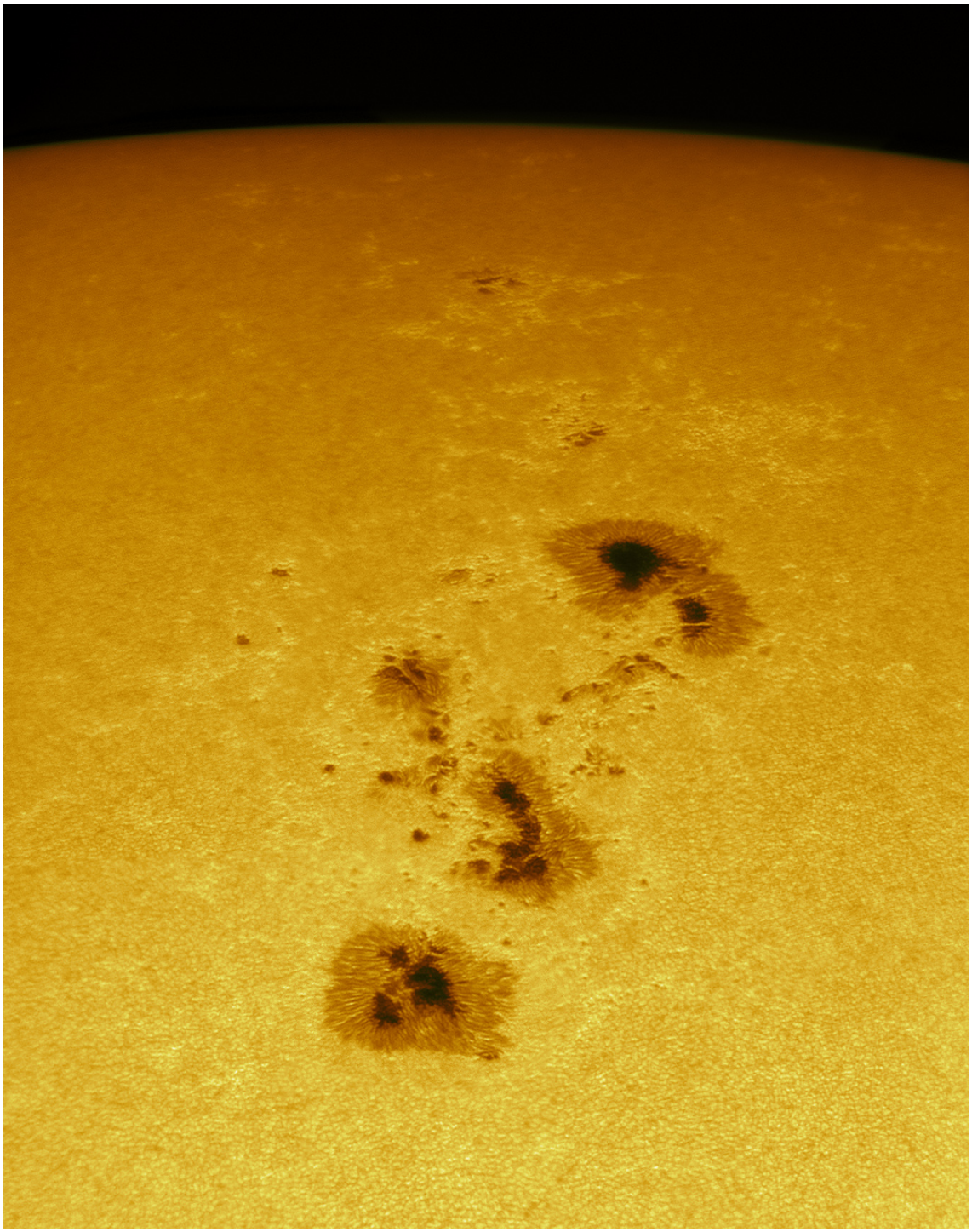


Figure 1.9: A piece of the non-uniform Fireball!
This is a high-resolution image of our Sun, captured by Alan Friedman on 07 May 2015. This high-resolution composite image shows the surface inhomogeneities on the Solar photosphere (the visible Solar surface). Spots, which are cooler and darker magnetised regions, are surrounded by bright faculae. The granular network cells formed by rising and falling plasma are also visible all over the photosphere. For more such images visit http://avertedimagination.com.

.

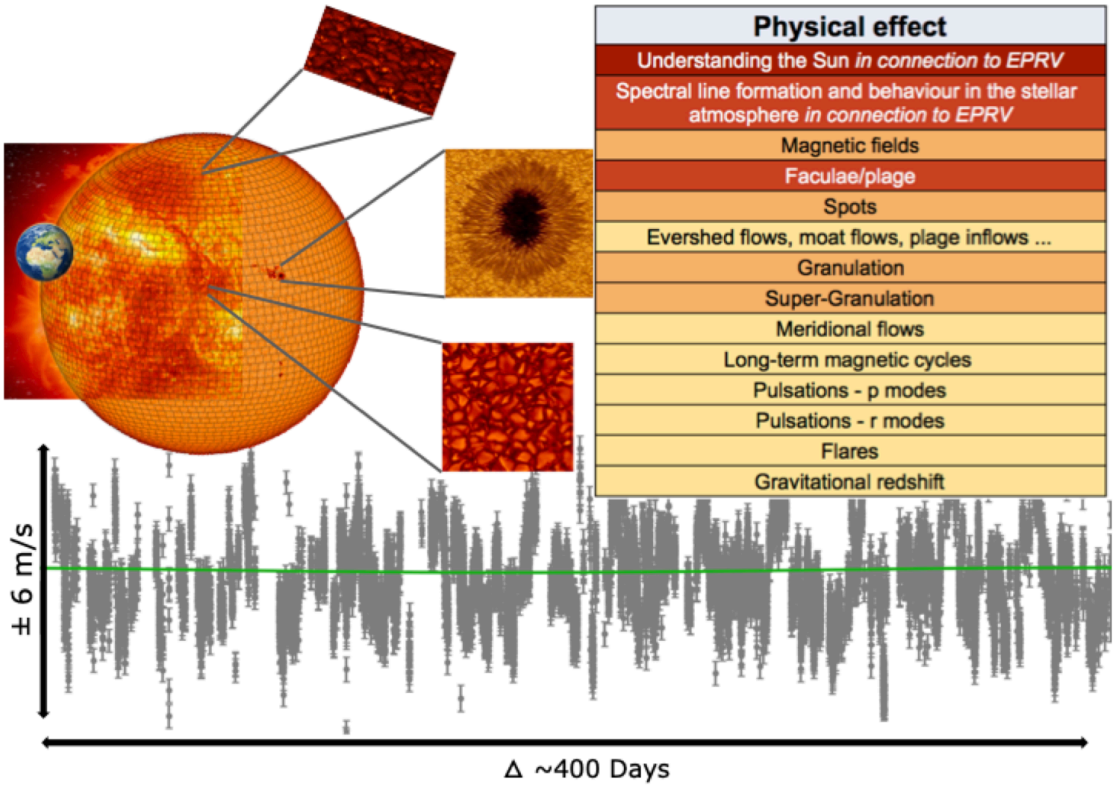


Figure 1.10: Figure and caption from the Extreme Precision Radial Velocity (EPRV) final report (Crass et al., 2021). Top left: Artist impression of the stellar variability challenge to EPRV. Top right: physical phenomena known to induce RV variability, colour-coded by importance to EPRV (ranging from red: urgent & critical; to yellow: long-term & significant). Bottom: HARPS-N RV observations of the Sun seen as a distant, point-like star (gray points; Collier Cameron et al. (2019)), and signal of an Earth twin (green line).

The Sun 'lives' from the energy generated in its core via nuclear fusion. The convective motions transfer energy from its radiative interior to its photosphere, where the plasma density is lower. It is this convective motion and its interaction with the magnetic fields that fuels the 11-year magnetic cycle, solar flares and many other processes (Hathaway et al., 2015). The fundamental cause of all phenomena referred to as "solar activity" is the Sun's large-scale magnetic field, which regenerates cyclically.

In the context of this thesis, I will only be talking in detail about processes like p-mode oscillations, granulation, short-term activity signatures of spots and faculae and long-term magnetic activity cycles, considering their contribution to the RV budget. Among these, only the latter two are the product of the stellar magnetic fields.

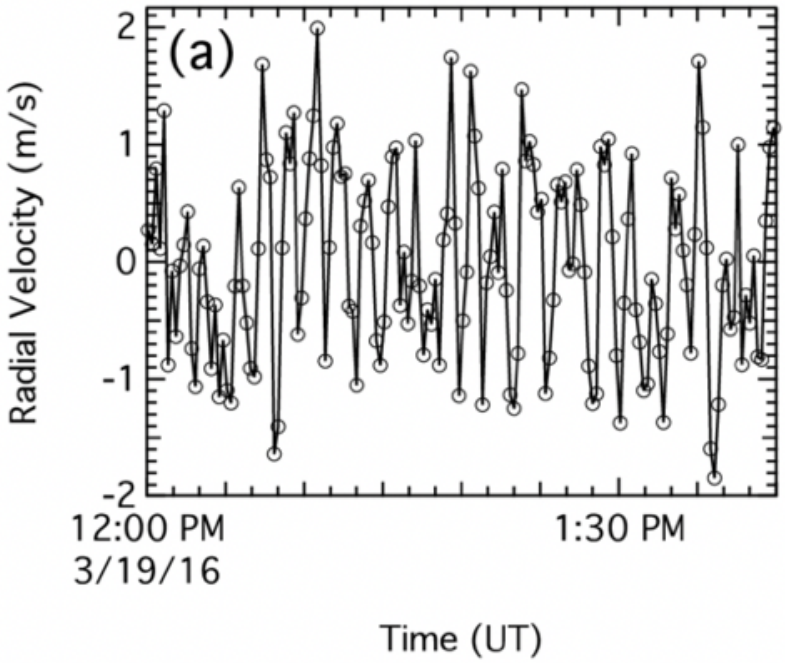


Figure 1.11: Solar radial velocities from two hours of 20-second solar exposures demonstrating solar p-mode oscillations (∼5 minute oscillations). Figure from Phillips et al. (2016).

Oscillations

Stars breathe! Constant pressure fluctuation within the star produces sound waves that travel through the solar/stellar interior and drive pressure-driven oscillations known as p-modes, which cause ripples on the surface of the Sun/stars (Ulrich, 1970). Helioseismologists and asteroseismologists use these oscillations to probe the solar/stellar structure, much like seismologists use seismic waves from earthquakes to study the interior of the Earth.

Thanks to the high-precision and extreme-precision spectrographs, we can now directly observe these acoustic modes. Figure 1.11 demonstrates that the RV amplitude of the solar p-modes is approximately 2 m s^{-1} when sampled at a 20 s cadence. These variations occur on a timescale of 5-15 minutes and produce RV variations with an amplitude of 0.1 to 4 m s^{-1} , depending on the stellar type and evolutionary stage (Brown et al. (1991); Schrijver & Zwaan (2000); Dumusque et al. (2011a) and references therein). As it can mask the RV signatures of small planets, it is now a standard procedure to obtain exposures of $>$ 10 minutes, to average out the RV effects of these short-frequency oscillations (Dumusque et al., 2011b). The ability to average out perturbations resulting from stellar oscillations using such specialised observational strategies, together with the stability and accuracy of new-generation high-resolution

spectrographs has resulted in the detection of small planets with masses between 2 and 10 $M_{\oplus}$.

Granulation

Granules are bright regions on stellar surfaces caused by upflowing hot plasma, encircled by dark intergranular lanes where cooler plasma is downflowing. Except for active regions where convection is noticeably lower, these convective phenomena are present throughout the stellar surface (Meunier et al., 2010; Gray, 2009). Granulation represents a discrete convective pattern with a diameter of less than 2 Mm and a lifetime of less than 25 minutes (Del Moro, 2004). These surface inhomogeneity features induce RV signatures due to ascending granules being blue-shifted and descending intergranular lanes being red-shifted. The solar observations from HARPS-N show that solar granulation induces RV variations with an amplitude of 0.4 m s^{-1} . Given the granular radial flow speeds of the order of 1 km s^{-1} , one could legitimately wonder why this figure is so small. The explanation is statistical: since the Sun has a million granules, the variations in the velocity field are averaged over roughly a million independent surface elements. Integrating over the visible hemisphere, this leads to a reduction in RV amplitude of order $\sqrt{n} \sim$1000.

We can modify our observing strategy to reduce their impact on RV measurements in a similar way that we do for p-modes. Currently, observations using high-precision spectrographs are strategically crafted to alleviate granulation effects on RVs by averaging two or three exposures of at least 15 minutes, taken $>$ 2 hours apart within a night.

There also exists a larger version of convective cells, referred to as supergranules, with sizes between 40 and 50 Mm in size and timescales of the order of a day (1.8 days for the Sun) (Del Moro, 2004). Although granulation has been extensively studied for the Sun and is the easiest to average due to its short associated timescales, super granulation is potentially more troublesome due to its longer lifetime Meunier & Lagrange (2019). Super granules cover the whole Sun, whose large sizes make them appear to be fewer in number at any given time. This complicates the process of averaging out, and their resultant RV jitter remains substantial (0.28 to 1.12 m s^{-1} ; Meunier & Lagrange 2019). Recently, super granulation was exposed to be a primary limiting factor in the detection of exo-Earths (Meunier & Lagrange, 2019), using simulated RV variations for a G2 type star.

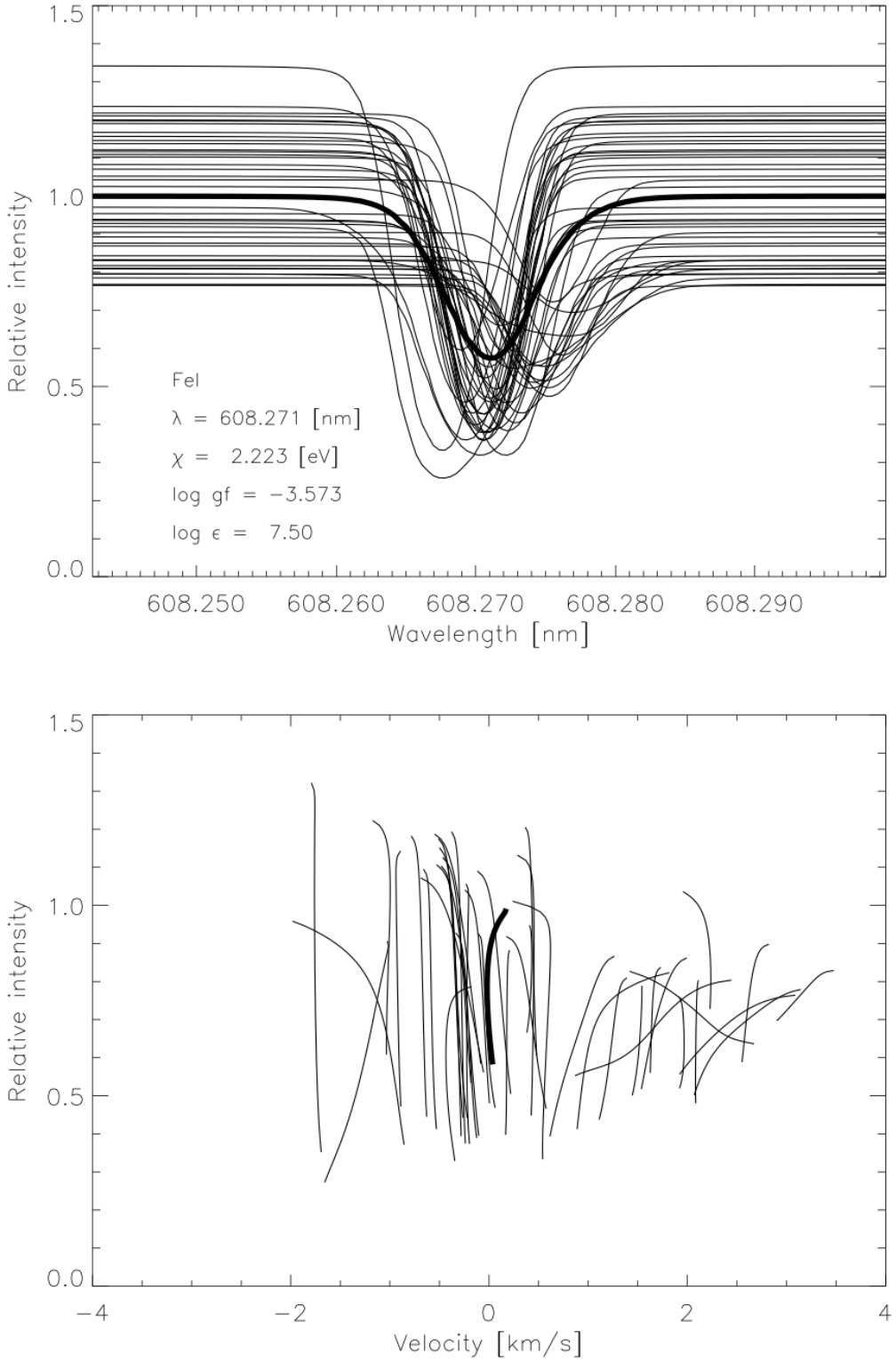


Figure 1.12: Spacially resolved spectral line profiles and bisectors for the FeI608.2 line. In comparison to the red-shifted intergranular lanes, the lines in the blueshifted granules are both stronger and have higher continuum intensities. The downflows contain the highest vertical velocities found in the photosphere. The averaged profile, or CCF, and its bisector are represented by the thick solid lines. Figure and caption from Asplund et al. (2000).

The manifestation of stellar granulation can be observed on either spatially resolved observables (like images, though this is currently only possible for the Sun) or unresolved observables (like spectral line profiles) in terms of their widths, shapes, and strengths. Image degradation resulting from limited instrumental resolution, instrument scattering, and atmospheric blurring and distortion in the case of Earth-based observations inevitably affects direct observations of sub-arcsecond size structures. Conversely, the spectral line shapes encode important properties like their velocity amplitudes and velocity-intensity correlations (Nordlund et al., 2009). While the amplitude of convective velocity field severely affects the widths of spectral lines, the correlations of velocity and temperature cause net blueshifts and introduce asymmetries of spectral lines (Gray, 2005). See Figure 1.12. Both of these contribute to an RV variation of sub-m s^{-1} amplitude.

Spots/Faculae

Stellar surface features such as spots, and networks of faculae induce photometric and spectroscopic variations modulated by the star's rotation (Kron, 1947; Dorren, 1987). Spots appear as dark regions on the surface of the star and have a smaller flux relative to the average flux, owing to their lower temperatures. These strong magnetic regions pose magnetic field strengths (2900 G; Hale (1908); Priest (1982)) thousands of times stronger than the magnetic field on Earth (0.25 to 0.65 G). Spots generally live for a few days, though very large ones can live for weeks. On the other hand, faculae (plural of facula, the Latin word for 'little torch') are brighter features mostly found in the intergranular lanes and surrounding spots. Although individual faculae only last a few hours, groups of faculae can live longer on the stellar surface (for several weeks and can even survive multiple stellar cycles!). Faculae outlast spots and always appear before them (Priest, 1982).

The short-term variations brought on by stellar rotation in the presence of active regions influence the RV precision in distinct ways and thereby severely hamper the detection of exoplanets. One effect arises from the temperature disparities between these active regions and the surrounding stellar surface, commonly referred to as the flux effect (Dumusque et al., 2014), which is the source of the most comprehensible perturbation affecting the RVs. The flux emitted by the approaching half of a rotating star is blueshifted, while the flux emitted by the receding half is redshifted. This is because one half of the disc is moving towards us and the other half is moving away. If the stellar surface is featureless, the Doppler shifts from

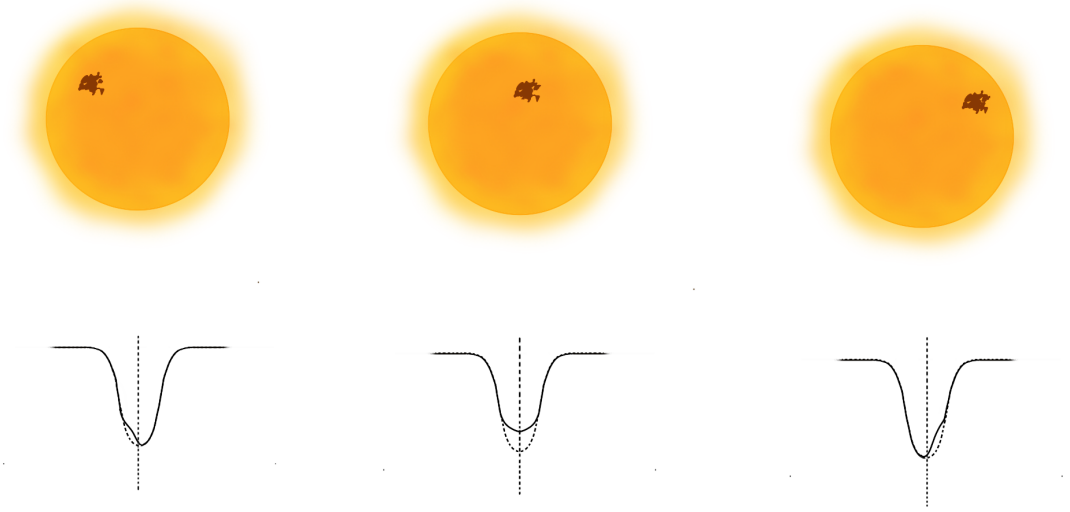


Figure 1.13: A schematic representation of how flux blockage due to star spots induces asymmetry to the CCF/spectral line shape, which results in an RV variation.

both sides cancel each other out and the spectral line profile is unaffected (dotted curve in Figure 1.13). However, spots and faculae disrupt the flux balance between the two hemispheres, whose rotation causes their Doppler velocities to be of opposite signs. This results in an asymmetry in the overall line profile's shape, slightly shifting its centroid. This also creates quasi-sinusoidal variations that mimic the Rossiter-McLaughlin effect brought on by a transiting planet, with a periodicity half that of the stellar rotation (Al Moulla et al., 2022; Aigrain et al., 2012; Dumusque et al., 2014).

Moreover, because of the strong local magnetic field, they inhibit the underlying convective motion, which results in a modulation with a stellar rotation period. This attenuates a portion of the granulation-induced convective blueshift (Gray, 2009). Hence, in contrast to the quiet photosphere, spots and faculae appear redshifted. Meunier et al. (2010) demonstrated that the dominant contribution in solar simulations is the suppression of convective blueshift, which, in contrast to the flux effect, cannot balance itself out when the approaching and receding hemispheres are equally covered. The suppression of convective blueshift can cause a net RV variation of several m s^{-1} as they emerge and vanish from the visible stellar hemisphere. On magnetically active stars, starspots and faculae can cause RV shifts of the order of 1–

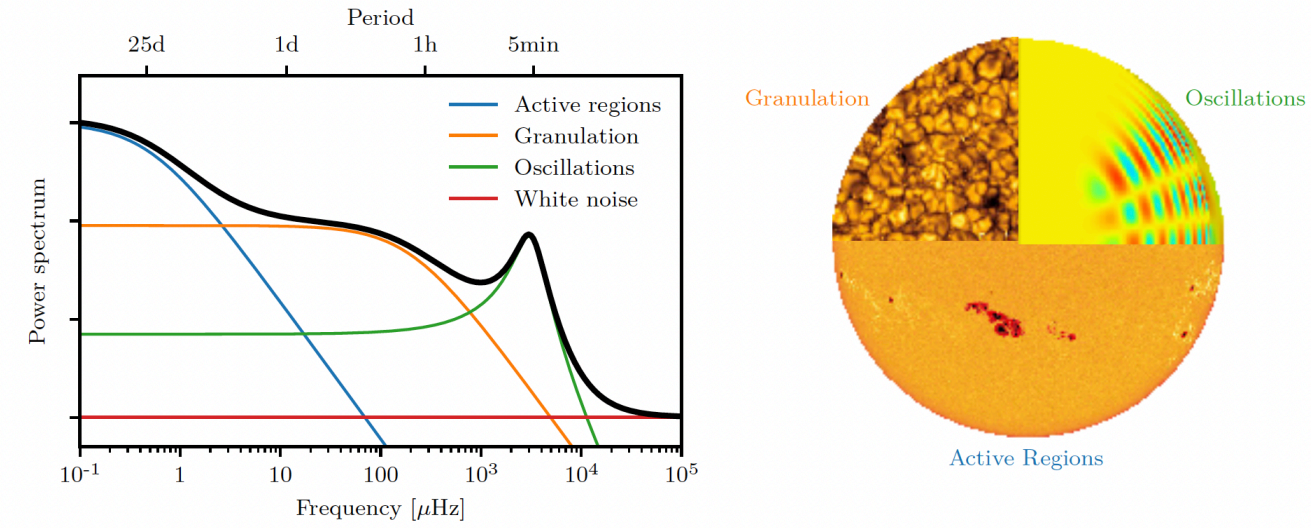


Figure 1.14: The power spectrum for the various activity-induced signals, illustrating the timescales linked to stellar activity. Figure from Faria (2018).

100 m s^{-1} (Wright, 2005). Collier Cameron et al. (2019) found that active regions' passage across the visible solar disc produces signals at 5 m s^{-1}. Even in the quietest stars, the intrinsic magnetic variability of the host stars generates changes of 1-2 m s^{-1} (Isaacson & Fischer, 2010).

Magnetic activity cycles

Since Schwabe (1844) first observed the roughly 11-year oscillation in Sunspot numbers, Sunspots have been regarded as the canon of solar variability (Leamon et al., 2020). An increase in Sunspots, faculae, and plages, as well as a rise in the frequency of violent events like prominences and coronal mass ejections, are indicators that the 11-year cycle is moving towards the activity maximum. These magnetic activity cycles span a period similar to those of Jupiter-like planets, and they have been observed in a large number of solar-type stars (Baliunas et al., 1995; Soderblom, 1985).

The detection of solar system analogues (low-mass, long-period planets) is affected by stellar magnetic cycles on similar timescales. From an extensive review, Santos et al. (2010) quantified the RV variations induced by stellar magnetic cycle to be of amplitudes $\sim$1 m s^{-1}. It was previously known that the flux in the Ca II H & K lines is correlated with the number of Sunspots, and thus can be used as an indicator of chromospheric activity (Leighton, 1959).

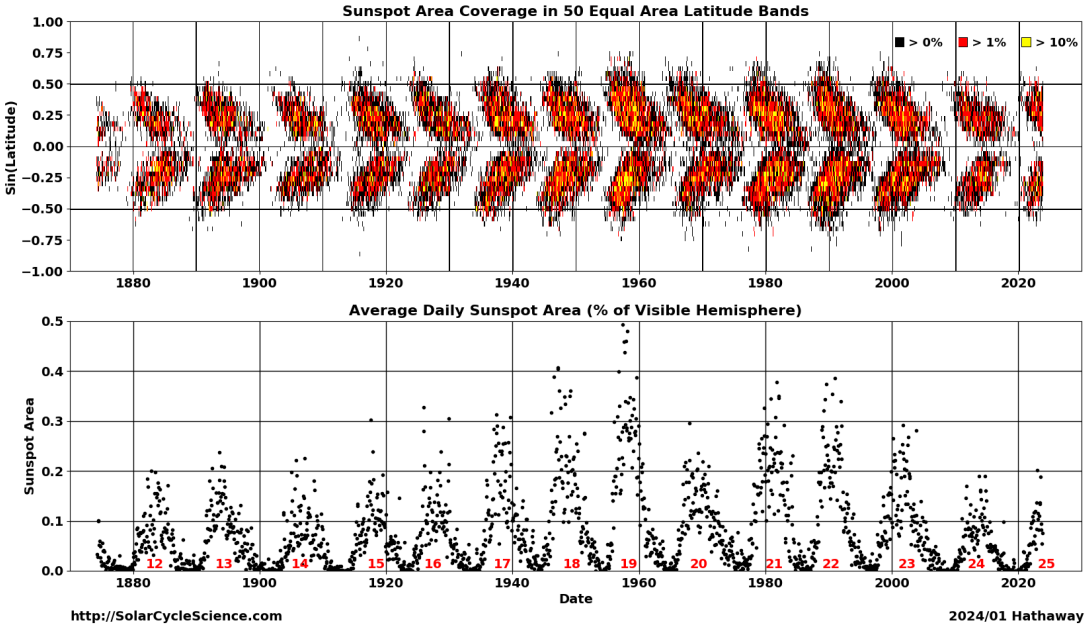


Figure 1.15: Butterfly diagrams showing the distribution and temporal evolution of the Sun's magnetic flux and Sunspot populations. Bands of Sunspots can be seen on both sides of the equator. The active regions appear at latitudes of roughly 30-35 degrees at the start of each cycle. As the cycle advances, active regions move closer and closer to the equator (Sporer's law). Figure and caption from http://solarcyclescience.com/solarcycle.html.

Therefore, simultaneous Ca II H & K measurements could be used to diagnose (and most likely correct) RV variations induced by magnetic activity cycles. Based on this understanding, the Mount Wilson survey (Vaughan et al., 1978; Vaughan, 1980; Vaughan et al., 1981) was carried out to study the rotational modulation using the Ca II H & K fluxes of late-type main sequence stars using the 1.5 m telescope at Mount Wilson Observatory.

1.4 Stellar activity mitigation in literature

In this section, I will give a brief rundown of the tools and techniques that have been developed to improve the reliability of the detection of planet signals in the presence of stellar activity. Starting with the early (traditional) attempts at decorrelating against activity indicators, we will discuss a variety of time-domain strategies such as the FF′ method and, single and multivariate Gaussian Process regression techniques, before offering a sneak peek at strategies that address stellar variability in the wavelength domain.

1.4.1 Strategic observations

Developing observational strategies to average out the radial velocity signatures of stellar noise is a useful approach, as was mentioned while talking about different types of stellar activity. Using real and synthetic RV measurements, Dumusque et al. (2011a,b) investigated various observational strategies that yield the smallest RV variation due to stellar noise resulting from the effects of granulation, oscillations, active regions, and instrumental errors. Due to the minutes timescale of p-mode oscillation, the effective observational approach employed in HARPS and HARPS-N -GTO observations is to obtain one measurement of 15 minutes per night. Although this has been effective in mitigating the impact of oscillation-induced RVs, it was not successful in averaging out granulation phenomena, owing to various timescales involved. According to Dumusque et al. (2011b), a more effective way to decrease granulation noise and enhance planetary detection limits is to increase the number of measurements made each night and bin those observations, which will sample the granulation time scales. Ever since, the adopted approach has been to bin the multiple 15-minute observations taken over the course of a single night, separated by at least two hours. When it comes to spots, obtaining three measurements per night of 10 min every three days, 10 nights a month seems the best-tested strategy (Dumusque et al., 2011a).

Even though all of these observational techniques have demonstrated their efficacy as anticipated, doing so requires a significant investment of telescope time. Furthermore, the averaging out strategies restrict our ability to comprehend the related phenomena.

1.4.2 Activity Indicators

We measure the RV of a star from the cross-correlation function (CCF), which represents the mean of all the lines of a spectrum. Activity indicators derived from the shapes of spectral lines (or CCFs) provide a direct connection to the intrinsic processes of stellar variability. These indicators are usually moments of the CCF or measures of its asymmetry. Moreover, the stellar chromospheric activity has been widely tracked using Hα lines and S-index, which is the ratio of the flux in the core of the Ca II H and K lines to the surrounding continuum. The basic idea is that these indicators are only affected by stellar activity, so any variations found in RV observations that are not seen in these indicators could indicate the presence of a planetary signal. This section focuses on how various types of stellar activity manifest in the different

CCF parameters. I will particularly focus on three CCF shape measures that have been used to identify activity-induced signals in RV data.

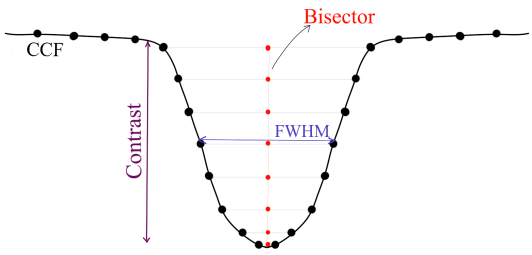


Figure 1.16: Illustration of various CCF shape parameters like FWHM, bisector and contrast.

Full-Width at Half-Maximum

The full width at half maximum (FWHM) of the CCF for stable spectrographs, like HARPS and HARPS-N, reveals profile variations brought on by stellar activity (Boisse et al., 2011)). As a facular region or spot traverses the stellar disc, the FWHM modifies to preserve the area that the line profile encloses (Figure 1.13). Variations in the FWHM are therefore correlated with RV perturbations resulting from the flux blocked by starspots on a rotating star.

FWHM has been used in many exoplanetary investigations as a 'traditional' proxy to probe the intrinsic stellar variability of the host star (Queloz et al. (2009); Hatzes et al. (2010), to name but a few.)

Bisector span (BIS)

Another CCF measure that points towards this direction is the Bisector Inverse Slope or BIS. A bisector is a continuous curve that divides the spectral line into two equal parts of equivalent width, by defining the locus of median points halfway between equal intensities on either side of the line (Figure 1.16). Due to convective asymmetry in solar and stellar line profiles brought on by the brightness and velocity structure of the photospheric granulation pattern, the CCF itself is not perfectly symmetric (Collier Cameron et al., 2019). BIS defines the difference in

line bisector velocities ($V_t - V_b$) in the (t)op=wings and (b)ottom=core portions of the CCF (Queloz et al., 2001). It is a measure of the asymmetry of the CCF/spectral line profile (Gray, 2009) while representing the apparent RV at different depths of the line (Dravins et al., 1981). If the CCF is a perfect Gaussian, then the bisector will be a straight vertical line that cuts the CCF through the centre, as shown by the red line in Figure 1.16. However, granulation causes asymmetric and wavelength-shifted profiles (Dravins et al., 1981). The bisector gets distorted and bent upwards as a result of the net convective blueshift caused by granulation (See the thick line in Figure 1.12). This is due to the blue-shifted photons from the brighter granules having a dominant effect because of the contrast between the darker descending flows in the intergranular lanes and the bright rising cores of the granules. The bisector will exhibit further minor distortions when active regions inhibit this net blueshift.

The BIS was first applied to exoplanet studies by Queloz et al. (2001) and has been used in several later studies as well. This activity proxy does have some limitations though. It is now known that BIS cannot offer adequate 'activity' information in the case of slow-rotating Sun-like stars with lower $v\ sin\ i$ (Haywood, 2016). For these slow rotators, the FWHM sometimes exhibits a stronger correlation than the BIS with the activity-induced signals observed in photometry (Queloz et al., 2009; Santos et al., 2014).

Equivalent Width

The equivalent width provides a precise definition of line width that is unaffected by line shape. The CCF's equivalent width (EW) can be defined as the width of a rectangle with length one and an area equal to the CCF's area. In other words, if A is the area of the CCF profile below the continuum level, then the equivalent width w is the width of the rectangle with the same area A.

The equivalent widths of the CCFs can change throughout the activity cycle due to the influence of magnetic fields and stellar spots (Spina et al., 2020). Collier Cameron et al. (2019) also noticed the same effect in HARPS-N solar data. Therefore, it can be used as an efficient activity cycle tracker. During the stellar magnetic cycle, the equivalent widths of lines can increase as a function of the chromospheric activity index, $\log R'_{HK}$. Over the course of the cycle, the lines originating from the outer layers of the photosphere experience the greatest fluctuations in their equivalent widths. Since stronger lines form higher above the stellar photosphere, the equivalent width value can be used as a proxy of the line optical depth (Gray,

1992). In contrast to the BIS, there is minimal indication of rotational modulation in the EW. It is hard to definitively determine whether the EW variation is primarily caused by cool stellar spots, the Zeeman broadening, or a combination of the two (Spina et al., 2020). According to Collier Cameron et al. (2019), the chromospheric network, which is relatively homogeneous over the Sun, maybe the source of the EW variation. It is also found that the EW is greater at the time of solar minimum.

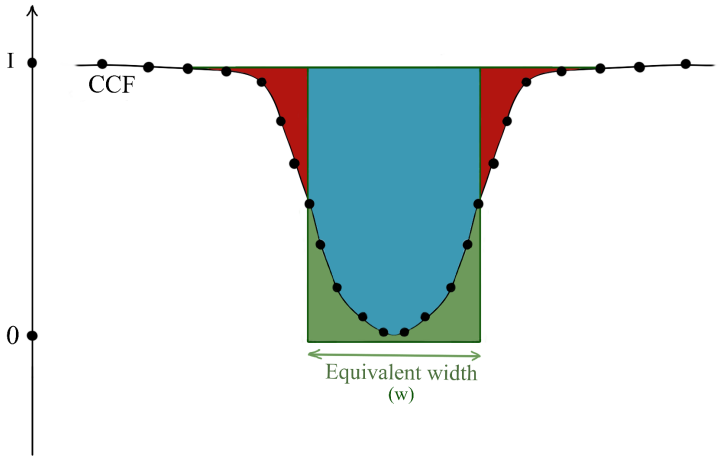


Figure 1.17: Illustration of Equivalent Width of a CCF profile. The concept is to construct a rectangular box with the same area of that of the profile, that extends from the continuum to the 0 flux line. The width of the resultant box measures the Equivalent width of the profile, indicating the strength of the line.
Area enclosed by the blue + red = Area of blue + green.

EW $\propto$ FWHM x Contrast

Since EW is proportional to the product of FWHM and Contrast (the amplitude of the Gaussian divided by its baseline level), EW information can be obtained from the stellar spectrum or the CCF without the need for additional observations. Contrast, along with FWHM and BIS, is automatically calculated by the data reduction pipelines of the new generation of spectrographs and is easily accessible.

1.4.3 Stellar activity mitigation in Time-domain

Various analysis techniques have been developed to understand and mitigate the impact of stellar activity effects. Up till recently, time-domain linear decorrelation against activity proxies has been the most effective method for modelling stellar contamination of RV data. The early efforts include decorrelation against proxy indicators of activity, or simply rejecting a planet candidate if the same periodicity was found in the Lomb-Scargle periodograms of FWHM, BIS or the chromospheric Ca II H&K emission flux, as in the RV (e.g. Hatzes et al. (2010)). Another time-domain method, introduced by Aigrain et al. (2012), makes use of well-sampled photometric light curves to predict the stellar variability expected in RV measurements. Known as the FF$'$ method, this approach uses the light curve (or photometric flux, F) and its first derivative (F') to approximate the effects of spots and faculae and requires no information about the rotation period or complex spot modelling. It does, however, require simultaneous high-precision photometry. Since such coordinated observations are difficult to obtain in practice, the FF$'$ method is likely to provide an incomplete representation of activity-induced RV variations (Haywood et al., 2014).

Haywood et al. (2014) pioneered the usage of Gaussian Process (GP) regression analysis, to model the quasi-periodic correlated noise introduced by the stellar activity in RVs. When using GP regression, the covariance structure of the activity signal is assumed to resemble that of a driven, damped harmonic oscillator, whose periodicity is defined by the stellar rotation, the damping timescale determined by the lifetime of active regions, and the harmonic complexity depends on the number of active regions present at widely separated stellar longitudes. For a detailed review of the technique, please refer to Chapter 2 of Haywood (2016) and Section 3.3 of Faria (2018). Although GP regression with a quasi-periodic kernel is successful at fitting rotationally modulated activity, other signals are also present. The 'too flexible' aspect of the Gaussian Process is its major strength and weakness. This motivated the evolution of GP into multi-variate GPs, capable of modelling the joint variations of RVs, the FWHM, BIS, and $\log R'_{\rm HK}$ simultaneously (eg: Rajpaul et al. (2015); Barragán et al. (2022)). Multi-variate GP requires the activity contribution to the RV signal to be consistent with the shift-independent proxy indicators. Moreover, time-domain approaches in general require strategic observations and a wealth of telescope time to densely sample the stellar rotation cycle and planetary orbits to inform the quasi-periodic model, which may not always be possible given the constraints of

telescope schedules and weather. Ideally, the spectrum itself would contain the actual stellar radial velocity. Decoding this information would reduce the requirement for dense sampling and make activity mitigation possible even with sparsely sampled historical datasets. This further motivates the call for developing complementary wavelength-domain approaches based on information retrieved from the stellar spectrum's instantaneous appearance.

1.4.4 Stellar activity mitigation in Wavelength-domain

A complementary approach to model stellar activity is to analyse the modes of variation present in the instantaneous spectral line profile or CCFs, independently of their temporal behaviour. While a planet produces a shift of all spectral lines, stellar activity affects the shape of the spectral lines. Building on this understanding, several RV analysis tools have been developed to separate the planet-induced and stellar activity-induced components of RVs. The FourIEr phase SpecTrum Analysis (FIESTA, Zhao & Tinney (2020); Jinglin Zhao. et al. (2022)) uses the real and imaginary parts of the Fourier transform of the line profile to disentangle apparent RV shift due to spectral line-profile variability from true Doppler shift. Concurrently, de Beurs et al. (2021) also developed machine-learning techniques such as linear regression and neural networks to separate activity signals from true center-of-mass RV shifts using the average spectral line-shape variations, without any time-domain information. Another recent investigation (Klein et al., 2022) employs Doppler imaging to simultaneously model the activity-induced distortions and the planet-induced shifts in the line profiles. It is challenging to differentiate between the effects induced by a planet and stellar activity, as the latter has varied effects on individual spectral lines (Dumusque et al. 2015).

In light of this, Collier Cameron et al. (2021) developed another algorithm (SCALPELS – Self-Correlation Analysis of Line Profiles for Extracting Low-amplitude Shifts) to decouple the RV information from the time domain to separate the apparent shifts caused by spectral line shape variability triggered by stellar activity, from the Doppler shifts caused by planetary orbital motion. This method of shape-shift separation provides accurate Doppler detection and characterization of exoplanets around well-observed, bright main-sequence stars across a wide range of orbital periods, particularly for lower mass planets, by analysing orthogonal modes of variation in the translationally invariant autocorrelation function of the line profile. Collier Cameron et al. (2021) demonstrated the algorithm using HARPS-N observations of the solar spectrum. They successfully verified and validated that the algorithm can accurately

detect multiple simulated planets injected into heliocentric solar observations, spanning a wide range of orbital periods.

1.5 Summary

Stellar activity has an RV signature which complicates the detection of planetary signals using the RV method. Because the magnetic activity of the host star occurs on identical timescales to small RV planets, it can entirely hide or mimic their RV signals. This imposes a severe challenge in the detection and precise characterisation of these planets. Several stellar activity mitigation tools operating in the time domain and wavelength domain have been developed and tested to date (Queloz et al. (2009); Aigrain et al. (2012); Hatzes (2014); Haywood et al. (2014); Rajpaul et al. (2015); Faria et al. (2018); Hara et al. (2017); Delisle et al. (2020); de Beurs et al. (2021); Barragán et al. (2022); Klein et al. (2022) are some of them). We have gained extensive knowledge and comprehension regarding the 'stellar activity problem' thanks to all of these studies. Having said that, we are still unable to explore planetary signals in the sub-m s^{-1} RV regime. We started with hot Jupiters and can now handle warm super-Earths and Neptunes discovered by *TESS*, but we are limited to those planets until we can surpass the 1 m s^{-1} barrier. Therefore, our goal of finding an Earth twin around another Sun remains unfulfilled, even with a sufficient number of recently commissioned gold standard instruments (with sub-m s^{-1} level RV precision). The RV signature of stellar variability is stopping us from harnessing the full potential of the state-of-the-art instruments to definitively enter the realm of terrestrial planet discovery (Figure 1.8).

1.6 This Thesis

In this Thesis, I present a novel workflow (TWEAKS) that I have developed over the past three years by integrating stellar activity mitigation in the time (KIMA) and wavelength (SCALPELS) domains. Chapter 2 will introduce you to the workflow and discuss the tools and techniques involved, in detail. In the following chapters, I will demonstrate how this technique improves the detection and mass determination of planets, particularly smaller ones. TWEAKS provides a strong foundation to expand in the direction of retrieving signals beyond 1 m s^{-1} (Anna John et al., 2023), regardless of the degree of intrinsic variability of the host star.

In Chapter 3, I will present the first in-depth investigative analysis (based on Anna John

et al. (2022)) carried out to study the potential of SCALPELS to 'clean up' the RV of a stellar target other than the Sun. We chose to study CoRoT-7, to present the 'high-activity star' case in this chapter. You will see TWEAKS in action on less active stars in Chapter 4. While HD 144579 presents the 'moderate activity star' case, Hd 166620 will present the 'quiet-star' case, based on the findings from Anna John et al. (2023). Chapter 5 showcases the contribution of TWEAKS in the detection and characterisation of multiple planets around various *TESS*, K2 and HARPS-N targets. This chapter will also benchmark TWEAKS with existing state-of-the-art planet detection tools that correct for stellar variability.

I will contextualise the results of this research in the Conclusions chapter. Additionally, the limitations of the current method and the ongoing initiatives to address the problem will be covered there.

2

TWEAKS: A pipeline to detect Earths around other Suns

In this chapter, I present a novel workflow called TWEAKS (Anna John et al., 2022, 2023), designed to precisely detect the orbits and masses of low-mass exoplanets around stars with different levels of intrinsic activity. TWEAKS is the jolly acronym for the **T**ime and **W**avelength-domain st**E**llar **A**ctivity mitigation using **K**IMA and **S**CALPELS. I will introduce various techniques used in combination during the process of separating the spurious stellar activity signals from the RV measurements, to achieve this detection precision. To begin with, I will explain the recently developed wavelength domain signal separation algorithm called SCALPELS (Collier Cameron et al., 2021), which serves as the foundation for TWEAKS. Then, I'll discuss KIMA (Faria et al., 2018), the nested sampling algorithm that makes up the time domain side of the TWEAKS coin.

Finally, I present the Gaussian mixture modelling, data splitting, and injection-recovery tests used to verify or disprove the coherency of any interesting signals found in the process.

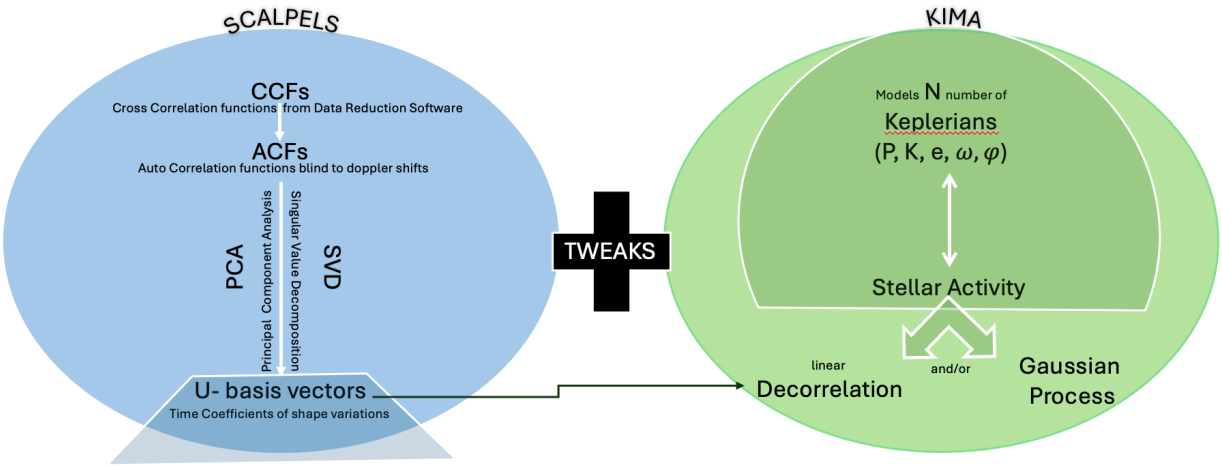


Figure 2.1: A simple schematic showing the TWEAKS architecture. TWEAKS models the observed RVs for the stellar activity component and planet component by integrating the wavelength domain signal separation tool SCALPELS (Collier Cameron et al., 2021), and time domain nested sampling planet searching tool KIMA (Faria et al., 2018).

I begin with a brief overview of how the TWEAKS pipeline functions. Subsequently, I will delve into the various concepts, algorithms, and procedures involved in detecting and measuring planet masses after mitigating the impact of stellar activity.

2.1 What is TWEAKS?

A succinct definition of TWEAKS would be as follows:

TWEAKS is an RV analysis pipeline designed for attaining sub-m s^{-1} RV detection thresholds (to detect low-mass planets) at long orbital periods, by combining wavelength-domain and time-domain stellar activity mitigation.

TWEAKS = SCALPELS+KIMA

This pipeline uses the SCALPELS technique, to separate the stellar variability component in the RVs, driven by spectral line-shape changes (Collier Cameron et al., 2021). The current version of SCALPELS works in the CCF domain. The temporal variability of the CCF contains both spectral line 'shape' changes (stellar activity component) and Doppler 'shifts' (planet component). SCALPELS constructs an orthogonal basis containing the coefficients of the first few principal components of shape-induced CCF variations, then project the raw RVs onto this

basis to obtain a time series of shape-induced RV variations. These are then scaled and fitted simultaneously with an orbit model to search for planet signals (Anna John et al., 2022, 2023).

The basis vectors (**U**-vectors) representing the shape components identified by SCALPELS are then used for stellar activity decorrelation using the KIMA nested sampling package (Faria et al., 2018).

Was that a complicated explanation for TWEAKS?

Ok, let's tackle the underlying techniques one at a time!

2.1.1 SCALPELS

Definition

Stellar magnetic activity produces time-varying distortions in photospheric line profiles in solar-type stars (See section 1.3). This changing profile shape produces systematic errors in high-precision RV measurements, limiting the efforts to detect low-mass exoplanets with orbital periods greater than a few tens of days and to measure their precise masses (Collier Cameron et al., 2021). To mitigate these measurement uncertainties, Collier Cameron et al. (2021) evaluated the changes in the spectral line shapes to influence the apparent RV. In contrast, changes in the first derivative of stellar distance (distance from the star's centre to the solar system barycentre, which varies due to the gravitational pull of the orbiting planet) with respect to time ($\frac{dl}{dt}$), induce only a shift in the profile without affecting the line shape or depth (See section 1.1.2).

i.e, stellar activity distorts the **shape** of spectral line profiles,

while , planet only induces a **shift** to the entire spectrum.

Collier Cameron et al. (2021) developed the SCALPELS algorithm as a wavelength-domain method for separating Doppler **shifts** due to planetary origin from spurious variations caused by line **shape** changes, arising from stellar activity and instrumental systematics.

Auto-Correlation Function (ACF)

The Cross-Correlation Function (CCF) represents the average shape of the absorption profiles across the whole spectrum. The temporal variability of the CCF contains both line shape changes and Doppler shifts. SCALPELS starts with a list of CCFs (a function of time and barycen-

tric velocity $C(v,t)$) for each observation epoch. The algorithm then computes the Auto-Correlation Function (ACF; Adler & Konheim (1962)) of the CCF. Hereafter, ACF refers to, ACF of the CCF. The ACF is effectively invariant to the stellar reflex motion with respect to the solar system barycentre. In other words, ACF is blind to **shifts**. SCALPELS uses this translational invariance property of the ACF to isolate the effects of shape changes in the CCF from shifts of dynamical origin.

ACF, $A(\delta v)$ can be described as the expectation value of the vector cross-product of the CCF with itself at a sequence of lags δv:

$$A(\delta v) = \mathrm{E}(\mathrm{CCF}(v) \cdot \mathrm{CCF}(v + \delta v)) \tag{2.1}$$

Collier Cameron et al. (2021) carried out a Principal-Component Analysis of the CCF and its ACF, using data from the HARPS-N solar telescope feed. They found that the majority of time variations of the CCF in general appear similar to those of the shift-invariant profile shape changes probed by the ACF.

The principal modes of variability in the ACF (or CCF) can be isolated by calculating the Singular-Value Decomposition (SVD) of the ensemble of ACFs or CCFs respectively and can be expressed as:

$$\mathbf{A}(\delta v_i, t_j) = \langle \mathbf{A}(\delta v_i) \rangle + \mathbf{U}_A(t_j) \cdot \mathrm{diag}(\mathbf{S}_\mathrm{A}) \cdot \mathbf{P}_\mathrm{A}(\delta v_i). \tag{2.2}$$

$$\mathbf{C}(v_i, t_j) = \langle \mathbf{C}(v_i) \rangle + \mathbf{U}_C(t_j) \cdot \mathrm{diag}(\mathbf{S}_C) \cdot \mathbf{P}_C(v_i). \tag{2.3}$$

The perturbations (line **shape** changes) resulting from stellar activity were then isolated by projecting the RVs onto the time-domain subspace spanned by the amplitude coefficients $\mathbf{U}_A$ of the ACF's principal components (Collier Cameron et al., 2021). On the contrary, dynamical **shifts** due to the planets are preserved when projected onto the orthogonal complement of the $\mathbf{U}_A$ time-domain subspace. We will see more about the projection process in the next subsection.

This is how SCALPELS works!

A schematic representation of the whole SCALPELS process is shown in Figure 2.2.

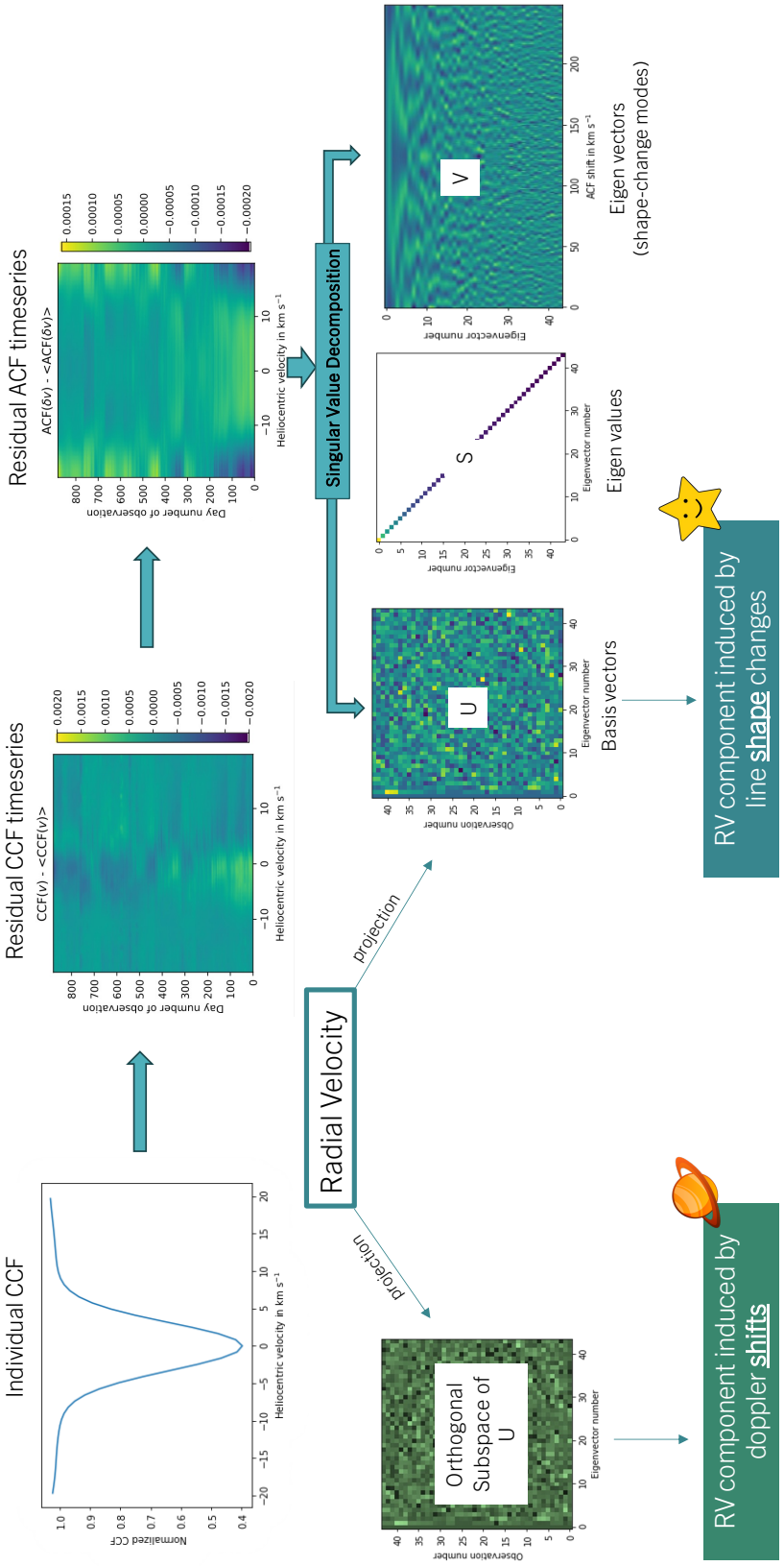


Figure 2.2: A schematic representation of SCALPELS separation as in Collier Cameron et al. (2021).

Signal separation via projection

The set of RV observations (obtained from the CCFs) have been treated as a vector $\mathbf{v}_{\rm obs}$. From which, the inverse-variance weighted mean $\langle v \rangle_{\rm obs}$ is subtracted and then this difference ($\mathbf{v}_{\rm obs} - \langle v \rangle_{\rm obs}$) is projected onto the $\mathbf{U}_A$- vector time-domain space (See 2.2).

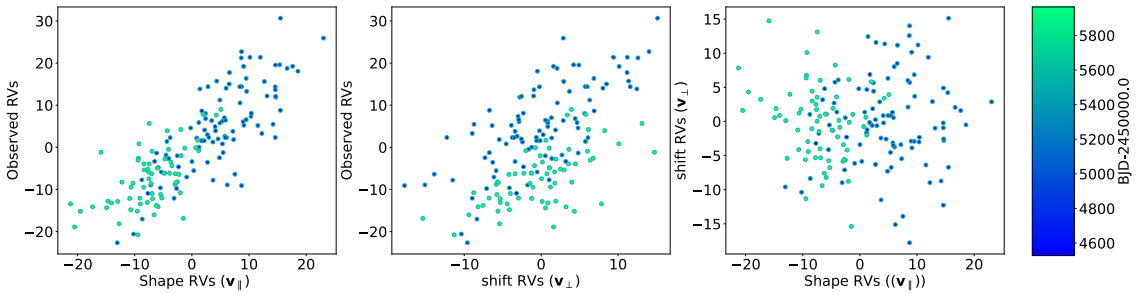


Figure 2.3: The correlation plots for the observed velocities (v_{obs}), shape-driven component ($v_\parallel$) and the shift component ($v_\perp$) for the HARPS RVs for CoRoT-7. Figure from Anna John et al. (2022).

The variation of the shape-driven component of the RV with time can be obtained from the sum of scaled velocity contributions from all principal components of ACF, $\mathbf{v}_\parallel = \mathbf{U}_A \cdot \hat{\alpha}$, where $\hat{\alpha}$ is the vector of response factors obtained by taking the inner product $\mathbf{U}_A^T$ with the mean-subtracted time series of observed velocities. The $\mathbf{v}_\parallel$ is found to provide a complete model of RV perturbations arising from the profile shape changes, to which the ACF is sensitive (Collier Cameron et al., 2021). The transformation from observed to shape-driven RVs can be considered as a linear projection of the observed velocities into a subspace spanned by the columns of $\mathbf{U}$:

$$\mathbf{v}_\parallel = \mathbf{U}_A \cdot \mathbf{U}_A^T \cdot (\mathbf{v}_{\rm obs} - \langle v \rangle_{\rm obs}). \tag{2.4}$$

These shape-driven perturbations are strongly correlated with the observed RVs, as seen from Figure 2.3, which could reliably reproduce the stellar activity-dominated long-term and short-term fluctuations (Collier Cameron et al., 2021). Even though this offers a linear decorrelation, the time coefficients of the SCALPELS basis functions exhibit time lags relative to the observed RV signals (See Figure 3.4), which needs to be addressed with caution. In addition to this, they can also be partially correlated with the orbit model, which is why in practice the amplitude coefficients of the orbit model and the decorrelation vectors must be computed simultaneously by weighted least squares, as in the CoRoT-7 analysis (refer to Section 2.1.5).

$\mathbf{v}_\perp$ is defined as the residual velocity which lies outside the $\mathbf{U}_A$ subspace, obtained from a projection into the orthogonal complement of the $\mathbf{U}_A$ subspace, where the planet signals are

sought:

$$\mathbf{v}_{\perp} = (\mathbf{v}_{\rm obs} - \langle v \rangle_{\rm obs}) - \mathbf{v}_{\parallel} = (\mathbf{I} - \mathbf{U}_A \cdot \mathbf{U}_A^T) \cdot (\mathbf{v}_{\rm obs} - \langle v \rangle_{\rm obs}) \tag{2.5}$$

To avoid overfitting, a reduced-rank version of **U** is employed. Collier Cameron et al. (2021) describe the determination of optimal rank using a leave-one-out cross-validation (LOOCV), by holding out each row $\mathbf{A}_j$ of the ACF in turn, and decompose the remaining rows to recompute the SVD. They reconstruct an estimate of $\mathbf{U}_j$ (of the missing j^{th} row of $\mathbf{U}_A$) by fitting the eigenvectors and eigenvalues to the j^{th} row of the ACF. After repeating this procedure for all rows, they rank the **U** based on the fidelity of reconstruction $\hat{\mathbf{U}}_k^{\mathbf{T}}$ to reproduce $\mathbf{U}^{\mathbf{T}}{}_{A,k}$, for k> 25 or so. Collier Cameron et al. (2021) found this method to be efficient in identifying the number of leading columns of **U** that contain significant profile information. The number of **U** vectors is then reduced further by reordering them in such a way that the Bayesian Information criterion (BIC=$\chi^2 + (k+1)\log n_{\rm obs}$) decreases to a minimum as quickly as possible. Refer to Section 3.3 of Collier Cameron et al. (2021) for more details about re-ordering and rank reduction.

2.1.2 KIMA

Definition

The basis vectors (**U**-vectors) representing the shape components identified by SCALPELS are then used for stellar activity decorrelation using the KIMA[1] nested sampling package (Faria et al., 2018), forming the backbone of TWEAKS. In Faria's words, "KIMA is a package for the detection and characterization of exoplanets using RV data. It fits a sum of Keplerian curves to a time series of RV measurements, using the Diffusive Nested Sampling algorithm (DNS; Brewer et al. (2010)) to sample from the posterior distribution of the model parameters. "

The evidence

KIMA employs Diffusive Nested Sampling (DNS Brewer (2014)) to sample the posterior distributions for each of the orbital parameters by modelling the RV data with a sum of up to N_p Keplerian functions. The joint posterior distribution is calculated using the prior $p(\Theta \mid M)$,

[1] Funfact: Joao P. Faria named this RV analysis package after the name of a (very) sweet drink produced in Azores, Portugal. I would call Kima, the Irn Bru of Portugal. I believe now you can feel the taste.

likelihood $p(D \mid \Theta, M)$ and evidence information $p(D \mid M)$ from the Bayes theorem:

$$\mathbf{p}(\Theta \mid D, M) = \frac{\mathbf{p}(\Theta \mid M)\ \mathbf{p}(D \mid \Theta, M)}{\mathbf{p}(D \mid M)} \tag{2.6}$$

where M is the model with Θ as the vector of all considered parameters and D is the RV data series. This gives KIMA the estimates for the Bayesian evidence ($\mathscr{Z} = p(D \mid M)$) of the data supporting each model, enabling the model comparison (Brewer et al., 2010). Essentially, evidence is the prior-weighted average of the likelihood. The value of the evidence increases for models with higher probability values across more of the parameter space and decreases for models with significant regions of low likelihood.

Bayesian model comparison involves comparing the posterior probability of the models, p(A | D) and p(B | D), given two distinct models A and B. Using the Bayes theorem and assuming equal prior probability for both models, KIMA computes Bayes factor (K), which is the ratio of the Bayesian evidence of the two competing models as,

$$\mathbf{ln\,K} = \mathbf{ln}\left[\frac{p(A \mid D)}{p(B \mid D)}\right] = \mathbf{ln}\left[\frac{p(D)_A}{p(D)_B}\frac{p(A)}{p(B)}\right] = \mathbf{ln}\left[\frac{p(D)_A}{p(D)_B}\right] \tag{2.7}$$

where $p(D)_A$ is the evidence of model A and $p(D)_B$ of model B. The ratio of posterior probabilities is therefore equal to the ratio of evidences when both models are equally likely apriori.

The ratio of Bayesian evidences between models with different N_p are compared similarly. Faria says, "Because KIMA uses the Diffusive Nested Sampling (DNS) algorithm, the evidence values are still accurate when the likelihood function contains phase changes which would make other algorithms (such as thermodynamic integration) unreliable."

The full model

Faria et al. (2018) developed KIMA as a transdimensional algorithm that calculates the fully marginalized likelihood, or evidence, for a model with a fixed number (N_p) of Keplerian signals or after marginalising over N_p. N_p is a free parameter for this latter task, and we sample from its posterior distribution using the trans-dimensional nested sampling algorithm of Brewer (2014).

For modelling, we use similar parameters as in Faria (2018); the number of planets N_p,

the orbital parameters for each of the N_p planets (P, K, e, ω, Φ), the extra white noise term s, and the systemic velocity v_{sys} of the star. In addition, we use the coefficients of leading SCALPELS **U**-vectors to model the stellar activity. These are the model parameters, whose joint distributions are sampled using the Markov Chain Monte Carlo (MCMC) algorithm. As a result, these parameters have prior and posterior distributions attached to them. The calculation of the Keplerian function is influenced by all these variables (apart from s).

Posterior analysis

Since N_p is an unknown parameter in most cases, the sampling algorithm needs to jump between candidate solutions with different numbers of planets. KIMA uses a birth-death MCMC within the nested sampling framework as proposed by Brewer (2014), which extends the traditional MCMC framework by allowing the dimensionality of the parameter space to change dynamically. This allows estimating the improvements in evidence values when the sampler switches between solutions with different N_p values. As an alternative, we can compare models with varying numbers of planets by fixing the values of N p in order and using the evidence ratio. But the trans-dimensionality of the DNS, allows the number of planets in the system N_p itself to be a free parameter in the analysis, and its posterior distribution to be estimated together with that of the orbital parameters. The same model comparison is then made possible by the posterior probability for each N_p value, which has the benefit of only requiring one algorithm run. Skilling (2004, 2006) proposed the Nested Sampling algorithm as a technique that directly calculates the evidence while also producing posterior samples as a by-product.

One of the main outcomes of this method is the posterior distribution for N_p. To decide on the number of detected planets, a probability ratio criterion for models is established with sequential numbers of planets (Figure 2.4), based on the scale for Bayes factors proposed by Jeffreys (1961). For N_p planets to be detected, the probability of N_p must be (at least) 150 times higher than the likelihood of (N_p - 1). See Table 2 in Standing et al. (2022) for more information. This criterion, which Faria (2018) refer to as the posterior ratio criterion, demands "decisive" evidence before a planet may be found, treating erroneous positives far worse than false negatives.

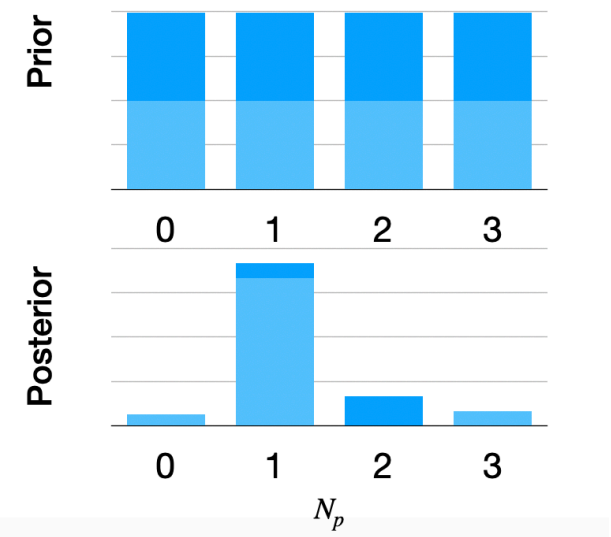


Figure 2.4: Schematic representation of the Uniform prior distribution for N_p, where it is considered as a free parameter with a maximum value of 3. A sample posterior distribution is also shown.

The 'known' planet

There is a 'known' object setting provided by KIMA, which allows one set of priors to be applied to the orbital parameters of a known planet while a different set of priors is used to explore the existence and parameters of putative planets (Standing et al., 2022). When a candidate Keplerian signal is proposed in the model, KIMA conducts fits for both the 'known' planet and the new signal. This option is useful if there is a transiting planet with precisely determined transit parameters or another 'known' Keplerian signal in the system that we can confidently fit out. Some undetectable planets might nevertheless exist around the star.

Decorrelating stellar activity

Linear decorrelation against any stellar activity indicators is another important feature KIMA provides. To do that, we must input an array of activity indicator/s as a function of time. The coefficient (or slope) of linear correlation with each indicator will be estimated together with the rest of the parameters of the nested sampling scheme. I used the basis vectors (**U**-vectors) representing the shape components identified by SCALPELS for stellar activity decorrelation. A Gaussian prior was used to model the amplitude coefficients of these decorrelation vectors.

2.1.3 The Gaussian Process

KIMA can also use a Gaussian Process (GP) with a quasi-periodic kernel as a noise model, to deal with activity-induced signals (Faria et al., 2018). If some of the rotationally modulated signals elude the SCALPELS signal separation, we activate this optional KIMA feature. The GP hyperparameters are inferred along with the orbital parameters. Note that all the parameters are estimated simultaneously. A variety of standard probability distributions are readily included in the package, making it simple for the user to establish priors for each of the parameters.

In the study of the very active star CoRoT-7 (See Section 3.2), we included a quasi-periodic GP, undertaking a similar approach to Haywood et al. (2014) and Faria et al. (2018), to account for the correlated noise. To investigate the improvement in the evidence, we modelled the planet orbits, SCALPELS basis vectors and the GP simultaneously, as opposed to a model without SCALPELS shape correction (Anna John et al., 2022).

The quasi-periodic component of the GP is defined by the kernel $k(t,t')$ (a.k.a. covariance function):

$$k(t,t') = {\eta_1}^2 \exp\left(-\frac{(\mathrm{t}-\mathrm{t}')^2}{2{\eta_2}^2} - \frac{2}{\eta_4^2}\sin^2\left(\frac{\pi(\mathrm{t}-\mathrm{t}')^2}{\eta_3}\right)\right) \tag{2.8}$$

The hyperparameters of $k(t,t')$ are represented by the variables η_1, η_2, η_3 and η_4. η_1 and η_2 define the amplitude of correlation between RVs at different time separations and the evolution timescale of active regions. The correlation timescale η_3 reflects the stellar rotation period. η_4 refers to the inverse harmonic complexity, which governs the variation timescale of GP latent functions in relation to η_3. Smaller values of η_4 indicate more short-scale structures within a single stellar rotation period. We used this GP to model the correlated noise occurring on time scales of the order of stellar rotation period and its harmonics.

2.1.4 False Inclusion Probability (FIP)

The next step is to evaluate any signal of interest based on its statistical significance after we have the posterior distribution from KIMA in hand. To do this, we applied the False Inclusion Probability (FIP) analysis formulated by Hara et al. (2022b). Essentially, FIP is the statistical measure of ***how often*** KIMA ***fails to find a signal***, in a given frequency bin. Please don't get

puzzled by my definition, you will understand what I meant after the next paragraph.

Hara et al. (2022b) presents this detection criterion as an efficient way to evaluate the reliability of significance levels, by effectively accounting for aliases and favouring to discard the presence of planets under a certain confidence level. We computed FIP in frequency space, with the bin size set to the Nyquist frequency resolution over the entire data duration. First, we calculated the True Inclusion Probability (TIP), as the fraction of all KIMA trial models (N_s) that contain planets whose periods fall in the frequency interval of interest. In other words, TIP is the number of models for which the orbital period (P) of a sampled planet falls within a given bin.

$$\mathbf{TIP} = \frac{\sum_{i=1}^{N_s} \mathrm{Pr_i(foreground)}}{N_s} \tag{2.9}$$

where Pr_i(foreground) is the probability that the point belongs to the foreground population, based on the Gaussian Mixture Modelling (See next section), and N_s the total number of trial models.

The FIP was then obtained as:

$$\mathbf{FIP} = 1 - \mathbf{TIP} \tag{2.10}$$

This allowed us to simultaneously search for multiple planets using a frequency window sliding across the full frequency range covered by the posterior distribution.

Let's put some numbers in now!

A FIP value of 0.83 indicates that only 17% of the models in the joint posterior samples contain a planet at this orbital period interval. The lesser the FIP, the more significant the signal is.

The process of significance determination of a given signal becomes challenging when signals exhibit fringing, caused by sampling patterns. Figure 2.5 shows the KIMA posteriors for CoRoT-7 as a case in point. The non-transiting planetary signals in CoRoT-7 appear to distribute power among the interference fringes due to the gap in observing seasons. Therefore, we employed Gaussian Mixture Modelling to identify all the posterior points associated with a specific signal's foreground population, without worrying about determining the ideal bin size each time.

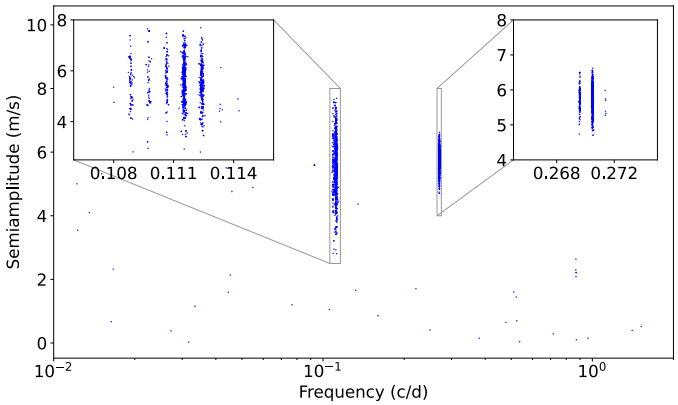


Figure 2.5: The semi-amplitude periodogram of CoRoT-7 HARPS data, displaying the joint posterior distributions obtained from KIMA, modelled with SCALPELS, GP and P_b as 'known'. The transiting 0.85 d planet is considered as 'known'. The inset plots zoom into the clusterings corresponding to the ∼9 d planet and ∼3.7 d planet (right), showing the interference patterns that occurred due to the widely separated observing seasons. Figure from Anna John et al. (2022).

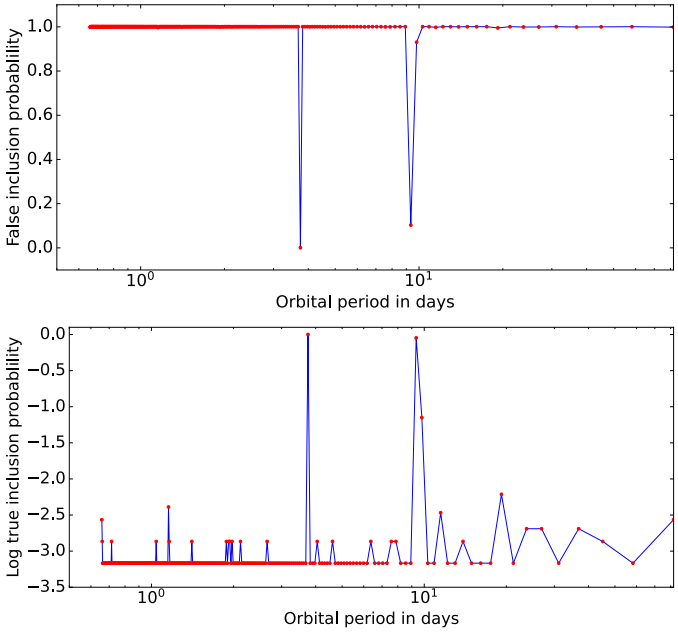


Figure 2.6: The FIP periodogram of CoRoT-7 HARPS data, computed using the joint posterior distributions obtained from KIMA, modelled withSCALPELS, GP and P_b as known. **Bottom**: TIP periodogram. Both periodograms strongly favour 2 planets at orbital periods 3.69 d and 8.96 d, in addition to the known transiting 0.85-day signal. Figure from Anna John et al. (2022).

Gaussian Mixture model

Gaussian Mixture Models (GMM) are used as an elegant clustering algorithm. GMM is a probabilistic model that assumes that every data point is generated by a combination of Gaussian distributions with unknown characteristics. In general, mixture models don't require knowing which sub-population a data point belongs to, allowing the model to explore the sub-populations automatically. Using GMM we precisely estimated the RV semi-amplitude for any

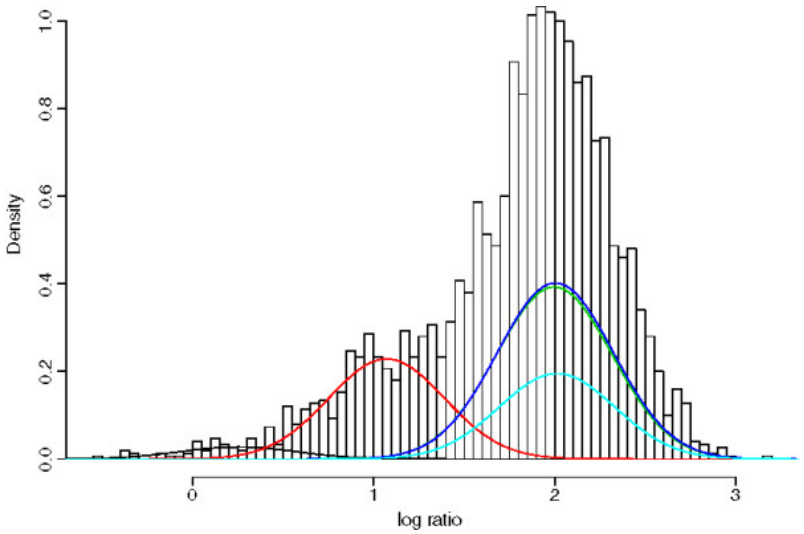


Figure 2.7: Figure from Chu et al. (2013) illustrating the application of Gaussian Mixture Modelling (GMM) to the investigation of genetic diseases. In this simulated mixture of Gaussian distributions, GMM distinguishes between a variety of discrete clusters (or distributions). The histogram is overlaid with the Gaussian mixture components, which are displayed in various colours. To precisely estimate orbital parameters, we use a similar approach to evaluate the likelihood that each point in the KIMA posterior distribution belongs to the foreground or background (prior) distribution.

signal of relevance found using TWEAKS, by isolating the cluster's foreground probability.

GMMs rely on the weighted average of the probability density functions, from various Gaussian distributions. To begin with, we determined the value of k, which represents the number of groups into which we want to divide the population. The value of k is two here (Foreground and Background). Then, we randomly initialise k centroids based on the number of clusters or groupings we desire. Following this, a cluster is created by assigning the data points to the nearest centroid. The data points are then reassigned and the centroids are updated. Up until the position of centroids is fixed, this procedure is iterative.

The probability density function of any Gaussian distribution in a one-dimensional space is given by:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu^2)}{2\sigma^2}} \tag{2.11}$$

where μ is the mean and σ^2 is the variance.

With two variables, we will have a 3D bell curve (Figure 2.8) rather than a 2D bell-shaped curve (Figure 2.7). The probability density function in two-dimensional space would then be given by:

$$f(x|\mu,\Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]} \tag{2.12}$$

where Σ is the 2x2 covariance matrix, μ is the 2D mean vector, and x is the input vector. The shape of the curve would now be determined by the covariance. For a dataset with d features, we would have a mixture of k Gaussian distributions, each with a distinct mean vector μ, and variance matrix Σ (where k is equal to the number of clusters). But hold on, how are the mean and variance values for every Gaussian determined?

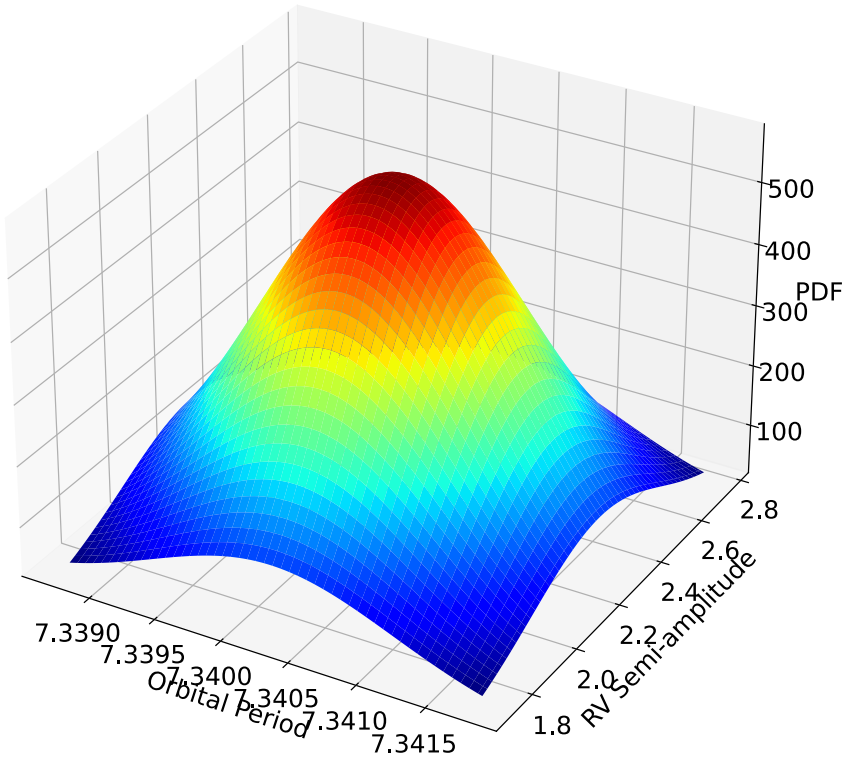


Figure 2.8: The probability density function in two-dimensional space for a candidate planet signal in HD 48948 system found using TWEAKS.

Expectation-Maximization (EM), a statistical approach for choosing the appropriate model parameters, is used to determine these values. The algorithm, in general, consists of two steps: the E-step and the M-step.

Suppose we have k clusters to assign. The mean and covariance values are μ_1, μ_2,.. μ_k and Σ_1, Σ_2,.. Σ_k, respectively, for k Gaussian distributions. The distribution's second parameter, which determines how many points are in it, also has another value.

To define the Gaussian distributions, we must now obtain values for all these parameters. We already chose how many clusters there would be, and we chose at random the mean, covariance, and density values. The E-step and M-step will follow soon after.

E-step

The likelihood that each point x_i is a part of the foreground distributions f_1, f_2,... f_k can be calculated as:

$$P_{fi} = \frac{\text{Probability } x_i \text{ belongs to foreground}, f}{\text{Sum of probability } x_i \text{ belongs to } f_1, f_2, ...f_k} \tag{2.13}$$

This value will be high if the correct cluster is allocated to the point, and low otherwise.

M-step

After the E-step, we return and update the Π, μ and Σ values as follows:

$$\Pi = \frac{\text{Number of points assigned to the foreground}}{\text{Total number of points}} \tag{2.14}$$

$$\mu = \frac{1}{\text{Number of points assigned to the foreground}} \Sigma_i P_{fi} x_i \tag{2.15}$$

$$\Sigma = \frac{1}{\text{Number of points assigned to the foreground}} \Sigma_i P_{fi} (x_i - \mu_c)^T (x_i - \mu_c) \tag{2.16}$$

We calculate the new probability for each data point using the updated values produced by this step, and we update the values iteratively. The process is repeated until the log-likelihood function is maximised.

We used a Gaussian mixture model (Feroz et al., 2019a), to compute the probabilities of posterior samples belonging to Gaussian foreground and background populations in the posterior parameter space, to exclude the posterior points coming from the uniform prior 'fog' (background). This way, we isolated the foreground probability of the cluster and obtained a reliable estimation for the RV semi-amplitude for any signal of interest. The application of GMM to determine the orbital parameters of planet candidates with improved uncertainties is demonstrated in Chapters 3 & 4.

2.1.5 Other techniques used in TWEAKS

l_1-periodogram

It is challenging to identify and disentangle the frequencies of signals arising from multiple harmonics. Despite being statistically significant, chances are not negligible that the tallest peak in the Generalised Lomb-Scargle periodogram(GLS) could be a spurious one due to the effects of aliasing. To address this problem, Hara et al. (2017) developed the ℓ_1-periodogram to search for periodicities in unevenly sampled RV time series, which could result in a periodogram similar to Lomb-Scargle periodogram but with fewer peaks due to aliasing. We employ the ℓ_1-periodogram in TWEAKS as a tool to direct the simultaneous sinusoidal planet orbit fitting discussed in the next subsection.

The ℓ_1- periodogram (Hara et al., 2017) is primarily designed to identify the sparsest set of orbital signals that fit the observed RVs. This sparsity is achieved by simultaneously fitting Keplerian signals on an optimally sampled grid of frequencies to the data, regularized by minimizing the ℓ_1 norm[2] of their amplitudes. Even though both ℓ_1-periodogram and Lomb-Scargle periodogram are defined on a frequency grid, the contrast arises as the former offers the representation of input signal as a sum of a number of sinusoidal components and therefore performs a simultaneous search for signals of all frequencies, where the latter fits a sinusoidal function at each frequency grid iteratively. Hara et al. (2017) present the algorithm as a tool to search for a representation of the input signal as a sum of a few sinusoidal components, that is, a representation which is sparse in the frequency domain.

As the primary step, the RV data is normalized to the mean of the data to get an estimate of stellar noise. The selection of frequency grids ω, covariance matrix $\mathbf{V}$ and the width of averaging interval η are done in succession. This algorithm suggests that RV measurements can be well approximated by the linear combination of a few vectors of $e^{-i\omega t}$ and $e^{i\omega t}$, where t is the array of measurement times (Hara et al., 2017). The noise is assumed to be drawn from a Gaussian with the covariance matrix $\mathbf{V}$. If the stellar rotation information $P_{\rm rot}$ is given, the covariance matrix uses a quasi-periodic kernel. The contribution of activity to the covariance matrix is then modelled as the product of an exponential correlation term and a quasi-periodic term.

[2] ℓ_1 norm is the sum of the magnitudes of the vectors in a space, a good measure of distance between them.

$$V_{ij}^{(\mathrm{act})} = \exp\left(-\frac{(t_i - t_j)^2}{2\tau^2}\right)\left(0.5 + \cos\left(\pi\frac{(t_i - t_j)}{P_{\mathrm{rot}}}\right)\right) \tag{2.17}$$

The full covariance matrix including the noise terms can be then expressed as:

$$\mathbf{V} = \delta_{\mathrm{k}} + \sigma_R^2 V_{ij}^{(\mathrm{act})} + \mathbf{I}\sigma_W^2 + \sigma_C^2 \tag{2.18}$$

where $\delta_{\mathbf{k}}$ is a diagonal matrix spanned by the RV uncertainties and σ_W is the white noise amplitude, which we considered as 0.1 m s^{-1} . σ_R stands for the red noise, which is 0.5 m s^{-1} here multiplied with an identity matrix **I**. In this simple covariance model, a block-diagonal calibration-noise covariance component σ_C is also included to account for the night-to-night calibration errors, which are known to be present in HARPS data at the $\sim$0.5 m s^{-1} level.

A remarkable advantage of the method is the use of ℓ_1 norm weighting to find periodicity in unevenly sampled signals. Even without any prior knowledge of the number of planets, the intrinsic signal information can be efficiently decoded. The algorithm assigns an amplitude to every frequency on the grid. It then tensions the goodness of fit to the data (using χ^2, which is the "ℓ_2 norm") against the ℓ_1 norm (median absolute deviation) of all the amplitudes on the frequency grid. The goal is to identify the minimal set of sinusoids that give an optimal fit to the data. The optimal solution is then found iteratively.

Moreover, we also used its capability to fit a set of externally-determined basis functions simultaneously with the orbit fit. If the columns of the reduced-rank matrix $\mathbf{U}_A$ are used for this purpose, this is equivalent to the simultaneous sinusoidal orbit fitting and SCALPELS decorrelation method described in the next subsection. This inter-operability with the SCALPELS enables us to perform shape-shift signal separation simultaneously with identifying the minimal set of orbit signals present.

Simultaneous sinusoidal planet orbit fitting

We employed the shape-signal separation simultaneously with fitting the orbits, given prior knowledge of the orbital periods obtained from the ℓ_1-periodogram analysis. A simultaneous modelling was chosen to be the best way to figure out the orbital solution, as a simple periodogram has the limitation of fitting only a single sinusoid per frequency sample, so the interaction between various signals may lead to inaccurate amplitude estimation. To determine the impact of the SCALPELS signal separation, Collier Cameron et al. (2021) used this

approach in the solar RV data, injecting a few planet-signals that should be orthogonal to all elements of $\mathbf{U}_A$. This prevented the signals from being partially absorbed during the SCALPELS projection.

Estimation of parameters and signal separation can be achieved in a single linear calculation, once the candidate signal periods have been determined via periodogram search or through prior knowledge of transits. In the case of CoRoT-7 (See Section 2.1.5), we have an estimate of the periods of the three candidate signals from the SCALPELS blind periodogram search and ℓ_1-periodogram, while the period and phase of the CoRoT-7b signal were obtained more precisely from transit observations.

To perform simultaneous orbit modelling, we followed the same method described by Collier Cameron et al. (2021). The net orbital velocity vector $\mathbf{v}_{\rm orb}$ for a set of n planet signals is modelled as the product of a set of coefficient pairs $\theta_{\rm orb} = \{A_1, B_1, \cdots, A_n, B_n\}$ with an array of time-domain function pairs $\mathbf{F} = \{\cos\omega_1 t_j, \sin\omega_1 t_j, \cdots, \cos\omega_n t_j, \sin\omega_n t_j\}$, where ω_k is the orbital frequency of the k^{th} planet:

$$\mathbf{v}_{\rm orb} = \mathbf{F} \cdot \theta_{\rm orb} \tag{2.19}$$

If we consider the planet orbits to be circular, $\mathbf{v}_{\rm orb}$ from Equation 2.19 could then provide a complete model of the RV data as the sum of shift (planet)-driven velocity variations and the shape (stellar activity)-driven velocity variations (Collier Cameron et al., 2021). Treating this as a least-square problem enabled us to solve for the unknowns, the amplitudes, and phases of $\mathbf{F}$. The vector $\theta_{\rm orb}$ can then be evaluated by minimizing the χ^2.

The detailed algorithm is given in Appendix A.

Stacked Bayesian General Lomb-Scargle Periodograms

We use the Stacked Bayesian General Lomb-Scargle (BGLS) periodogram method to probe the statistical implication of the planet candidate signals identified using TWEAKS. Stacked BGLS was developed for identifying the periodicities caused by stellar activity, by tracking the signal-to-noise ratio (SNR) of the detection over time (Mortier & Collier Cameron, 2017). Adding in more observations should increase the power or SNR measure of the signal if the signal is real and coherent. On the contrary, if the signal is incoherent or short-lived, adding more data would cause the SNR or Lomb-Scargle power to decrease (e.g., Howard et al., 2011; Hatzes, 2013; Mortier & Collier Cameron, 2017). Hence, signals arising from stellar

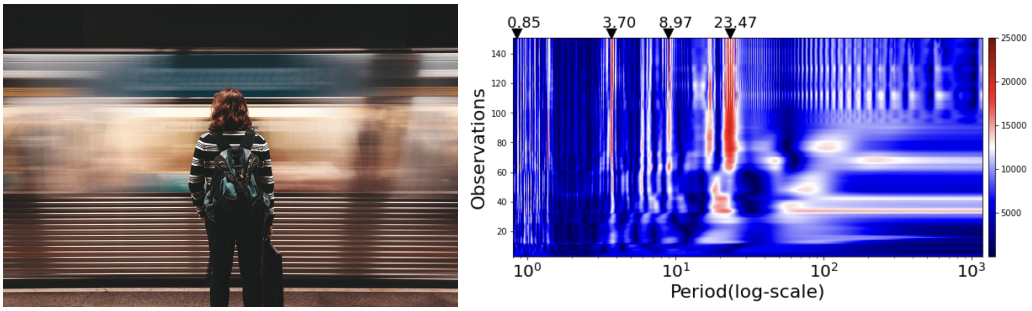


Figure 2.9: Left: This long exposure image demonstrates how stacking many short exposure images blurs a moving (incoherent) object while keeping the background object (coherent) sharp (Picture credit: Google). The Stacked Bayesian General Lomb-Scargle (BGLS) periodogram applies a similar concept to separate the stellar variability (incoherent) signals from the planetary (coherent) signals in the RV data, by stacking periodograms generated in each step of adding in more observations. **Right:** The Stacked BGLS periodogram for CoRoT-7 with the number of observations in the y-axis plotted against the period in d displayed on a logarithmic scale. The colour gradient designates the logarithm of probability, where the redder is more promising. The orbital periods of planet signals (0.85 d, 3.70 d, and 8.97 d) show a gradual increase in power on the addition of subsequent data points, whereas the stellar activity signal at the rotation period (23.47 d) shows a decrease in power after ∼100 observations. Figure from Anna John et al. (2022).

activity can be identified by monitoring their variable and incoherent nature. To track the significance of periodic signals, the BGLS periodogram was calculated for a smaller data subset and recalculated repeatedly with subsequent addition of observations.

Mortier & Collier Cameron (2017) recommend that one should not solely rely on this approach to establish the planetary nature of a signal. This can be used as an additional test to reinforce a finding. Therefore, we included SBGLS in our workflow as a confirmation tool to validate the coherent (or planetary) nature of any signal of interest we may detect during the process. Figure 2.9 shows the SBGLS periodogram for CoRoT-7, re-inforcing the coherency of candidate planet signals found using the TWEAKS pipeline.

2.2 Summary

We employed TWEAKS as an RV analysis pipeline to search for planetary signals below the 1 m s^{-1} RV barrier imposed by stellar activity (Figure 1.8). This pipeline uses the SCALPELS technique to separate the stellar variability component in the RVs, driven by spectral line-shape changes (Collier Cameron et al., 2021). SCALPELS constructs an orthogonal basis containing the coefficients of the first few leading principal components of shape-induced CCF variations, then project the raw RVs onto this basis to obtain a time series of shape-induced RV variations. The coefficients of these basis vectors are then used to model the stellar activity component while solving simultaneously with the orbit parameters (Anna John et al., 2022, 2023).

We used the ℓ_1 periodogram, a compressed-sensing method developed by Hara et al. (2017), to identify the sparsest set of orbital periods that could simultaneously fit the SCALPELS shift signal. The resulting sparse solution will eliminate unwanted alias signals and reveal candidate planet signals. The signals favoured by ℓ_1 periodogram are then used to guide the simultaneous planet orbit and stellar activity modelling to study the impact of SCALPELS separation.

The basis vectors (**U**-vectors) representing the shape components identified by SCALPELS are then used for stellar activity decorrelation using the KIMA nested sampling package (Faria et al., 2018). KIMA employs Diffusive Nested Sampling (DNS; Brewer (2014)) to sample the posterior distributions for each of the orbital parameters by modelling the RV data with a sum of up to N_p Keplerian functions. The range of posteriors obtained from KIMA is then used for model comparison, Evidence analysis and detection limit studies described in Chapters 3 and 4. The precise orbital parameters for any candidate planet signals that passed the coherency tests (SBGLS periodogram and/or data splitting tests) are then determined by applying Gaussian Mixture Modelling on the posteriors.

2.3 Some useful references

Here is a collection of resources I've gathered over the past few months with aid from others which I hope you will also find helpful.

- The following sources discuss the variety of techniques used in the TWEAKS pipeline:

  Collier Cameron et al. (2021) introduces SCALPELS and its application to the HARPS-N solar data. You can access the basic code here.

  Faria et al. (2018) details KIMA nested sampling package, the code available at https://github.com/j-faria/kima.

  Mortier & Collier Cameron (2017) presents the Stacked BGLS technique, the algorithm can be accessed here.

  Hara et al. (2017) discusses the ℓ_1 periodogram.

  If you are keen on learning the basics of Gaussian Mixture modelling, pay a visit to this webpage.

3

CoRoT-7: The very active star

In this Chapter[1], I present the work I have done towards characterising the planetary system CoRoT-7. CoRoT-7 is an active star host to a small hot Neptune and the first Earth-size exoplanet ever discovered (Léger et al., 2009). The system was first discovered via the transit method, early in the CoRoT space mission. It was then followed up with radial velocity observations using SOPHIE, HARPS, and HARPS-TERRA to determine the precise mass and to look for more planets in the system.

Using TWEAKS I confirmed the existence of a third planet in the system, whose existence was previously in doubt. In addition, I also obtained refined mass measurements for all planets in the system and characterised the transiting planet as a rocky Super-Earth with 10 σ precision.

[1]This Chapter uses material from and is based on, Anna John et al., 2022, MNRAS, Volume 515, Issue 3, pp.3975-3995

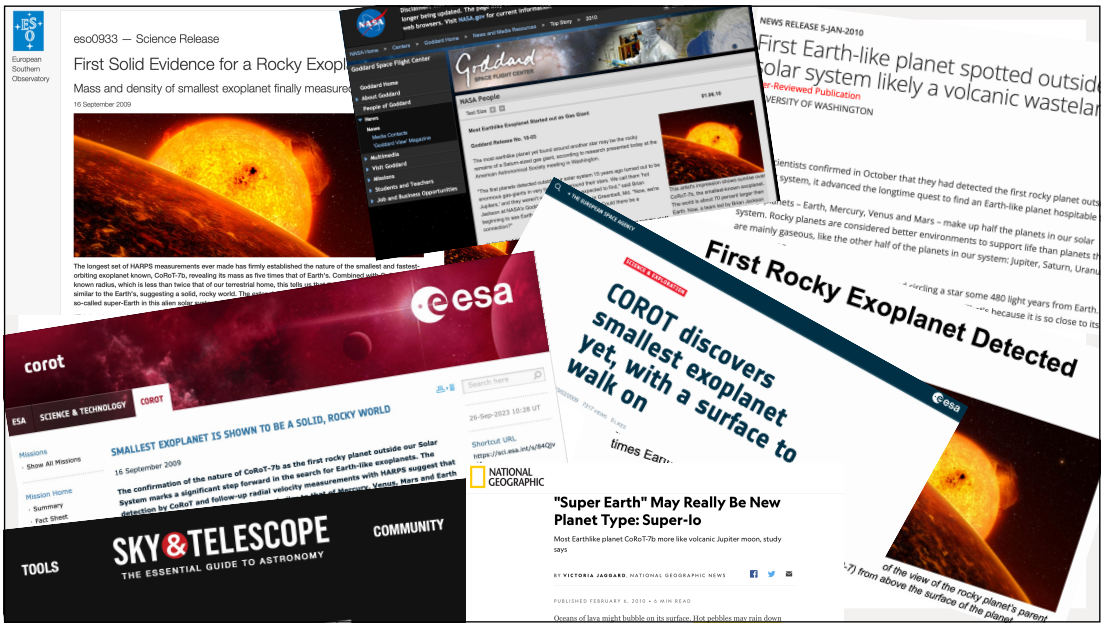


Figure 3.1: A compilation of online press releases regarding the discovery of CoRoT-7b, the first transiting rocky Super Earth! This illustration was inspired from a similar one in Haywood (2016).

3.1 Why CoRoT-7?

Following the discovery of the super-Earth CoRoT-7b, numerous investigations have generated different inferences for the number and masses of planets present in the CoRoT-7 system, owing primarily to the star's high level of intrinsic activity. This system has an extensive history, which I present in Section 3.1.1 before reporting on my analysis in the subsequent Sections.

Within Section 3.2 we describe the outcomes of TWEAKS analysis for CoRoT-7 as different subsections. We also examine the evidence for a third planet in the system and the impact of the number of non-transiting planets on the planetary mass of transiting planet CoRoT-7b extensively, followed by a discussion and conclusion in Section 3.2.6.

3.1.1 CoRoT-7 planetary system in literature

CoRoT-7 has been in the limelight since the discovery of the planetary system orbiting this bright star, 157.57 parsecs (Gaia Collaboration et al., 2023) away. CoRoT-7 is a very active G9V type main-sequence star of V magnitude = 11.7, slightly cooler ($T_{\rm eff} = 5101.2^{+186.4}_{-44.4}$ K, (Gaia Collaboration et al., 2021)) than our Sun. The stellar rotation period was measured to be 23 days, from CoRoT photometry. Using gyrochronology (Barnes, 2007), which calculates a star's age as a function of its rotation period and colour, (Léger et al., 2009) estimated the age of

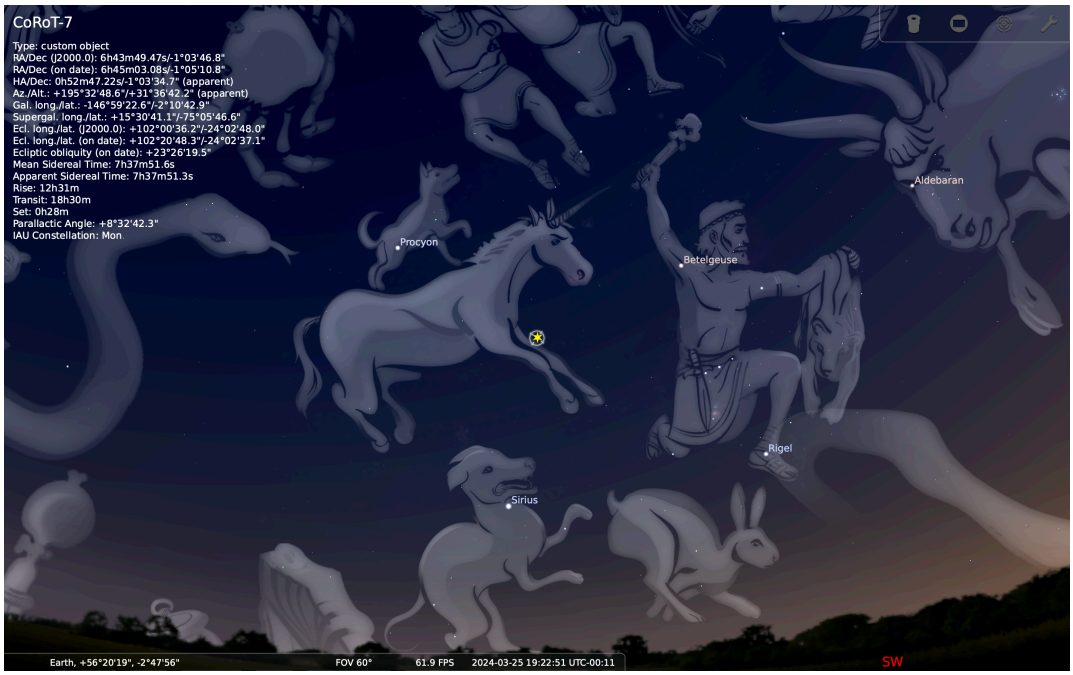


Figure 3.2: A Stellarium guide showing CoRoT-7 (marked by the yellow star in crosshair) in the Monoceros constellation on a March sky, from Scotland. Please note: the size of the star is not to scale & the star is too faint to be seen with the naked eye from Earth.
Fun fact: Monoceros is the Latin word for Unicorn, the one-horned horse, which occurs to be the national animal of Scotland.

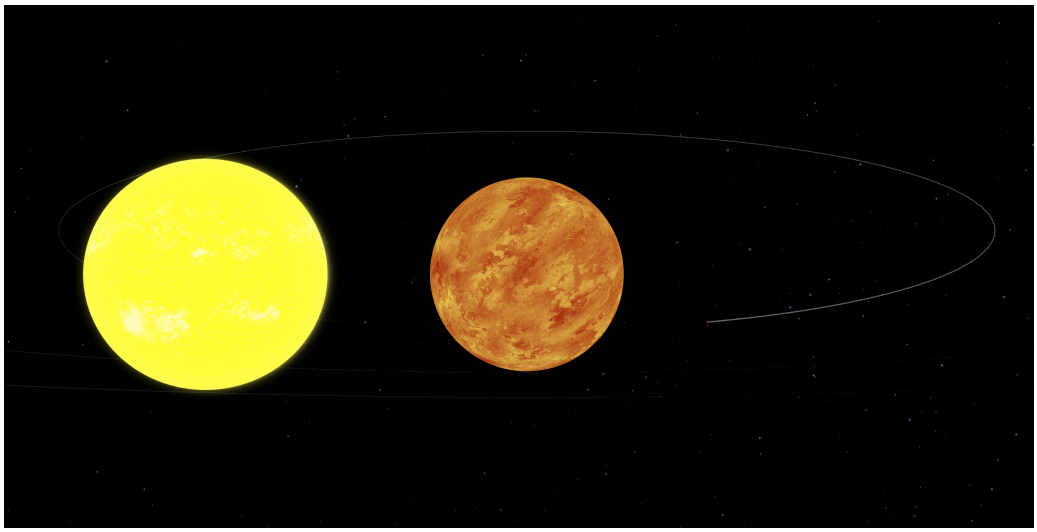


Figure 3.3: Size of CoRoT-7 ($0.93R_{\odot}$) portrayed in comparison to Sun (left). Credits: NASA

CoRoT-7 to be 1.7 ± 0.3 Gyr. Léger et al. (2009) reported the detection of the first ever known transiting super-Earth, CoRoT-7b with an orbital period of 0.85 d and a measured radius of 1.68

± 0.09 $R_\oplus$. Following this discovery, a 4-month intensive HARPS RV follow-up campaign was launched to measure the mass of CoRoT-7b. These observations were carried out under ESO Program IDs 082.C-0120, 082.C-0308(A) and 282.C-5036(A) using the HARPS instrument on the ESO 3.6 m telescope at La-Silla. CoRoT-7 was observed for 3 nights in 2008 March and then observed continuously from 2008 November to February 2009, spanning over 4 months. The findings of this run are reported in Queloz et al. (2009). Given the huge modulations in the CoRoT photometry, they anticipated that the RV fluctuations would be heavily influenced by stellar activity. The star's lightcurve (2008-2009 CoRoT run) reveals modulations owing to starspots of up to 2%, indicating that CoRoT-7 is more active than the Sun (The Sun's highest reported fluctuation in irradiance is 0.34%, Kopp & Lean 2011). Using the measurements from this RV campaign, Queloz et al. (2009) revealed CoRoT-7c, the second planet in the system, with an orbital period of 3.69 d. This made CoRoT-7 the first star discovered with a planetary system made of two short-period super-Earth mass planets with one in transit (Queloz et al., 2009).

This entire chapter describes our analyses aiming to provide a better understanding of the planetary system hosted by CoRoT-7. The following two subsections explore the two major conundrums that have been heating up scientific debates for more than a decade.

1) Is there a 3rd planet, CoRoT-7 d?

2008-9 RV observations

After the discovery of the first two inner planets (transiting CoRoT-7 b & non-transiting CoRoT-7c), the presence of a potential third planet, CoRoT-7d, with an orbital period of 9.02 d and a mass of 16.7 ± 0.42$M_\oplus$was reported by Hatzes et al. (2010), using the 2008-2009 HARPS RV data. Hatzes et al. (2010) applied a pre-whitening procedure to the BIS, FWHM and CaII H & K emission lines (spectral quantities used as measures of the host star's activity) owing to their in-dependency to the planetary dynamics. As no significant signal came up with a period of 0.85 and 3.69 days in any of these stellar activity indicators, the planetary candidacies of CoRoT-7b & CoRoT-7c were strengthened. However, the calculated planetary masses were not in good agreement with those from the previous studies. This discrepancy in the mass calculation could possibly be rooted in the difference in the methods adopted to cancel out the effects of stellar activity signals from the RV data (See Table 3.1 for the masses).

Furthermore, no signal was found in the periodogram of activity indicators at 9.02 days, and thus this RV signal was attributed to a third-planetary candidate in the system, CoRoT-7d. If the third companion could be confirmed, then CoRoT-7 may represent a case of an ultra-compact planetary system (Hatzes et al., 2010). The 9.02 d signal was previously detected by Queloz et al. (2009) as well, but had been attributed to a 'two frequency beating mode' arising from an amplitude modulation of a 61-day stellar activity signal, found during the pre-whitening procedure. It was therefore deemed to be associated with stellar activity. Subsequently, due to the presence of strong stellar and instrumental variations with unknown spectral characteristics, Pont et al. (2011) claimed that the 2008-2009 data cannot be used to search for additional (non-transiting) planets in the 3–10 d period range and that claims of the detection of such planets (CoRoT-7c and CoRoT-7d) do not withstand scrutiny. This suggestion inspired further studies to push for new CoRoT-7 observations.

More RV observations in 2012

Aiming to resolve the disparities, simultaneous photometric and spectroscopic observations were obtained in 2012 from CoRoT and HARPS (ESO Program ID 088.C-0323) for 26 consecutive clear nights in a row from 2012 January 12 to February 6, with multiple well-separated measurements on each night adding up to 77 observations in total (Haywood et al., 2014). These RV data were reprocessed in the same way as the 2008-2009 data (Queloz et al., 2009), using the HARPS data processing pipeline. Altogether, the CoRoT-7 system was observed for RV studies using HARPS in 2008, 2009 and 2012 with a total of 177 measurements.

The purpose of the 2012 observations was to observe the system simultaneously with CoRoT and HARPS and to use photometry as a proxy for activity-driven RV fluctuations. Along with updating the planetary parameters for CoRoT-7b, Haywood et al. (2014) analysed this system further for the evidence of the third planetary companion with the period of 9.02 days, reported by Hatzes et al. (2010). Although the Lomb-Scargle periodogram showed a large peak in the 6-10 day range, the marginal likelihoods computed from their GP regression analysis supported the two-planet solution over a three-planet solution that included the 9-day signal. Hence, they concluded the 9.02-day signal to be more likely associated with the second harmonic of the stellar rotation at $\sim$7.9 d. However, they successfully confirmed CoRoT-7c along with CoRoT-7b and improved their planetary parameters.

Later, Faria et al. (2016) carried out a model-comparison study employing GP regression

to model the activity alongside simultaneous orbit fitting in a nested-sampling scheme using the full CoRoT-7 RV data set. While they found evidence for a weak signal at 9 days, the odds' ratio again favoured the 2-planet solution. Tuomi et al. (2014) also presented an analysis of the HARPS-TERRA (Template-Enhanced RV Re-analysis Application) velocities of CoRoT-7 that suggested a system of two planetary companions (CoRoT-7b & CoRoT-7c), possibly three (CoRoT-7d).

2) Mass of the transiting Super-Earth, CoRoT-7b

From the plethora of literature, CoRoT-7 has been known to have two planets, one being the first to be found with a density similar to that of Earth, CoRoT-7b. CoRoT-7b had the smallest exoplanetary radius known at the time of discovery. However, the mass estimates for this planet reported in the literature did not converge (Table 3.1). This is plausibly due to the activity level of the star that contributes a significant amount of RV "jitter" and how the various methods correct for this ambiguity. It also depends on how many non-transiting Keplerian signals are included in the model. The method of removal of the activity signal will also affect the RV amplitude, which in turn can bias the mass determination of companions. Most of the mass determinations listed in Table 3.1 used the same HARPS RV data set, which highlights the trouble in planets' mass determination around an active star like CoRoT-7 (Hatzes et al., 2011).

Collectively, these studies outline the critical role of stellar activity and non-transiting signals in contaminating the RV observations not only in this particular system but also in general. Therefore, it is crucial to understand this role to enhance our ability to detect low-mass planets and thereby measure their masses precisely.

The goal of this chapter is to understand and mitigate the effects of stellar activity in the RV measurements of CoRoT-7 using TWEAKS and;

- refine the mass and characterise the transiting planet CoRoT-7b
- determine whether CoRoT-7 d is a planet or a stellar activity artefact

We used the very same set of HARPS RV measurements, that were used to obtain the wide range of masses ($2.3M_{\oplus}$to $8M_{\oplus}$) given in Table 3.1.

Table 3.1: A timeline of the wide range of reported planet masses, especially for CoRoT-7b. It is evident that the mass of CoRoT-7c is nearly consistent. The measured orbital periods for all (proposed and confirmed) planets are also included.

	Planetary masses ($M_\oplus$)			Orbital periods (in days)		
Authors	CoRoT-7 b	CoRoT-7 c	CoRoT-7 d	CoRoT-7 b	CoRoT-7 c	CoRoT-7 d
Léger et al. (2009)[a]	< 21	–	–	0.85 d	–	–
Queloz et al. (2009)[b]	4.8 ± 0.8	8.4 ± 0.9	–	0.85 d	3.69 d	9.02 d (deemed as activity)
Hatzes et al. (2010)[b]	6.9 ± 1.4	12.4 ± 0.42	16.7 ± 0.42	0.85 d	3.69 d	9.02 d
Ferraz et al. (2011)[b]	8.0 ± 1.2	13.6 ± 1.4	–	0.85 d	3.69 d	–
Boisse et al. (2011)[b]	5.7 ± 2.5	13.2 ± 4.1	–	0.85 d	3.69 d	–
Pont et al. (2011)[b]	2.3 ± 1.8	–	–	0.85 d	–	–
Hatzes et al. (2011)[b]	7.4 ± 1.2	–	–	0.85 d	3.69 d	9.02 d
Haywood et al. (2014)[c]	4.7 ± 0.9	13.5 ± 1.08	–	0.85 d	3.69 d	8.58 d (deemed as activity)
Tuomi et al. (2014)[d]	4.8 ± 2.3	11.8 ± 4.1	15.4 ± 6.1	0.85 d	3.71 d	8.89 d (unsure about origin)
Barros et al. (2014)[e]	5.5 ± 0.8	–	–	0.85 d	–	–
Faria et al. (2016)[c]	5.5 ± 0.8	12.6 ± 0.7h	–	0.85 d	3.69 d	8.58 d (deemed as activity)

Data used : [a] CoRoT photometric LRa01& SOPHIE RV data, [b] HARPS RV 2008-9, [c] HARPS RV 2012, [d] HARPS-TERRA RV 2008-9, [e] CoRoT photometric LRa06

3.2 Results

In this section, I discuss results from the first in-depth investigative analysis to study the potential of TWEAKS to "clean up" the RVs of a stellar target and thus improve the fidelity of RV measurements. We applied the TWEAKS pipeline (Chapter 2) to the archival HARPS data on the CoRoT-7 system, which has been studied extensively with time-domain methods, but for which the precise system architecture remains a subject of debate.

3.2.1 Spectroscopy Time-series

We used the CCFs obtained from the HARPS Data Reduction Software (DRS) as the input for the TWEAKS pipeline. The cross-correlation procedure was performed using a K5 spectral mask. In addition to measuring the RV, the DRS returns several shape diagnostics for the CCF and flux indices for the known activity-sensitive spectral lines. The RV data and supplementary information including the S-index, BIS, Area and FWHM from 2008-9 and 2012 observing campaigns are shown in the top and bottom panels of Figure 3.4. The quantity 'Area' is the product of the FWHM of the CCF and its contrast, i.e., the fractional depth of the CCF measured relative to the continuum level. The y-axis values for each parameter in both observing seasons are set to have the same limits for easy comparison.

The supporting measurements of the 'activity indicators' such as the FWHM and BIS of CCF and the Ca II H & K values from HARPS allow us to examine the intrinsic stellar variability. Queloz et al. (2009) suggest that one can check two simple characteristics to look for changes in the shape of CCF: the width parameter FWHM and its BIS, computed following Queloz et al. (2001). The chromospheric activity is typically quantified by the S-index (Wilson, 1978). The S-index varies linearly with the chromospheric emission flux in the cores of the Ca II H & K lines in the near ultraviolet (Salabert et al., 2016). From both panels of Figure 3.4, the mean value for the S-index can be calculated as 0.3569 in the 2008-9 season and 0.2759 in 2012 and is denoted by the grey horizontal dashed line. The lower mean value in 2012 indicates that the star was less active during 2012.

Similarly, the RV measurements possessed an RMS scatter of 10 m s^{-1} in the 2008-9 run and a lesser scatter value of about 6.8 m s^{-1} in 2012, showing that the amplitudes of RV variability were lower in 2012 than 2008-9. The mean levels of RV amplitudes were comparable (31.193 and 31.182) during the two campaigns. However, both sets exhibit obvious multi-

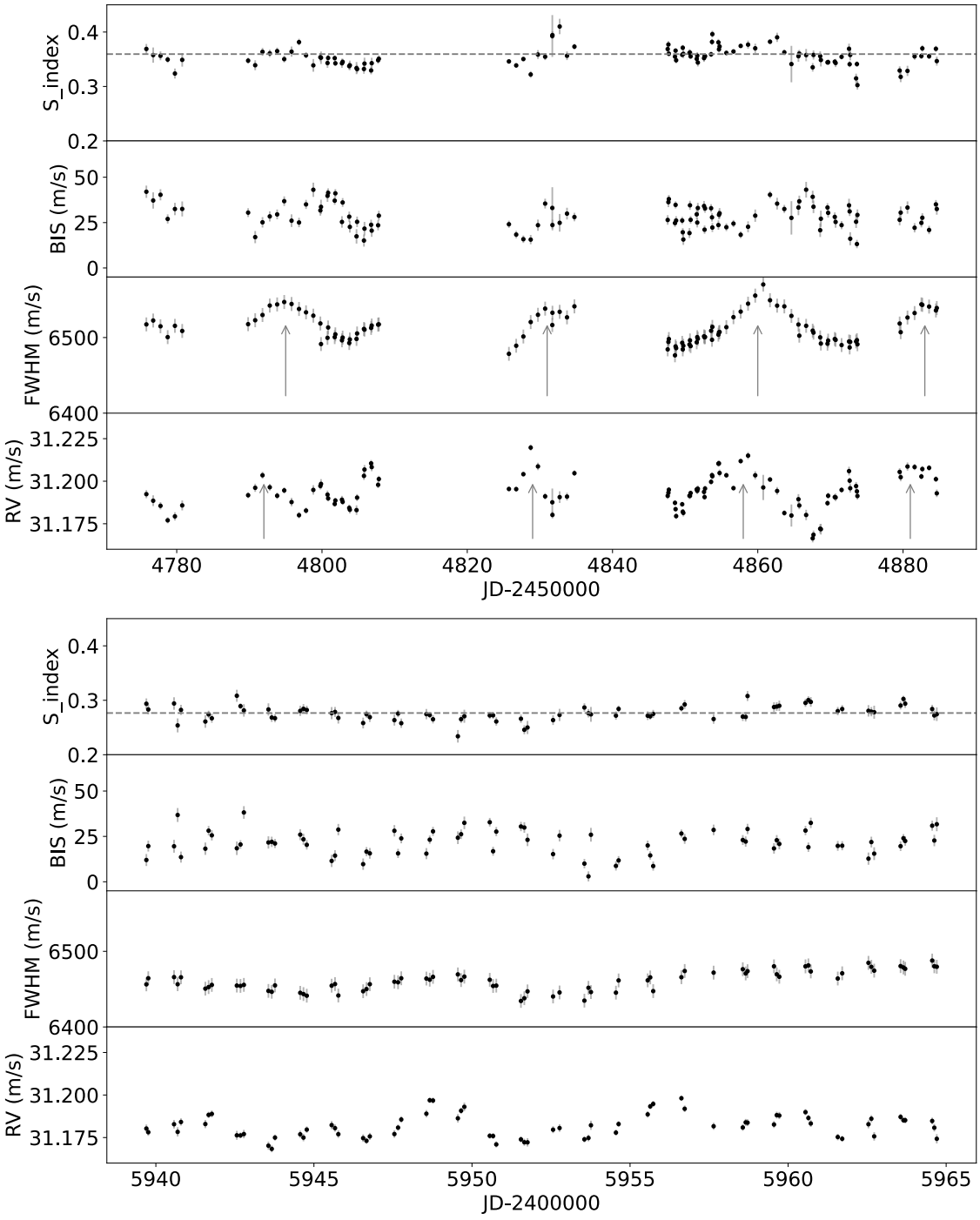


Figure 3.4: The spectroscopic information measured during the HARPS run spanning over November 2008 and February 2009 is given in the top figure. The RV, FWHM, Area, BIS and S-index data have been plotted against the Barycentric Julian date and shown in each panel from bottom to top. The figure at the bottom shows the respective representation of spectroscopic data throughout the 2012 HARPS run. The dashed grey lines running horizontally in the top panels of both figures show the mean values of the S-index, indicating the difference in the level of stellar activity during both observing runs. There are arrows given in the RV, FWHM and Area panels of the left figure, pinpointing that the RV peaks occur $\sim$ 2 days before the FWHM and Area peaks. The respective uncertainties are shown using grey error bars. Figure from Anna John et al. (2022).

periodic variability structures. The FWHM variations exhibit a fairly smooth periodic pattern, reflecting the photometric rotation period of ∼23 days (Léger et al., 2009) in both seasons.

A close inspection of Figure 3.4 shows a temporal offset between FWHM, Area and the RV. The RV maxima occur ∼2 days before those of FWHM and Area. In the Sun, Collier Cameron et al. (2021) observed similar time lags between the RV, FWHM and BIS. Such temporal shifts between RV and several activity proxies were reported in previous studies as well (e.g., Santos et al., 2014; Queloz et al., 2009). While Queloz et al. (2009) suggest these time-structured variations between these two parameters to be associated with star-spot-related variability in CoRoT-7, Collier Cameron et al. (2021) note that the same phenomenon is present even when the RV signal is dominated by facular suppression of the convective blueshift. Haywood et al. (2014) detail the suppression of convective blueshift by active regions surrounding star-spots to have a much greater impact on RV than the flux blocked by star-spots, particularly in slowly rotating stars like the Sun. CoRoT-7 is much more active than the Sun and likely spot-dominated.

3.2.2 SCALPELS periodogram analysis

We applied SCALPELS to the barycentric RVs of CoRoT-7, and examined the results using a Generalized Lomb-Scargle (GLS) periodogram (Zechmeister & Kürster, 2009) for different subsets of the data (2008-9, 2012 and combined; See Figure 3.6), given the varied level of activity in both seasons as discussed in Section 3.2.1. The first two components (blue and orange) in Figure 3.5 & 3.6, show the observed barycentric RVs from which their mean is subtracted ($\mathbf{v}_{\rm obs} - \langle v \rangle_{\rm obs}$), and the shape-driven ($\mathbf{v}_{\|}$) RV perturbations derived using the SCALPELS projection respectively. The green curve shows the cleaned shift-driven velocities, obtained after subtracting the shape-driven variations from the observed RVs.

Blue = Observed RVs

Orange = SCALPELS shape or activity component

Green = SCALPELS shift or planet component

The shift periodograms (green) in Figure 3.6 show strong peaks at 3.69 days, lining up with the orbital period of CoRoT-7c from the literature (See Table 3.1). CoRoT-7b's orbital period (0.85 days) is close to one day, and its peak in the periodogram is obscured, yet visible within the forest of aliases around one day.

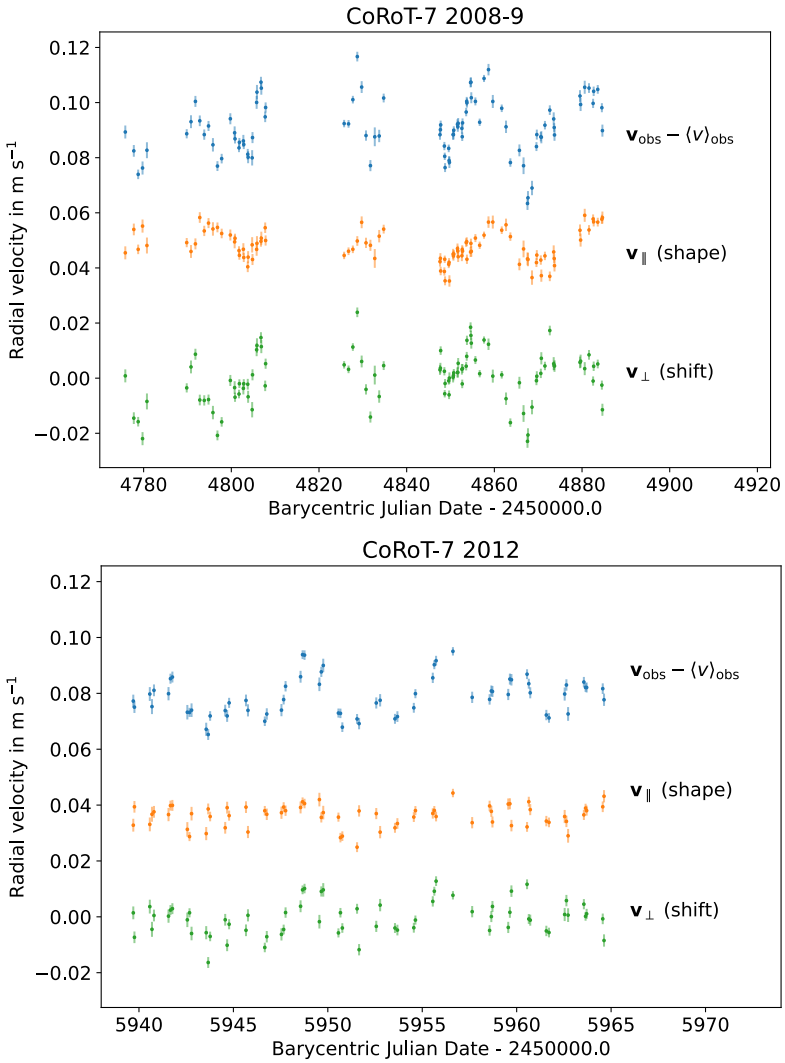


Figure 3.5: The top panel corresponds to the data from 2008-9 run, while the bottom panel shows 2012 data. In both panels, the blue curve at the top is the barycentric RV subtracted from its own mean. The middle orange curve represents the shape-driven component obtained from the SCALPELS projection, while the green curve manifests the 'cleaned' RV, which is the shift-driven component. Please note that an offset of ±60 m s^{-1} is introduced for clarity. The relatively small uncertainties are shown as semi-transparent error bars to avoid overcrowding. Figure from Anna John et al. (2022).

Along with these two confirmed planet signals, a third signal with a period of 8.96 days and an amplitude of ∼4 m s^{-1} also appears in all shift-driven periodograms. This peak is similar to the 9.02 days signal attributed to CoRoT-7d by Hatzes et al. (2011). We found no significant signal at the 8.96-day period in the activity-induced shape-driven periodogram, regardless of the dataset chosen. This argues against an activity-related origin proposed in several previous studies.

We would also like to point out the discrepancies existing in the orbital period of CoRoT-7d evaluated in previous studies. While Hatzes et al. (2010) postulated CoRoT-7d as a potential

third planet with an orbital period of 9.02 ± 0.019 days, Haywood et al. (2014) and Faria et al. (2016) found a signal at a period of 8.58 days which is almost half a day shorter than the period reported by Hatzes. Tuomi et al. (2014) spotted a signal at a period of 8.89 ± 0.01 days, but were unsure about its nature. We note that all these candidate periods from the literature correspond to individual 'fringes' in the interference pattern shown in Figure 2.5.

23 day periodicity

A strong peak in the combined observed periodogram shows up at around 23 days with an amplitude of 6.42 m s^{-1} , which corresponds to the stellar rotation period pinned down from photometry by Lanza et al. (2010); Haywood et al. (2014) etc. Although a shift signal at this period is present, the appearance of a shape-driven signal at the same period casts doubt on a planetary interpretation of this signal. While the presence of a shift signal without a counterpart in the shape signal points strongly to a planetary origin, we cannot completely ignore the possibility of some form of activity in CoRoT-7 producing a shape change in the profile that closely mimics a shift. Therefore, we have to investigate further the nature of the 23-day signal.

The complex frequency structure of the periodogram for the combined RV datasets (Figure 3.6) arises from both cycle-count uncertainty and aliasing. In principle, all of this structure should be attributable to a finite number of orbital signals combined with a quasi-periodic activity signal and noise. Given the difficulty in interpreting the system using standard GLS periodograms, especially when the window function is complex and there are multiple, non-harmonic signals present, we are inclined to move on to the next steps in TWEAKS.

3.2.3 l_1-periodogram

Figure 3.7 shows the ℓ_1- periodograms (refer to Section 2.1.5 for description), after un-penalizing the 0.85-day signal of the transiting planet. In the ℓ_1- periodogram terms, this means that we are always including the 0.85 days planet in the fit. While the top panel represents the ℓ_1-periodogram for the raw RVs, dominated by the long-term activity signals, the shift RV counterpart is shown in the bottom panel of Figure 3.7. After simultaneously decorrelating the stellar activity-induced shape variations using the SCALPELS **U** vectors, the two non-transiting planet candidates dominate the periodogram. Now the 4 signals with the greater significance and lowest false alarm probabilities are found at periods 3.70 days, 8.96 days and a closely

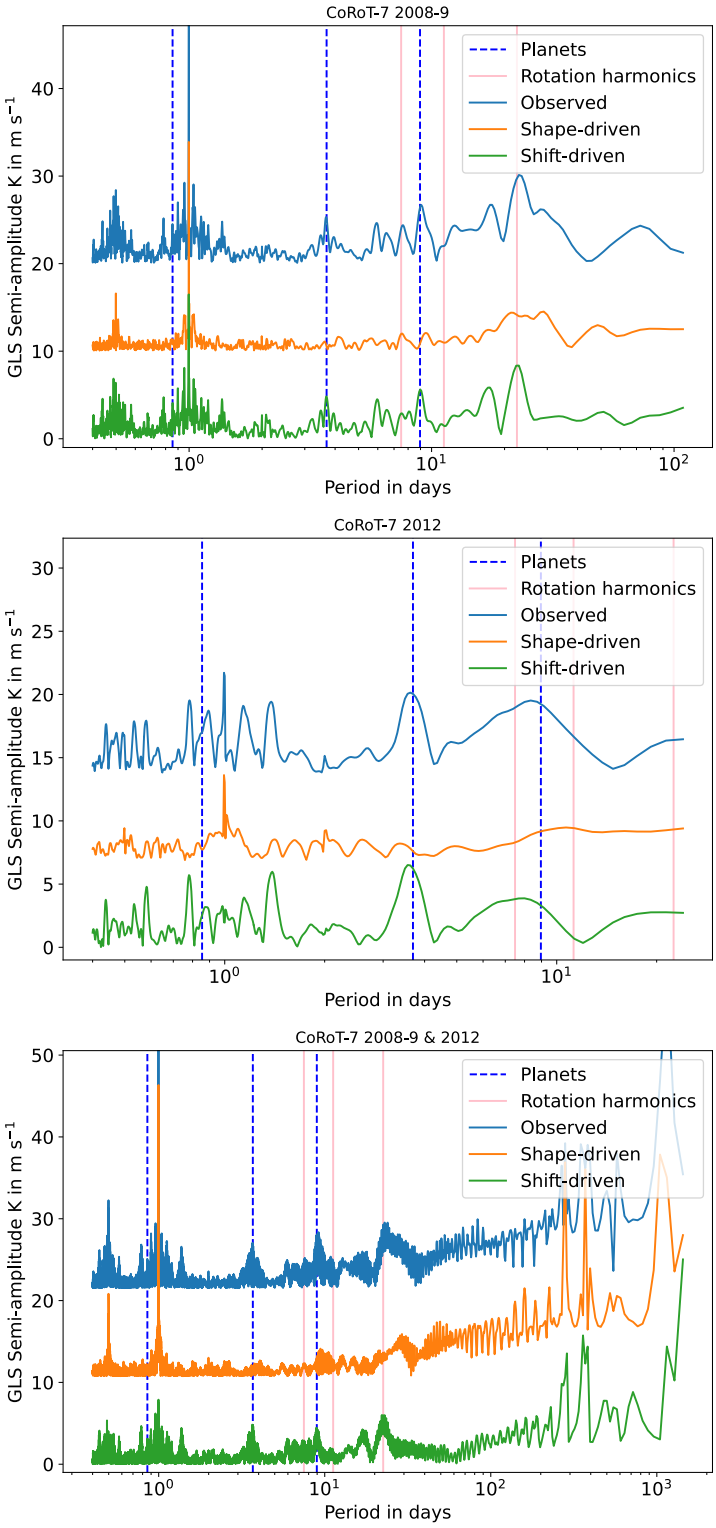


Figure 3.6: The left, middle and right figures represent 2008-9, 2012 and combined data sets respectively. In each figure, the top periodogram (blue) is for the measured velocities derived from the CCF. The middle trace (orange) is for the shape-driven velocities ($\mathbf{v}_{\parallel}$) obtained from the SCALPELS projection. The third periodogram (green) is the shift-driven velocities ($\mathbf{v}_{\perp}$) remaining after the subtraction of shape-driven velocities from the observations. Light blue dotted vertical lines denote the orbital periods of all confirmed and potential planetary companions. Figure from Anna John et al. (2022).

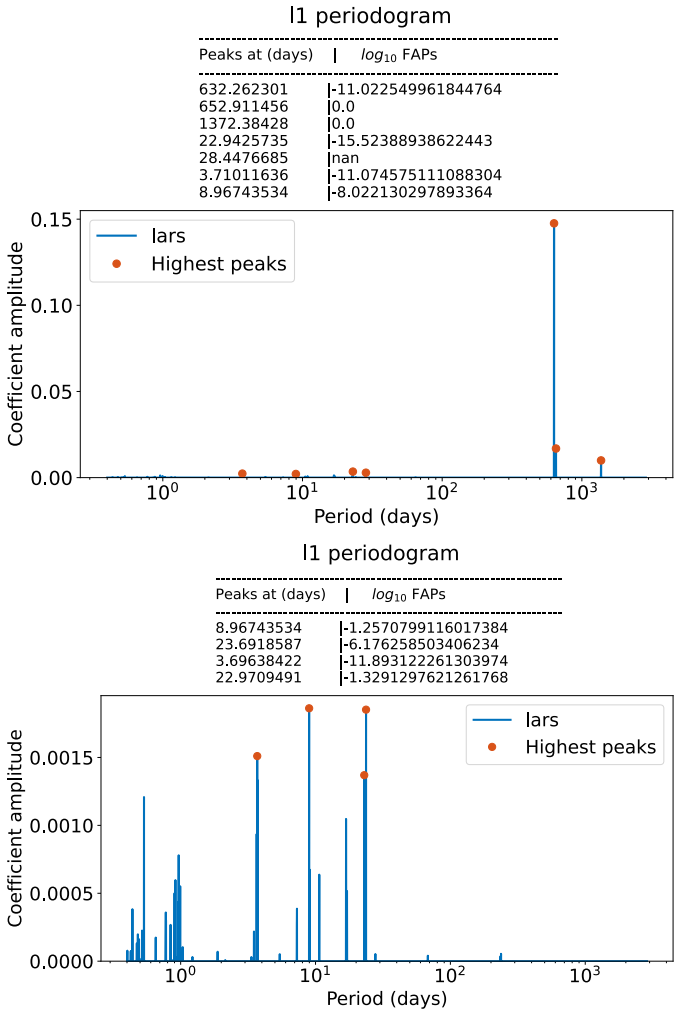


Figure 3.7: Top: The ℓ_1- periodogram of the raw RVs of CoRoT-7 after unpenalizing the transiting 0.85 days signal of CoRoT-7b, with the strongest spikes marked with red dots at the top. Traces of peaks for the planetary signals of CoRoT-7c (3.7 days) and CoRoT-7d (8.9 days) along with the stellar rotation signal (∼23 days) are present after the dominant long-term activity signals. **Bottom**: The ℓ_1-periodogram for the raw RVs after SCALPELS decorrelation vectors were fitted simultaneously as the planet signals, showing the strongest peaks at the orbital period of planets (8.96 and 3.71 days) and the stellar rotation period with the activity trends suppressed significantly. The corresponding False Alarm Probabilities (FAP) are also listed above each panel. Figure from Anna John et al. (2022).

spaced pair (23.69 and 22.94 days) around the stellar rotation period. The long-term signals are weakened considerably. All other strong peaks that showed up in the periodogram are identifiable as one-day aliases ($|\frac{1}{P} \pm 1|^{-1}$) of these four signals.

The presence of two closely-spaced signals near 23d in both RV sets suggests that, in the ℓ_1 periodogram's sparse representation of the data, the change in the amplitude of this signal between 2008-9 and 2012 is modelled as a 1990 days beat pattern between the two closely-spaced sinusoids.

The presence of low-amplitude peaks in Figure 3.7 may also indicate the presence of

correlated, non-Gaussian noise, which we expect to be the case here. The ℓ_1- periodogram is successful at shortlisting the most plausible candidate planet signals in our data and suppressing the parasitic signals. To validate the physical origin of these signals, we computed additional ℓ_1 periodograms for the shape-driven RVs and the activity indicators, as shown in Figure A.1. None of these show traces of peaks at or around 0.85, 3.69 and 8.96 days, adding to the case for their planetary origins, especially for the third signal. Having identified the most plausible signals, we now turn to fitting them simultaneously with SCALPELS.

3.2.4 Simultaneous sinusoidal planet orbit fitting

In Section 3.2.2, we projected the observed RVs, including planet signals, into the space spanned by the SCALPELS **U**-vectors. Although we discovered no connection between the shape and shift RVs (See 2.3), if any of the planet signals are correlated with any of the basis vectors, they may be partially absorbed in the shape signal. This is problematic, as we cannot guarantee orthogonality between any irregularly sampled orbital superposition with the components of the SCALPELS **U** vectors.

The orbital perturbations of any planet and the SCALPELS projection process must therefore be modelled self-consistently for the signal separation to recover their semi-amplitudes as reliably as possible. For this reason, Collier Cameron et al. (2021) solved simultaneously for shape-driven variations and a set of circular Keplerian orbits with known periods. Here we use a similar approach, using the set of orbital frequencies obtained with the ℓ_1- periodogram as inputs. The approach of obtaining a simultaneous solution involved determining the shape-driven variations from the difference between observed RVs and the orbit model velocities (Collier Cameron et al., 2021).

The assumption of circular orbits for candidate signals of the three planets can be justified by considering the tidal circularization timescales. According to Dobbs-Dixon et al. (2004) we could assume negligible eccentricities for the signals at 0.85 and 3.69 days, having periods less than 6 days and hence accounted for by dissipation of tidal disturbances within their envelopes that are induced by their host stars. For the third signal at a period of 8.96 days, we calculated the circularization timescale as described in Dobbs-Dixon et al. (2004), and found a value of ∼610 Gyr for an assumed planetary tidal quality factor $Q'_p = 10^6$. We also relaxed the prior on eccentricity for the analysis in Section 2.1.2 to allow the data to find the feasible value. The obtained value for eccentricity falls around 0.057±0.062, again supporting the assumption of

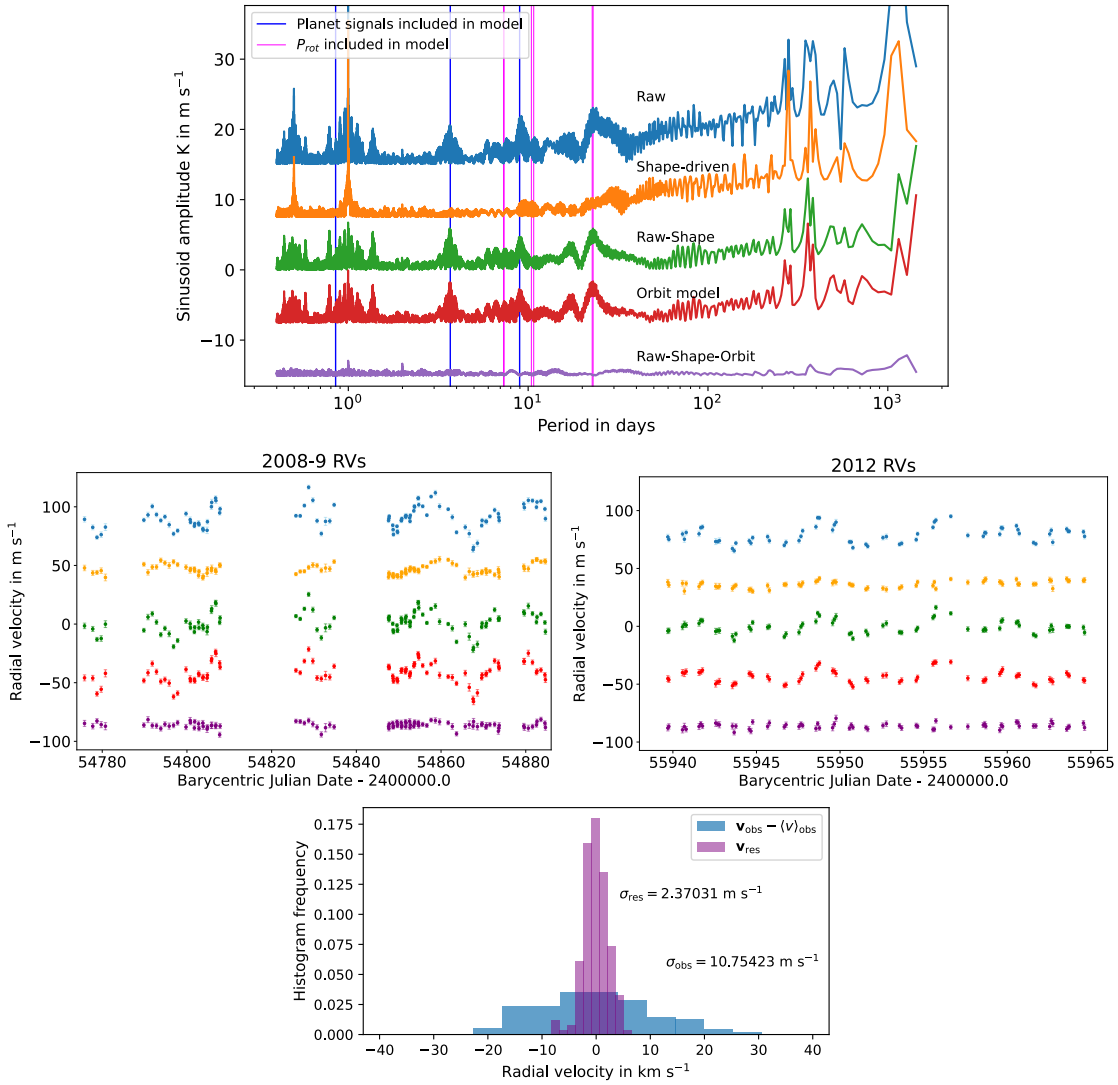


Figure 3.8: The periodograms of RVs derived from the barycentric CCFs when signal separation and orbit fitting of the sinusoidal signals of periods derived from the analysis associated with Figure 3.7 have been performed simultaneously. **Top**: The top three traces are as defined in the caption of Figure 3.6. The fourth trace (red) shows the periodogram of the fitted model of all the signals, including the 3 candidate planet signals (blue lines) and 3 pairs of closely spaced beat periods associated with the rotation period and its harmonics (magenta lines) found by the ℓ_1- periodogram. The orbital model (red) is now much closer in appearance to the shift signal (green), leaving little residuals in the bottom trace (purple), obtained after subtracting both shape-driven and orbital RV models from the observed RVs. Each trace is offset by $\sim$ 10 m s^{-1} for better illustration. **Bottom**: The resultant histogram of residual RVs shows significantly reduced RMS scatter. Figures from Anna John et al. (2022).

circular orbits.

The top panel of Figure 3.8 displays the SCALPELS RV periodograms. Four strong sinusoidal signals with non-zero FAPs were found from the ℓ_1- periodogram search. These periodograms (Figure 3.8) are obtained from simultaneously modelling the orbits of those signals while performing the signal separation. The orbit of the transiting planet signal was also modelled simultaneously with the two pairs of periods at $P_{\rm rot}/2$ and $P_{\rm rot}/3$. The orbital periods of candidate planets are marked by vertical blue lines at 0.85 days, 3.71 days, and 8.96 days, and beat periods for stellar rotation periods and harmonics by magenta lines at 22.94, 23.75, 10.42, 10.74, 7.30 and 7.35 days.

The top three traces are as defined in Section 3.2.2, with the only difference being the shape-driven periodogram here is obtained by subtracting the orbit model from the raw RVs. The fourth trace (red) shows the periodogram of the fitted orbit model of the candidate signals. The bottom periodogram in purple shows the residuals after subtraction of both the shape-driven and orbital RV models. It is worth noting the level of matching between both periodograms (green & red), even with the finest details, in this case.

A closer look at the orange and green periodograms reveals that the balance of the signal separation has been altered from those in Figure 3.6, by modelling the orbital motion explicitly at known periods. The algorithm efficiently reduces the correlation between the modelled orbit signals and modes of velocity variation driven by line shape changes alone. It is therefore preferable to fitting orbits to the shift RVs derived from a blind SCALPELS analysis.

To determine the orbital velocity amplitudes that would be recovered without applying profile-shape corrections to the velocities, we reduced the number of principal components $k_{\rm max}$ to zero. We found that, when the sinusoids are fitted to the observed RVs and left uncorrected for line-profile shape variations, the semi-amplitudes of the recovered signals depart from the semi-amplitudes recovered during the simultaneous modelling (see columns 2 and 4 of Table 3.2).

We also compared the recovered semi-amplitudes of the four strongest signals when fitted with and without the 23.69 days beat signal found in the ℓ_1- periodogram search, in the appendix (See Table A.1). The last two columns of both tables list the semi-amplitudes and associated uncertainties for the shift-driven RVs acquired in both cases. While the differences in the RV amplitudes for planets b, c and d between the 4-signal model and the 5-signal model

Table 3.2: The periods and semi-amplitudes of the strongest signals in the periodograms of raw ($k = 0$) and shape-corrected apparent shift-velocities ($k = k_{\max}$) from simultaneous modelling of CCF shape changes and planetary motion, made with prior knowledge of the periods for candidate planets (from the ℓ_1- periodogram) and pairs of beat period at the stellar rotation period and its harmonics are listed below.

P(days)	K(m s^{-1}) $k = 0$	σ_K(m s^{-1}) $k = 0$	K(m s^{-1}) $k = k_{\max}$	σ_K(m s^{-1}) $k = k_{\max}$
0.8535	4.185	0.263	3.552	0.306
3.6963	5.484	0.254	6.012	0.312
8.9674	6.877	0.324	5.482	0.380
10.4269	0.472	0.569	1.156	0.606
10.7409	4.215	0.535	2.560	0.597
7.3039	2.173	0.536	2.391	0.620
7.3583	3.004	0.570	3.276	0.655
22.9425	6.915	0.288	4.891	0.353
23.6918	4.077	0.273	3.311	0.348

are small when compared with their uncertainties, the 5-signal model gives a more faithful approximation to the behaviour of the 23-d activity signal between seasons.

The 23-day peak in the shift signal is therefore best modelled as the beat pattern between the two closely spaced periods, requiring 2 frequencies to be fitted for the stellar rotation period. In the 4-signal case, when a single sinusoid is fitted at this frequency, the 23-day peak shows strong cycle-count interference. We attribute this effect to the 1990-day beat period between the 22.7 days and 22.9 days signals being greater than the entire observation span of ∼1195 days. When both 23-day periods found by the ℓ_1- periodogram are fitted, the form of the 23-day peak in the model signal (red) is much closer in appearance to the shift signal (green) (See Figures 3.8 and A.2). However, there is still some power in the residuals at $P_{\rm rot}/2$ and $P_{\rm rot}/3$ (marked with pink vertical lines in Figure A.2), when only fitting 5 signals. This indicates that a purely sinusoidal model for the residual activity signal, even with two closely spaced frequencies, might be insufficient.

Therefore, it's better to model the shift-like component of the activity-driven RV signal in such a way as to minimise the residuals and make the residuals as flat as possible. This motivated us to include pairs of sinusoids for the $P_{\rm rot}/2$ and $P_{\rm rot}/3$ as shown in Figure 3.8. The RMS scatter is now significantly reduced to 2.37 m s^{-1} . The resultant orbital periods, semi-amplitudes, and uncertainties of the sinusoidal signals included in the orbit model are listed in Table 3.2.

3.2.5 Nested Sampling using GP +SCALPELS

As seen in Figure 3.6 the 23-day stellar rotation signal leaks into the shift component. This signal appears in both shape and shift, in contrast to the 0.85, 3.69, and 8.95-day signals, which only appear in the shift component and lack significant counterparts in the shape signal. I would like to emphasize that the plausibility of any planet candidates found in the shift component is not strongly affected unless there is some kind of activity that produces a shift signal, but no shape signal at certain periods. In addition, the existence of an activity-related shift signal is not a surprise, given that part of it is produced by the suppression of photospheric granular motion.

However, to account for any activity-induced shift-like signals that may elude the SCALPELS analysis, any remaining rotationally-modulated signals were modelled using GP regression applied to the RVs. This is where we explored the possibility of integrating SCALPELS with an efficient time-domain approach called KIMA.

Using KIMA (Faria et al., 2018), we sampled the posterior distribution of orbital parameters along with a complete Keplerian solver and a Gaussian Process. The posterior distribution constrained by the RV data was then used to probe the credible number of planets (given the data), the orbital parameters and the planetary masses. The priors on the hyperparameters of the GP were carefully tailored for CoRoT-7, as given in Table 3.3. To account for the stellar activity, we decorrelated the SCALPELS U-vectors $\mathbf{U_1}$, $\mathbf{U_2}$ & $\mathbf{U_3}$, as we found that the three leading principal components are the strongest contributors to the SCALPELS shape component.

Knowing the parameters of one planet provides a small amount of information about the parameters of another planet (Faria et al., 2016). Keeping this in mind, we treated the transiting planet CoRoT-7b as a 'known planet', adopting the updated transit parameters in Barros et al. (2014). By putting tight priors on the orbital period P and mid-transit time T_0 we ensured that the phase of this transiting planet signal is consistent with the photometry. For the known planet, we put a lower limit on K derived from the mass it would have if it were made of compressed water, with a radius of 1.528 $R_\oplus$ (Barros et al., 2014) that is around 1.56 $M_\oplus$ (according to the Growth model interpretation of mass distribution proposed by Zeng et al. (2016)), giving an RV amplitude of $K_{min} \sim 1.08$ ms^{-1}. The upper limit was set to be about a physically-motivated number of $K_{max} \sim 10$ m s^{-1} corresponding to the 10 $M_\oplus$ expected for a pure iron planet of that size.

Table 3.3: Prior distribution for model parameters. $\mathscr{LU}$ - Log Uniform ; $\mathscr{MLU}$ - Modified Log Uniform; $\mathscr{K}$ - Kumaraswamy ; $\mathscr{U}$ - Uniform ; $\mathscr{G}$ - Gaussian.

Notations	**Orbital parameters**	Priors
P	Orbital Period	$\mathscr{LU}$(0.5,100)
K	Semi-amplitude	$\mathscr{MLU}$(1,25)
e	Eccentricity	$\mathscr{K}$(0.867,3.03)
ϕ	Orbital phase	c
ω	Longitude of line of sight	$\mathscr{U}$(0, 2π)
	GP parameters	
η_1	Amplitude of correlations	$\mathscr{LU}$(1.6,2.7)
η_2	Decay timescale	$\mathscr{LU}$(15,60)
η_3	Correlation period	$\mathscr{U}$(19,30)
η_4	Periodic scale	$\mathscr{U}$(-1,0)
s	Extra white noise	$\mathscr{LU}$(0.5,max v)
	For known planet	
P	Orbital Period	$\mathscr{G}$(0.85359, 5.7e-7)
K	Semi-amplitude	$\mathscr{MLU}$(1.08,10)
e	Eccentricity	$\mathscr{U}$(0, .1)
t_0	Time of mid-transit	$\mathscr{G}$(54398.0776, 0.0007)
ω	Longitude of line of sight	$\mathscr{U}$(-π,π)
β	**Activity indicators**	$\mathscr{G}$(0,10)

Taking into account the existing knowledge of the rotation period from the literature (e.g., Queloz et al., 2009; Haywood et al., 2014) and the present work, we restricted the GP hyper-parameter η_3 (representing the stellar rotation period) to a sensible range between 19 and 30 days. We adopted a prior period distribution for non-transiting planets with periods between 0.5 days and the duration of the 2008-9 data, i.e., 100 days.

As an initial experiment, the number of planets N_p was set to be a free parameter in the MCMC, with a uniform prior between 0 and 5. Considering N_p as an unknown parameter, KIMA uses a birth-death MCMC proposed by Brewer (2014), which allows estimating the improvements in evidence values when the sampler switches between solutions with different N_p values.

The resultant joint posterior distributions for the semi-amplitudes of the signals are shown in the top panel of Figure 3.9. We found that two non-transiting planets are favoured by the data, considering the ephemeris of the transiting planet CoRoT-7b as 'known' in the model. Amidst the posterior spread in the parameter space, these are seen as heavily populated regions at $P_c = 3.69$ days and $P_d = 8.96$ days.

Reviewing this posterior distribution triggered another experiment by fixing the number of

planets as N_p=2 with the rest of the priors as the same as in Table.3.3. The resultant posterior distribution of RV semi-amplitudes is represented in the bottom panel of Figure 3.9, with well-constrained peaks around $P_c = 3.69$ days and $P_d = 8.96$ days, within the orbital period uncertainties reported in the literature. Supporting the presence of CoRoT-7c and CoRoT-7d, the same constrained detail can be seen in the posterior distribution of eccentricities also (See Figure A.4). From the 10000 effective samples in the joint posterior distribution, the evidence $\log\mathcal{Z}_2$ for this model was calculated as $\log(p(D\,|\,M))$ = -475.46.

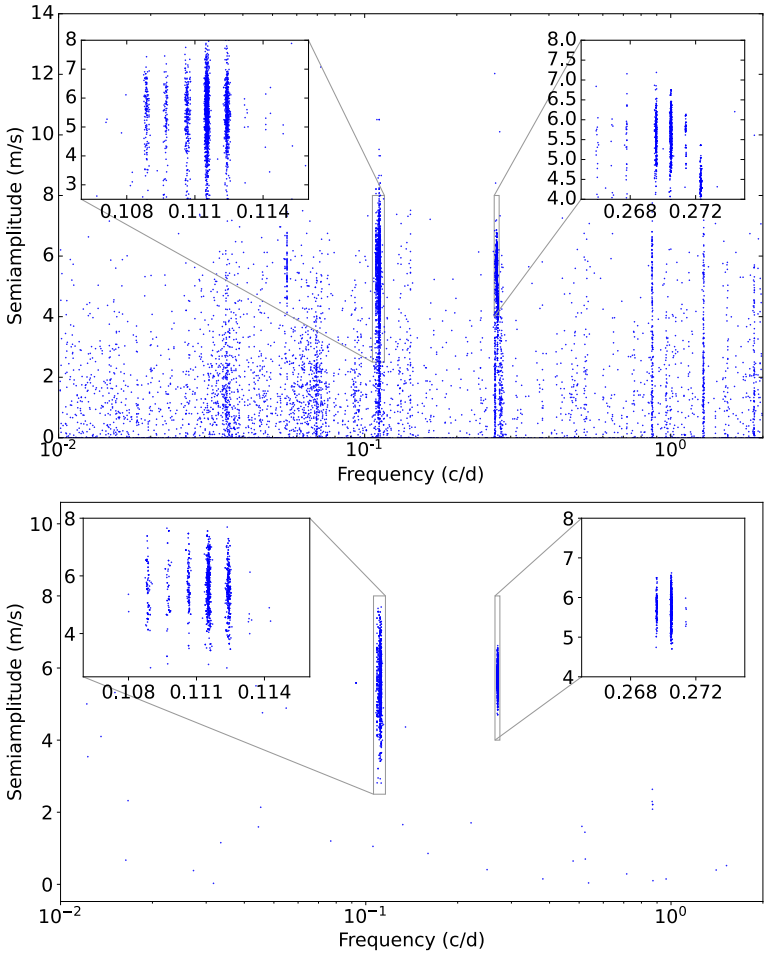


Figure 3.9: The posterior distribution for semi amplitudes (ms^{-1}) in log frequency space (cd^{-1}) is shown. The transiting 0.85 days planet is considered as 'known'. The inset plots zoom into the clusterings corresponding to the ∼9 days planet and ∼3.7 days planet (right), showing the interference patterns occurred due to the widely separated observing seasons. **Top** : Setting the number of planets, N_p as a free parameter with uniform priors between 0 and 5. **Bottom** : N_p is fixed at 2. Figure from Anna John et al. (2022).

If we fix N_p=1, we recover the orbital parameters for CoRoT-7c only, similar to those reported by Faria et al. (2016). The evidence is degraded to $\log\mathcal{Z}_1$ = -479.20 for this model (please note the minus sign). The difference in log evidence between the model with 2 non-

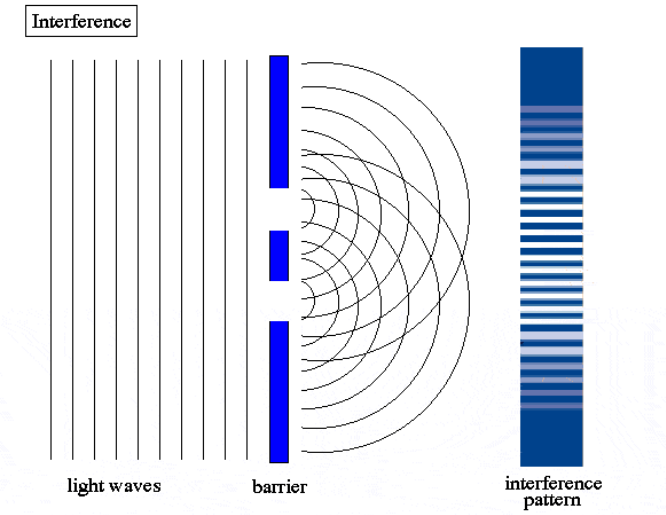


Figure 3.10: Interference pattern of coherent light obtained after passing through two slits. Figure credit: Google.

transiting planets and the model with 1 non-transiting planet is 3.8. This is in the substantial support range of the Jeffreys scale for Bayesian evidence comparison (Jeffreys 1961). These results therefore indicate a 2-planet model to be 42.31 times ($\mathscr{Z}_1$-$\mathscr{Z}_2$) more probable than a 1-planet model. We emphasize that, when saying a 2-planet model, we are referring to the non-transiting planets CoRoT-7c and CoRoT-7d, as the model considers the orbital parameters of transiting planet CoRoT-7b as known.

After 'confidently detecting' CoRoT-7b and CoRoT-7c, Faria et al. (2016) pointed out that it is more likely that there are four planets according to the posterior distribution. Their posterior showed smaller peaks around ∼9 days and ∼2 days. However, adding the SCALPELS basis vectors to the model for activity decorrelation enabled us to confidently recover the orbital signal of CoRoT-7d at 8.96 days. We were unable to recover any signal around 2 days, as it is removed by SCALPELS for being some artefact arising from activity, aliasing or stellar rotation harmonics.

We also compared the run 'with SCALPELS' against the one 'without SCALPELS', to validate the improvement in the model with the inclusion of the SCALPELS shape-signal decorrelation vectors along with a GP. The results are given in Table 3.4. We found that de-trending the RVs for line shape variations using the SCALPELS basis vectors yields a model that is 4.14 times more probable than a model that does not include these activity variations.

Table 3.4: The mass and mean-density (column 4 & 5) calculated for the transiting planet CoRoT-7b from the posterior distributions of orbital period and RV semi-amplitudes (column 2 & 3) considering different models are listed. Important note: CoRoT-7b is considered as a known planet in all these models with transit parameters from Barros et al. (2014) (See Table 3.3), who updated the planet radius as 1.528 ± 0.065 $R_\oplus$. N_p is the fixed number of planets in each model. The first three rows give results from models with GP + SCALPELS and different numbers of planets, whereas the last three rows are the results for different models with GP alone and different N_p.

Model	P_b[days]	$K_b[ms^{-1}]$	M_b $[M_\oplus]$	ρ_b[g cm^{-1}]	log $\mathscr{Z}$	$\mathscr{Z}_n$-$\mathscr{Z}_{n+1}$
GP +SCALPELS						
N_p=1	0.854 ± 6.01e-7	3.943 ± 0.559	5.595 ± 0.794	8.641 ± 0.269	-479.20	–
N_p=2	0.854 ± 5.87e-7	4.291 ± 0.460	6.056 ± 0.653	9.355 ± 0.235	-475.46	42.31
N_p=3	0.854 ± 5.78e-7	4.278 ± 0.453	6.048 ± 0.643	9.322 ± 0.234	-474.58	2.40
GP only						
N_p = 1	0.854 ± 5.48e-7	4.007 ± 0.550	5.679 ± 0.781	8.786 ± 0.264	-478.58	–
N_p = 2	0.854 ± 5.53e-7	4.293 ± 0.468	6.087 ± 0.664	9.394 ± 0.236	-474.87	40.89
N_p = 3	0.854 ± 5.72e-7	4.331 ± 0.476	6.146 ± 0.676	9.459 ± 0.236	-474.27	1.82

Planet signals and interference pattern

The zoomed-in view of the over-density regions in Figure 3.9 revealed interference patterns associated with CoRoT-7c and CoRoT-7d at 3.69 days and 8.96 days. Let's go back to our high school Physics textbook and look at Young's double slit experiment, where allowing coherent light to pass through two tiny apertures produces interference patterns (dark and bright fringes, Figure 3.10). This occurs because each slit acts as a wave source, facilitating waves from the two slits to interact constructively or destructively at the screen. It is interesting to consider the two seasons of observations in 2008-9 and 2012 as the two slits; a 'double-slit experiment', in which the path lengths between the light waves from two slits give rise to an interference pattern due to the phase shift. Here, the sharp peaks are separated by a frequency $\sim \frac{1}{1200}\ cd^{-1}$, accounting for the time separation (1200 days) between two observing campaigns. This demonstrates the coherency of these signals across seasons, as both signals are detected unambiguously in both seasons.

FIP and TIP calculation

We performed the FIP analysis (Section 2.1.4) in the frequency space, setting the bin size equal to the Nyquist frequency resolution of the 2008-9 season rather than the full duration of the data. Since the probability of each planet is spread over several interference fringes, we should marginalize the uncertainty in the cycle count to establish the existence or otherwise of the planet. A frequency bin size of 0.005 cycles per day captures all probabilities while fully resolving the structure of the 2008-9 periodogram. This allowed us to pool the power distribution across different fringes.

We searched for various planets simultaneously using a frequency window sliding along the entire frequency space spanned by the posterior distribution of the 2-planet model (with CoRoT-7b being known). Consequently, two strong peaks were found, at 3.697 days and 8.965 days, both with significantly minimal FIP and maximal TIP, reinforcing their planetary nature (Figure 2.6). To avoid being misled by the shared power among adjacent interference peaks linked with relevant signals, a basic histogram was used to find these most probable periods, See Figure A.3.

Being a trans-dimensional sampling algorithm, KIMA enabled us to directly sample the joint posterior of the number of planets (N_p) and other orbital elements. Several ambiguous

detections are prompted at spurious periods due to the cross-talk between different aliases. FIP and/or TIP attenuated the invalid detections associated with these periods. Interesting is the enfeebled TIP and very large FIP (∼1) at the stellar rotation period of ∼23 days and its harmonics, providing more evidence against the planetary nature of a body at this period. From Figure 3.8 alone, one could argue that since the stellar rotation period (∼23 days) is seen in the shift signal, any other period that also survives when the shape signals are subtracted could also be activity-driven. However, a better validation for the dynamical origin of the candidate signals is achieved here by incorporating a GP to account for any remaining quasi-periodic signals. It is promising to see that both the non-transiting planet signals with orbital periods 3.69 and 8.96 days survive the GP and produce a 2-slit interference pattern as shown in Figure 3.9.

3.2.6 Stacked Bayesian General Lomb-Scargle periodogram

To investigate the stability of the periodic signals found in the previous analyses, we used a stacked normalized BGLS periodogram (Mortier & Collier Cameron, 2017) of the shift-driven RV data. Following the method of Mortier & Collier Cameron (2017), we derived the stacked periodograms of CoRoT-7 from the cleaned shift-driven RVs obtained after the SCALPELS shape-signal separation. The BGLS periodogram was initially obtained from the first *4* data points and then refined progressively by adding more one by one. This analysis was done for different subsets of data, obtained from the two observing runs (2008-9 & 2012) separately (Figure A.6) and also for the entire data set as shown in Figure 3.11. Figure 3.11 is colour-coded based on the log probability of the periods in the BGLS periodogram. One should expect to see the SNR of a planetary signal increasing asymptotically in proportion to the square root of the number of data points added.

Fringes began to appear in the stacked BGLS periodogram when both data sets were analysed together, as seen from Figure 3.11. These are caused by the uncertainty in the cycle count elapsed across the three-year gap separating the 2008-9 and 2012 RV campaigns (Mortier & Collier Cameron, 2017). However, the combined data set works well for tracking the candidate signals.

Except for the 23 days signal, all other candidate signals stand out with clearly growing probability features at periods 0.85, 3.69 and 8.97 days, supporting the coherent nature of these signals. It is worth noting that the 23.47 days signal grows unsteadily, with the strength

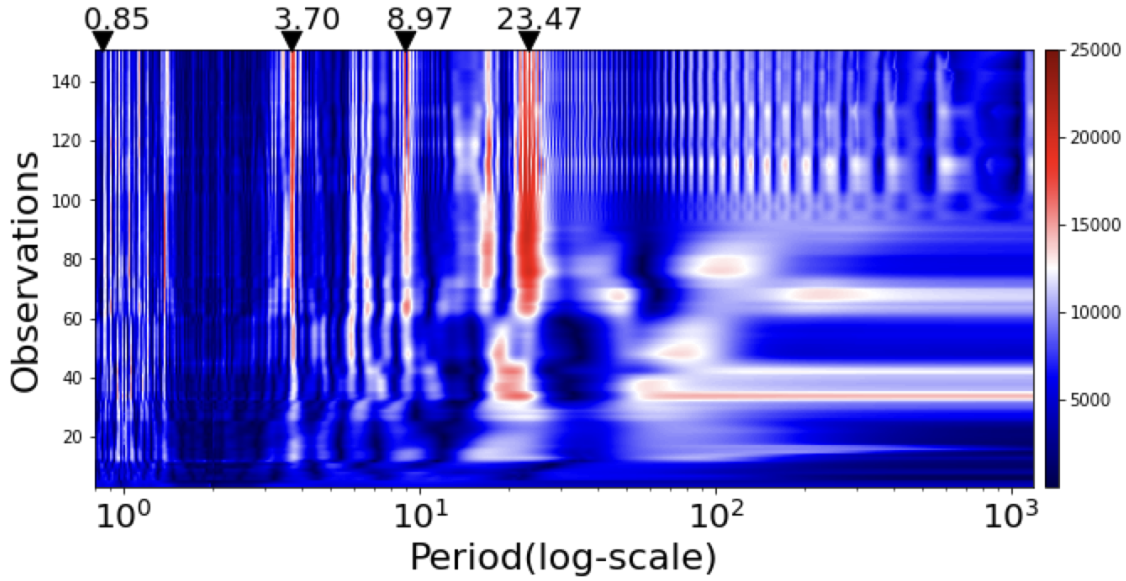


Figure 3.11: The Stacked Bayesian General Lomb-Scargle periodogram for CoRoT-7 with the number of observations in the y-axis plotted against the period in days displayed in logarithmic scale. The colour gradient designates the logarithm of probability, where the redder is more promising. Figure from Anna John et al. (2022).

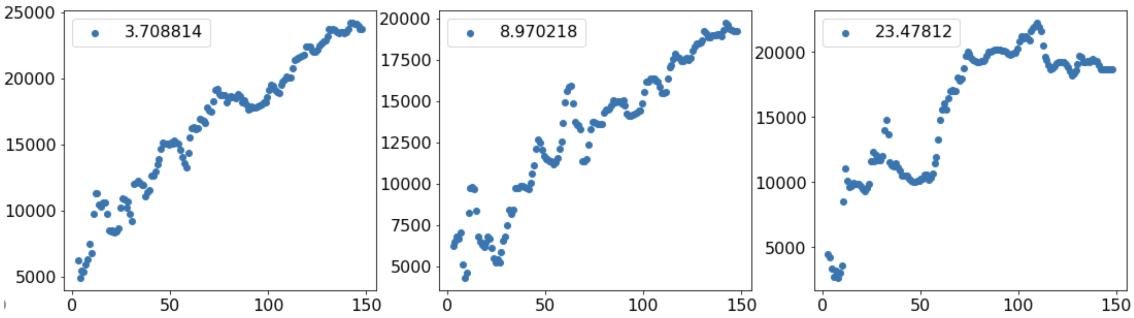


Figure 3.12: x — axis: Number of observations, **y — axis**: signal-to-noise ratio. The SNR plotted against the number of observations for the optimal periods of individual candidate signals at periods 3.70 days, 8.97 days and 23.47 days. Figure from Anna John et al. (2022).

maximizing around 70 to 100 observations and decreasing later on, see also Figure 3.12. We looked particularly into the 8.97 days signal to check its behaviour after adding more data points from the 2012 campaign and found that the SNR constantly increases with more observations, as expected for a real and coherent signal.

Tracking the significance of detection

We tracked the significance of each strong signal in Figure 3.11 by plotting the SNR against the number of observations (Figure 3.12). The optimal frequency for each candidate signal was obtained from the dominant peaks in the top line of the stacked BGLS periodogram.

Each point in Figure 3.12 represents the maximum SNR for the set of all RVs up to that point in time. We particularly looked into the 3.69 and 8.97 days signals and found that they exhibit a monotonic rise towards asymptotic square-root behaviour, confirming that they are strictly periodic and were present throughout the observations, hence substantiating the existence of CoRoT-7c and CoRoT-7d. As a result of the steady growth, their detections reach $\sim 25\sigma$ and $\sim 20\sigma$ respectively.

On the other hand, the detection significance of the 23.47 days signal first grows rapidly but then levels out and even drops at some point when more observations are added. This is a clear sign of an incoherent signal. We have the advantage of having longer baseline which makes this analysis relevant, which would not be the case for shorter baselines and longer orbital periods. This evidence again strengthens the activity origin of the 23-day signal, due to its unstable nature throughout the period of observation.

3.2.7 Mass & Mean density of CoRoT-7b

As discussed in Section 3.1, the mass of the transiting planet CoRoT-7b has been a subject of debate for more than a decade. Table 3.1 shows the wide range of masses as reported in the literature. This can be attributed to the relatively high level of activity in the host star. Table 3.4 shows different posterior distributions obtained for semi-amplitudes for different KIMA models. Using model comparison we found that the model with *GP + SCALPELS +* $N_{p=2}$ *+CoRoT-7b*$_{known}$ is the optimal one. The posterior distributions for the GP parameters for this analysis is shown in Figure A.5. A significant improvement in the evidence ($\mathscr{Z}$) was observed when fixing the number of non-transiting planets at 2. This model yielded an RV semi-amplitude of 4.291 ± 0.46 m s^{-1} for CoRoT-7b.

The planetary mass was calculated using this semi-amplitude, assuming a circular orbit. The orbital inclination i and stellar mass M_* were obtained from Barros et al. (2014) as 80.98 ± 0.51 degree and 0.915 ± 0.017$M_\odot$ respectively. We determined the mass of CoRoT-7b to be 6.056 ± 0.653$M_\oplus$, making it a super-Earth. This mass measure is within the uncertainty limit of several previous studies (Faria et al., 2016; Hatzes et al., 2010). Incorporating with the planet radius of 1.528 ± 0.065 $R_\oplus$ from Barros et al. (2014), the planet bulk density was calculated as 9.355 ± 0.235 g cm^{-1}.

We used the MANIPULATEPLANET tool developed by Zeng et al. (2016) to get an estimate

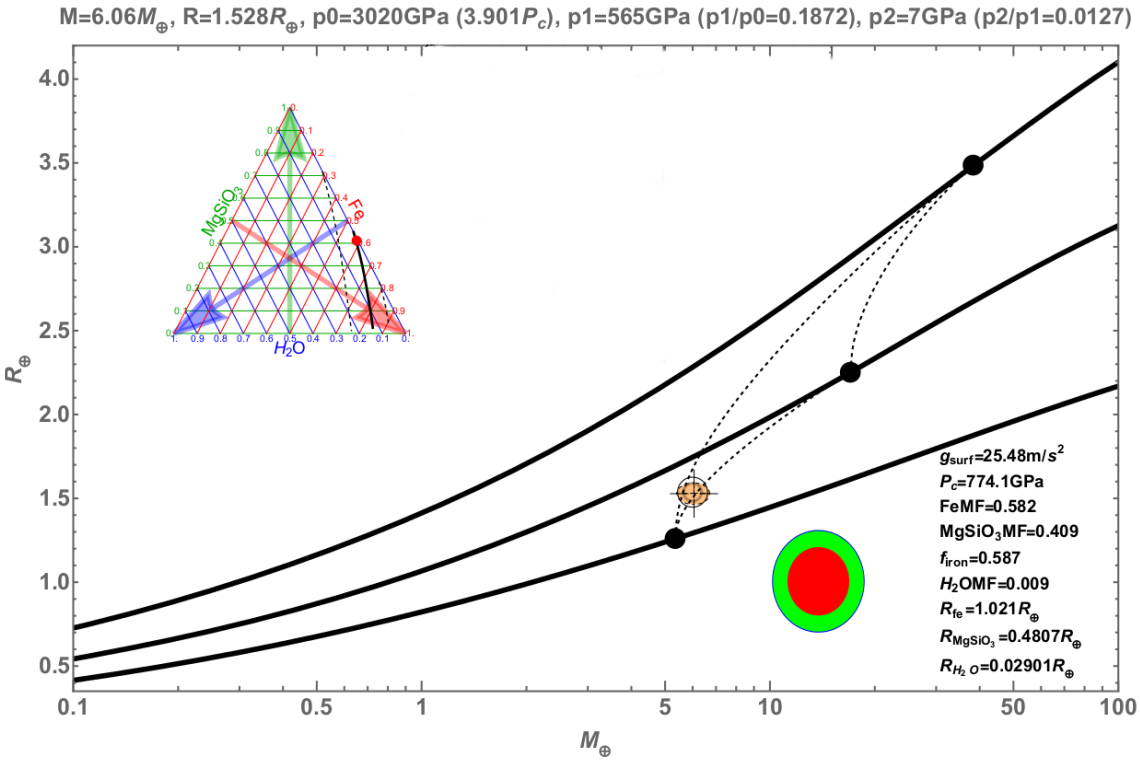


Figure 3.13: The most plausible composition of the transiting planet CoRoT-7b backed by the derived mass and radius for minimum core pressure is shown as obtained from the *manipulateplanet* tool developed by Zeng et al. (2016). The thick black curve in the ternary diagrams shows the degeneracy of the tri-layer model due to the trade-off between the three components, but not due to the mass or radius measurement uncertainties. Figure from Anna John et al. (2022).

of the planetary composition. The derived value of mass and a minimum core pressure (3020 GPa) supports a planet with an iron core (Fe: 58.7% by mass), Silicate mantle ($MgSiO_3$: 40.9%) and a thin layer of water on the surface (H_2O: 0.9%), favouring the rocky composition discussed in the literature.

For the given value of the mass and radius of the planet, we varied the core pressure in the model to investigate the degeneracy in the composition of the planet up to a maximum supported core pressure of 3346 GPa. This maximum core pressure proposes a dense iron core (Fe: 96.3 %) immersed in deep ocean mantle (H_2O: 12.6 %) with a compromised silicate layer($MgSiO_3$: 3.2 %). Also, the lower value of planet mass (M-1σ) suggests a composition that can be up to 78.6 % Fe, 6.2% $MgSiO_3$ and 15.1% H_2O. However, none of the above compositions seems to be feasible, considering the high equilibrium temperature (1756 K) of CoRoT-7b estimated assuming a zero albedo (Barros et al., 2014). Therefore, the most plausible composition occurs to be the one discussed earlier in this section, shown in Figure 3.13 with little or no surface layer of water. It should be borne in mind that the MANIPULATEPLANET tool assumes a condensed volatile layer typical of a cold planet. More recent studies, (e.g.

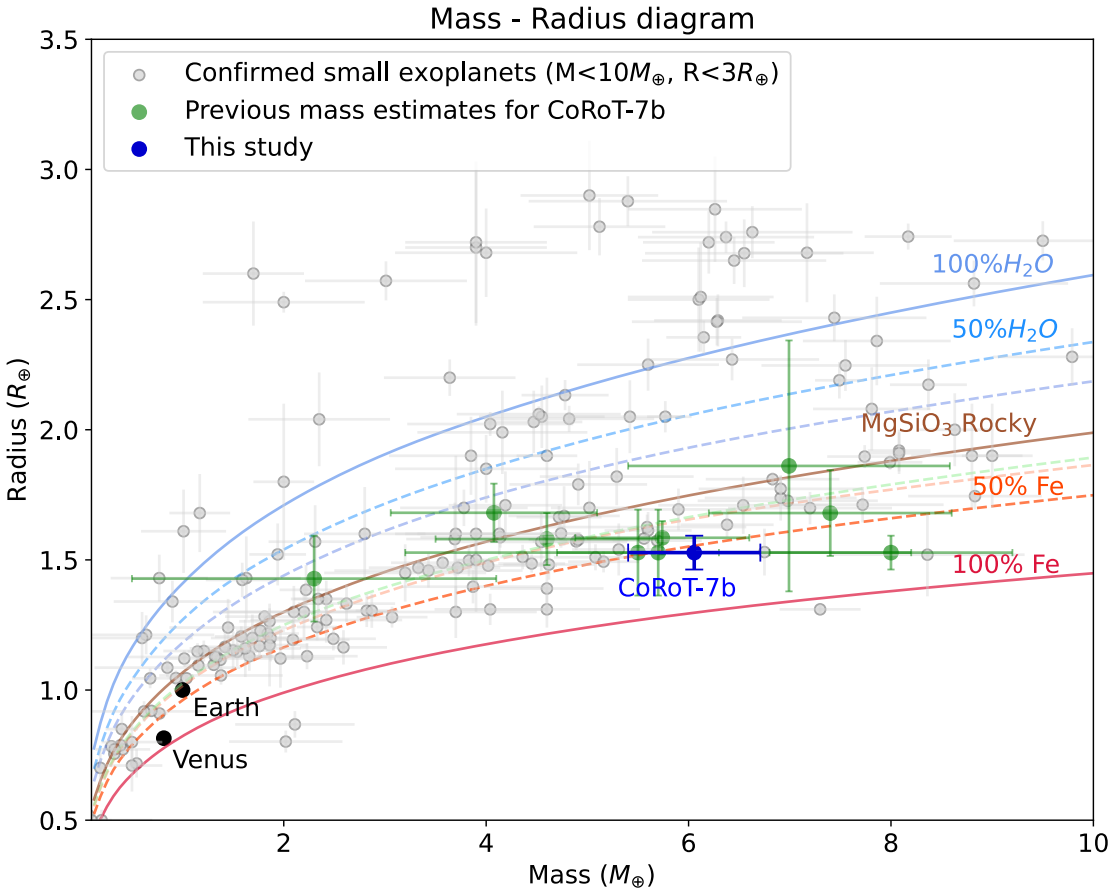


Figure 3.14: Mass–radius relation for planets with masses <10 $M_\odot$ represented by grey circles. The plot also includes Earth and Venus, for reference. The lines show models of different compositions, with solid lines indicating single-composition planets (either H2O, MgSiO3, i.e., rock, or Fe). The dashed and dotted lines indicate Mg-silicate planets with different amounts of H_2O and Fe. The green circles show the widely different mass estimates for the transiting planet CoRoT-7b from the literature, where the refined mass estimate from this study is marked in blue, with significant improvement in uncertainty. We want to point out the reason for a few points falling in a horizontal line with our estimate being the similar value of planetary radius considered in the corresponding analyses. Figure from Anna John et al. (2022).

Table 3.5: Planetary parameters for CoRoT-7b, CoRoT-7c and CoRoT-7d obtained from the best model.

Planet	P[days]	K[m s^{-1}]	M [$M_\oplus$]
CoRoT-7b	0.8535±5.87e-7	4.291±0.46	6.056±0.653
CoRoT-7c	3.697 ± 0.005	5.757 ± 0.298	13.289 ± 0.689
CoRoT-7d	8.966 ± 1.546	5.525 ± 0.792	17.142 ± 2.552

Zeng et al., 2016, 2021) show that a substantial water layer would establish a deep steam atmosphere in a planet as strongly irradiated as CoRoT-7b, increasing the radius well beyond the observed value.

We also explored the masses of known planet CoRoT-7b for different models, as tabulated in Table 3.4. As seen from the table, the mass estimation improves and settles for the transiting + 2-planet model incorporated with GP regression and the SCALPELS decorrelation. This approach takes into account the contributions of all additional planets whose existence may not be established conclusively, but whose RV variations influence the mass determination of the known planet(s).

3.2.8 Masses of non-transiting planets

As listed in Table 3.5, we also computed the masses of non-transiting planets from the RV semi-amplitudes obtained from the analysis described in Section 2.1.2. The inner non-transiting planet CoRoT-7c, with an orbital period of 3.69 days, was found to have a mass of 13.289 ± 0.689 $M_{\oplus}$. This planet could be structurally similar to Uranus ($\sim$ 14.5 $M_{\oplus}$) in our Solar System. The RV semi-amplitude corresponding to the second non-transiting planet CoRoT-7d at 8.97 days suggests a planetary mass of 17.14 ± 2.55 $M_{\oplus}$. The mass of this long-period planet is comparable to that of Neptune ($\sim$ 17.1 $M_{\oplus}$). However, the bulk density and composition of these non-transiting planets are unknown, due to the lack of radius information from the photometry.

3.3 Summary

In this chapter, we discuss the results of applying the TWEAKS to extract precise RV estimates for the planetary candidates using the HARPS archival data of an active star, CoRoT-7.

I definitively determined the masses of three planets orbiting CoRoT-7, including the mass of the first-ever detected super-Earth, CoRoT-7b, whose mass has been a subject of debate for more than a decade. The relatively high level of magnetic activity in CoRoT-7 made this mass estimation extremely challenging, leading to a broad range of mass estimates in the literature (Queloz et al. (2009); Hatzes et al. (2010, 2011); Pont et al. (2011); Boisse et al. (2011); Ferraz et al. (2011); Haywood et al. (2014); Tuomi et al. (2014); Faria et al. (2018)). I used the same archival HARPS RV data of CoRoT-7 to disentangle the apparent RV variations caused

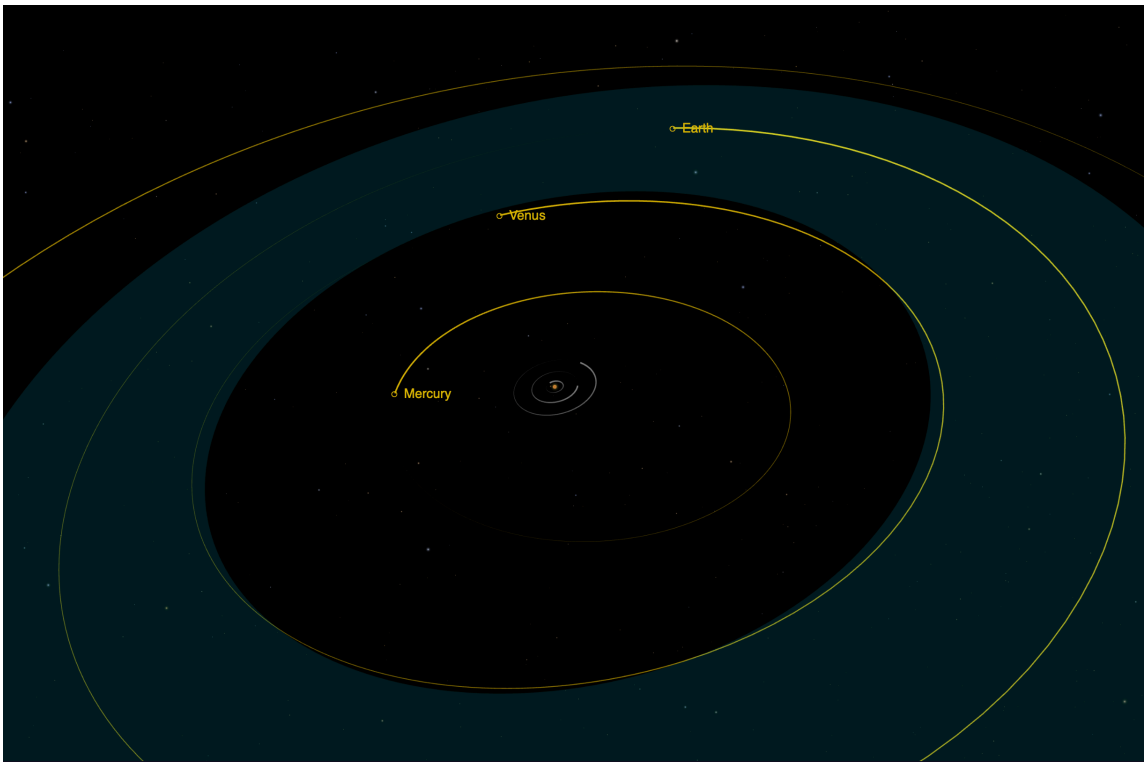


Figure 3.15: The CoRoT-7 3-planet system depicted in comparison to the inner Solar System. The three 'white' orbits represent the orbits of CoRoT-7 b,c and d. The estimated habitable zone is shown by the green annulus. Credits: NASA

by spectral line-**shape** changes (stellar activity component) from th Doppler **shifts** (planet component), as in Collier Cameron et al. (2021). I then performed a rigorous search for the Doppler shifts, using a trans-dimensional nested sampler (Faria et al., 2018).

This novel planet search approach including Bayesian model comparison allowed me to confirm the planetary status of the third planet CoRoT-7d, whose planetary origin had been mistaken for a stellar activity artefact for the past decade. Combining an efficient stellar activity model and considering the contributions of all additional (non-transiting) planets, I re-determined the masses of CoRoT-7b, CoRoT-7c, and CoRoT-7d to a precision of 11%, 6% and 15% respectively (See Figure 3.14). This allowed me to characterize CoRoT-7b as a rocky Super-Earth, with an iron core and a silicate mantle (Anna John et al., 2022). My work definitively illustrated how the detection, precise mass measurement, and characterization of exoplanetary systems are made challenging by stellar variability, and the consequently limited knowledge of the presence of additional planets in the system, particularly for the low-mass planets whose RV amplitudes are near the detection threshold.

4

HD 144579 & HD 166620: Moderate and less active stars

In this chapter, I extend the application of TWEAKS to two HARPS-N Rocky Planet Search (RPS) targets, HD 144579 & HD 166620[1]. These targets with moderate and low levels of stellar activity were carefully chosen to complement the very active star case represented by CoRoT 7.

I found no significant planet detections in either of these stars' data. This study provided valuable insights into the importance of correcting for intrinsic stellar variability while trying to achieve RV precisions of sub m s^{-1} levels demanded for the detection of Earth analogues. Achieving a detection precision comparable to the calibration precision of HARPS-N takes us one more step closer towards the detection of Earth-twins around other Suns using Extreme Precision RV (EPRV) instruments.

[1]This Chapter uses material from, and is based on, Anna John et al., 2023, Volume 525, Issue 2, October 2023, Pages 1687–1704. Parts of the investigation described in Section 4.1.4 were carried out by co-authors in the published version of the work.

4.1 Why HD 144579 & HD 166620?

In the previous chapter, we discussed how intrinsic stellar variability hampers our ability to detect planetary signals and in turn, challenges their precise mass determination. We applied TWEAKS to alleviate the extremely complex scenario presented by stellar activity in the highly active star CoRoT-7. Now, let's see how this stellar activity mitigation strategy would behave on stars with different intrinsic activity levels. We chose HD 144579 and HD 166620 from the HARPS-N RPS campaign to represent a 'moderately active star' and a 'less active star' case.

4.1.1 HARPS-N RPS program

HD 166620 and HD 144579 have been observed for more than a decade with HARPS-N (Cosentino et al., 2012), as part of the HARPS-N collaboration Rocky Planet Search (RPS) initiative. The RPS programme was initiated in 2012 to search for small planets orbiting bright, nearby stars (mainly K and early M dwarfs) when the *Kepler* field was unobservable. The major objective of the RPS program is to conduct a systematic search for low-mass planets around nearby quiet stars that are visible from the northern hemisphere via extensive monitoring of the RV, with extremely high precision (Motalebi et al., 2015). Targets for this programme are monitored 2-3 times per night (to mitigate granulation), with exposure times of 15 minutes, which is the cumulative open shutter time during multiple sub-exposures taken to prevent saturation of the brightest spectral orders. This helps to average out the p-mode oscillations (Chaplin et al., 2019), and thereby reduce the impact of stellar variability with brief typical timescales.

4.1.2 Observations

A total of 1025 observations of HD 166620 were obtained between 2013 March and 2023 March, and 888 observations of HD 144579 were taken between 2013 February and 2022 March. For the mitigation of p-mode oscillations and granulation, we use the daily binned observations in this study, which reduces the total number of RV observations for HD 166620 and HD 144579 to 293 and 256, respectively (Figure 4.1). After binning, the radial velocity RMS for HD 166620 reduced from 1.8 to 1.45 m s^{-1} and a reduction from 3.5 to 1.29 m s^{-1} was observed for HD 144579.

The HARPS-N Data Reduction Software (DRS), version 2.3.5 (Dumusque et al., 2021), was used to reduce the spectra used in this research. The DRS uses a mask (K2 & G9 spectral

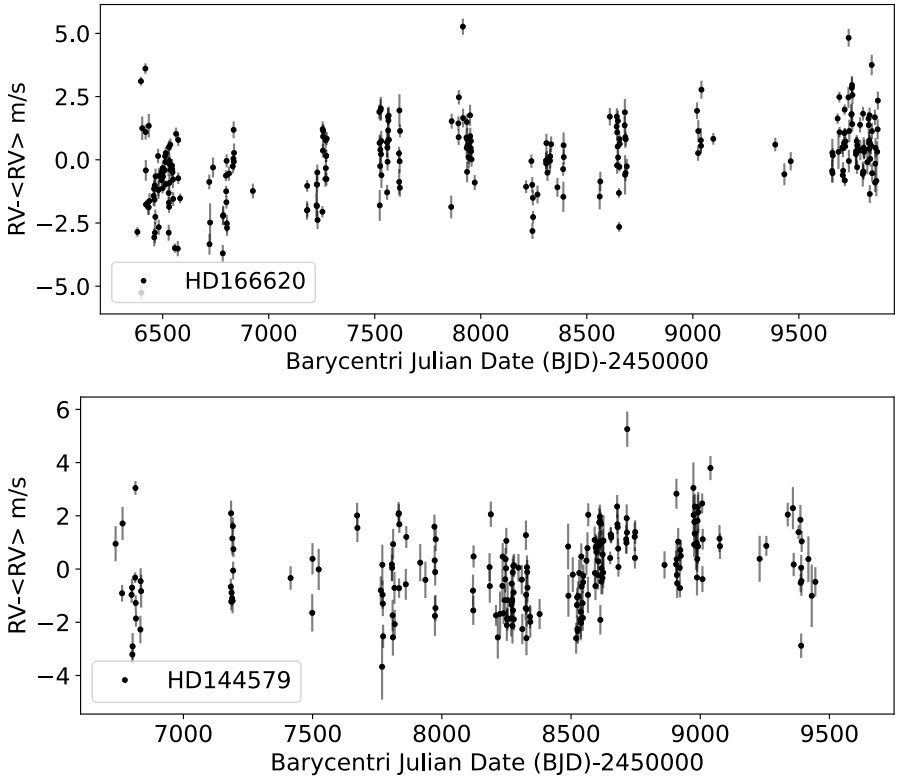


Figure 4.1: Nightly binned radial velocity observations (median-subtracted) for HD 166620 and HD 144579 measured from HARPS-N spectra are shown in the top and bottom panels, respectively. Figure from Anna John et al. (2023).

types were used for HD 166620 and HD 144579, respectively) composed of a multitude of lines that span the entire HARPS-N spectral range to calculate the Cross-Correlation Function (CCF), which is a representation of the average shape of the absorption profiles across the whole spectrum. Fitting a Gaussian to the CCFs then determines the stellar radial velocity, Full-width at Half Maximum (FWHM) and CCF Area (product of FWHM and the central line depth). In this analysis, we use the CCFs as they are obtained from the DRS. The DRS pipeline also provides supplementary information such as the spectral bisector span (BIS), Hα, Na-Index, and Mount Wilson S-index for each observation. These additional data products are widely used as stellar activity indicators.

4.1.3 HD 166620: The Maunder Minimum star

What is Maunder Minimum?

The most well-known characteristics of our Sun are its Sunspots and its magnetic activity cycle, regularly waxing and waning with a period of roughly 11 years. The 11-year Sunspot cycle was

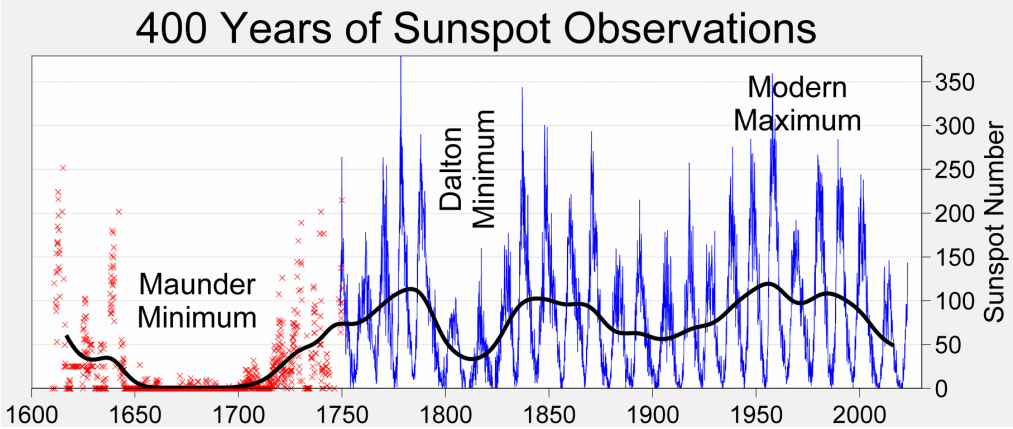


Figure 4.2: The Maunder Minimum was shown in a 400-year history of Sunspot numbers, when Sun was in a prolonged magnetic activity minimum. The Sun's journey towards a grand minimum and its transition from cyclic activity to a protracted era of flat activity is seen. This plot is created by Robert A. Rohde as part of the Global Warming Art project.

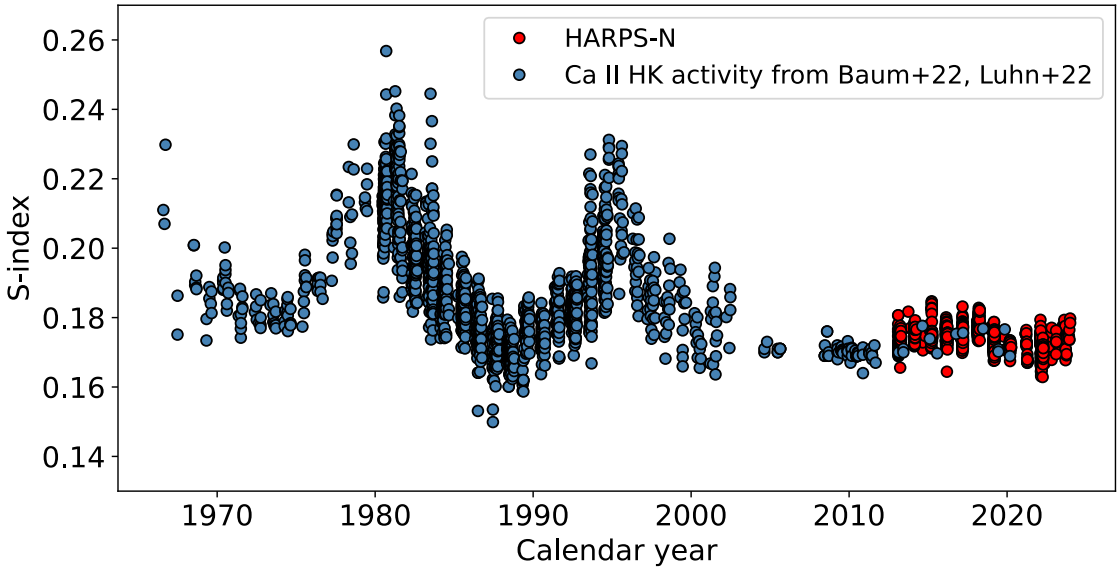


Figure 4.3: The Ca II H&K S-index data of HD 166620 from Mount Wilson and Keck-Hires, displaying the transition into the Maunder Minimum (e.g., Luhn et al., 2022). The 10 years of S-index data obtained from HARPS-N, shown by the overlapping red points reinforces the star's present Maunder minimum phase. We adjusted for the offset in the HARPS-N data with a constant shift to have a consistent median S-index value with the Mount Wilson data set. Altogether we present 57 years of stellar chromospheric activity record for HD 166620 here. Figure from Anna John et al. (2023).

discovered by Heinrich Schwabe in 1843, about 180 years ago. About 130 years later, Eddy (1976) assembled convincing historical evidence spanning centuries, to demonstrate the validity of the "prolonged Sunspot minimum" period from 1645 to 1715, which was conjectured by Gustav Spörer in 1889 and E. W. Maunder in 1894. Eddy (1976) termed this time frame when Sunspots were extremely rare, as the Maunder Minimum (MM). Less than fifty Sunspots

were observed altogether during 28 years (1672–1699) within the MM (Figure 4.2). In comparison, throughout a comparable time span, one would typically observe between 40,000 and 50,000 Sunspots. There have been active studies going on to identify a Maunder minimum analogue/candidate since the late 1900s (Soderblom, 1985; Baliunas et al., 1995; Shah et al., 2018; Baum et al., 2022).

Recently, HD 166620, a K2V type star with $V_{mag} = 6.38$ (Motalebi et al., 2015),was found to be the first Maunder Minimum analogue (Baum et al., 2022). This star has been subjected to several rigorous studies owing to its riveting photometric properties (See Section 4.1.4). The Mount Wilson program (1966–2001) (Duncan et al., 1991) studied the stellar chromospheric activity of HD 166620, leaving us with an extensive record of the stellar chromospheric activity (Baliunas et al., 1995). Baum et al. (2022) updated the stellar parameters for HD 166620 and identified this star as a possible Maunder minimum candidate based on the five decades of CaII H&K measurements from the Mount Wilson Survey, and continued observations at Keck as part of the California Planet Search programme (Wright, 2004; Isaacson & Fischer, 2010). Recently, Luhn et al. (2022) presented additional Mount Wilson data that definitively traced the transition from cyclic activity to a prolonged phase of flat activity.

Baum et al. (2022) and Luhn et al. (2022) concluded HD 166620 to be the first unambiguous Maunder-Minimum analogue, as shown by its S-index activity time series, as it changed from a cycling to a flat activity state. We reproduce Luhn et al. (2022)'s data along with the 10 years of S-index data from HARPS-N here in Figure 4.3. This 'flattened' state, which appears to have started between 2001 and 2005, is explained by Luhn et al. (2022) as a less active phase of the star with fewer residual active areas to account for the even lower variability. Luhn's S-index data and the HARPS-N S-index data show it persisting to the present (Figure 4.3). This span of low activity includes the period of time from 2012 to the present, during which the radial velocity of HD 166620 has been monitored intensively for planetary reflex-motion as part of the HARPS-N Rocky Planet Search (RPS) programme. We also have the advantage of looking for planets around this star during its magnetically quiescent phase, expecting reduced stellar variability.

The second target of this study is HD 144579. In contrast to HD 166620, no in-depth literature is presently available about this extremely bright G8V star with $V_{mag} = 6.65$. This star is slightly more active than HD 166620. Figure 4.4 shows the S-index measurements of

both stars. The stellar parameters obtained as part of this study together with the Gaia Early Data Release (Gaia Collaboration et al., 2021) for both the stars are given in Section 4.1.4 and Table 4.2.

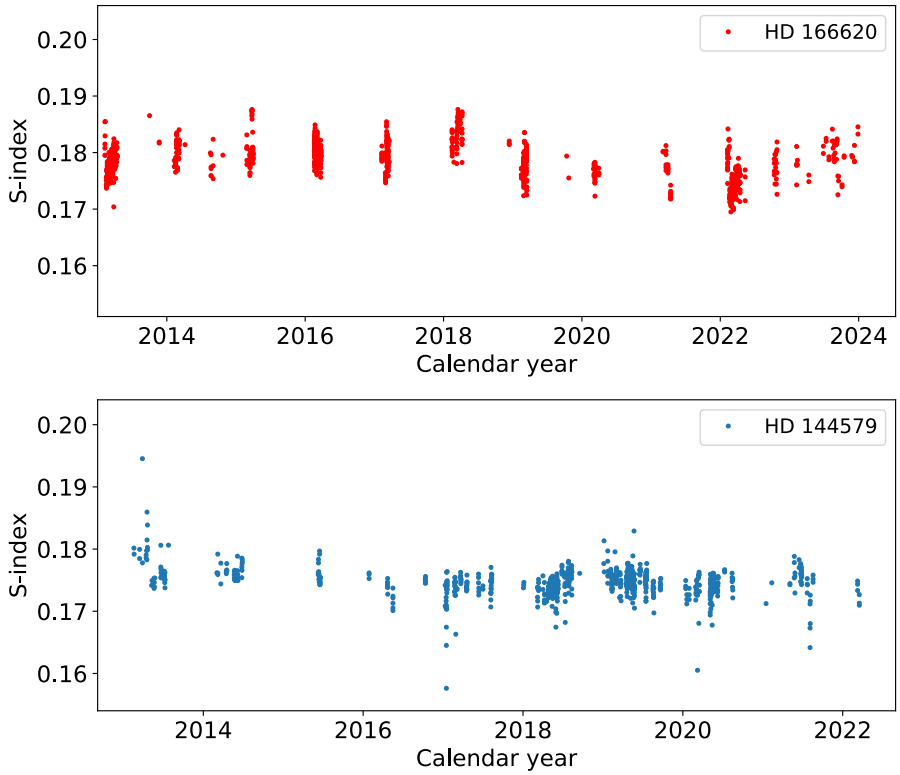


Figure 4.4: The HARPS-N S-index histories of both stars from 2012 to the present, on a common vertical scale. This shows the cycle that is present in HD 166620 and HD 144579, and provides an activity benchmark against which to compare the SCALPELS shape signals shown in Figure 4.5.

4.1.4 Stellar characteristics of HD 166620 and HD 144579

Galactic velocities for both stars were derived using the Gaia DR3 (Gaia Collaboration et al., 2021) data and in particular the trigonometric parallax, the radial velocity, the positions and the proper motions. Following Johnson & Soderblom (1987) we calculated the galactic velocities U, V, and W, as reported in Table 4.2. These are expressed in the directions of the Galactic center, Galactic rotation, and north Galactic pole, respectively, and we did not subtract the solar motion from our calculations. Using this kinematical information, we can estimate which population in the Galaxy our stars belong to. Using the descriptions in Reddy et al. (2006), we find that there is 98.5% and 90.8% probability that HD 166620 and HD 1445797, respectively, belong to the thin disc.

We utilised our HARPS-N spectra (details below) to estimate the stellar atmospheric param-

Table 4.1: Stellar parameters obtained for HD 166620 and HD 144579 from three independent methods. Our final adopted parameters, listed in Table 4.2, are the weighted average of the results from these three methods.

Parameter	HD 166620	HD 144579	Method
T_{eff} (K)	5005 ±116	5323 ±70	ARES+MOOG
	4966 ±70	5248 ±70	CCFPams
	4997 ±50	5316 ±50	SPC
log(g)(cgs)	4.82 ±0.22	4.84 ±0.12	ARES+MOOG
	4.70 ±0.20	4.56 ±0.20	CCFPams
	4.43 ±0.10	4.60 ±0.10	SPC
[Fe/H]	-0.21 ±0.06	-0.63 ±0.05	ARES+MOOG
	-0.25 ±0.05	-0.70 ±0.05	CCFPams
[m/H]	-0.19 ±0.08	-0.62 ±0.08	SPC

eters: effective temperature (T_{eff}), metallicity [Fe/H], surface gravity, microturbulent velocity and projected rotational velocity. Following Mortier et al. (2020), three independent methods were used for obtaining these parameters: (1) ARES+MOOG[2], using a stacked spectrum for each star, where the stitched spectra were simply added together after putting them in the lab-frame. Parameters were estimated using a curve-of-growth method based on neutral and ionised iron lines (Sousa, 2014). (2) CCFPams[3], using the Cross-Correlation Functions and applied an empirical relation, found in Malavolta et al. (2017). (3) Finally, spectral synthesis was applied on all individual spectra using the Stellar Parameter Classification tool (SPC - Buchhave et al., 2012).

Surface gravities are notoriously hard to measure accurately from spectra. The values resulting from ARES+MOOG and CCFPams were corrected following Mortier et al. (2014) while the value for SPC was additionally constrained using YY isochrones (Yi et al., 2001). Precision errors for effective temperature, surface gravity and metallicity from ARES+MOOG and CCFPams were inflated for accuracy by 60 K, 0.1 dex and 0.04 dex respectively following Sousa et al. (2011). Errors for the values of SPC are also accounted for accuracy (Buchhave et al., 2012). We note that ARES+MOOG is the only method to derive microturbulent velocity (ξ_t) while SPC is the only method that derived the projected rotational velocity ($v \sin i$).

The stacked spectra were additionally used to derive individual chemical abundances for magnesium, silicon and titanium. To calculate these abundances, ARES+MOOG was used with more details provided in Mortier et al. (2013). These three abundances, in combination with

[2]ARESv2: http://www.astro.up.pt/~sousasag/ares/; MOOG 2017: http://www.as.utexas.edu/~chris/moog.html

[3]https://github.com/LucaMalavolta/CCFpams

Table 4.2: Stellar parameters obtained for HD 166620 and HD 144579 from the literature and from the present study. All methods obtaining the parameters are described in Section 4.1.4.

Parameter	HD 166620	HD 144579	Reference
RA (J2000)	18:09:37.4162	16:04:56.7936	Gaia Collaboration et al. (2021)
DEC (J2000)	+38 27 27.9980	+39:09:23.4346	Gaia Collaboration et al. (2021)
μ_α (mas/yr)	-316.454 ± 0.018	-570.872 ± 0.016	Gaia Collaboration et al. (2021)
μ_δ (mas/yr)	-468.348 ± 0.020	52.633 ± 0.017	Gaia Collaboration et al. (2021)
G	6.130 ± 0.003	6.461 ± 0.003	Gaia Collaboration et al. (2021)
B	7.28 ± 0.02	7.388 ± 0.065	Wenger et al. (2000)
V	6.39 ± 0.02	6.655 ± 0.014	Wenger et al. (2000)
J	4.952 ± 0.0278	5.182 ± 0.02	Cutri et al. (2003)
H	4.458 ± 0.192	4.824 ± 0.017	Cutri et al. (2003)
K	4.232 ± 0.021	4.755 ± 0.016	Cutri et al. (2003)
W1	4.236 ± 0.279	4.733 ± 0.178	Cutri et al. (2021)
W2	3.996 ± 0.13	4.601 ± 0.083	Cutri et al. (2021)
W3	4.235 ± 0.014	4.777 ± 0.015	Cutri et al. (2021)
Spectral type	K2V	G8V	This work
Parallax (mas)	90.12 ± 0.02	69.64 ± 0.014	Gaia Collaboration et al. (2021)
Radial Velocity (km s^{-1})	-19.51 ± 0.12	-59.44 ± 0.12	Gaia Collaboration et al. (2021)
U (km s^{-1})	16.80 ± 0.05	-35.88 ± 0.04	This work
V (km s^{-1})	-31.34 ± 0.10	-58.51 ± 0.07	This work
W (km s^{-1})	0.25 ± 0.05	-18.57 ± 0.09	This work
$T_{\rm eff}$ (K)	4989 ± 48	5296 ± 37	This work
$\log(g)_{\rm spec}$ (cgs)	4.65 ± 0.10	4.67 ± 0.08	This work
[Fe/H] (dex)	-0.21 ± 0.04	-0.65 ± 0.04	This work
$[\alpha/{\rm Fe}]$	0.20	0.26	This work
vsini (km s^{-1})	< 2.0	< 2.0	This work
ξ_t (km s^{-1})	0.45 ± 0.31	0.62 ± 0.12	This work
Mass ($M_\odot$)	$0.76^{+0.032}_{-0.019}$	$0.73^{+0.022}_{-0.013}$	This work
Radius ($R_\odot$)	$0.77^{+0.007}_{-0.006}$	$0.76^{+0.005}_{-0.004}$	This work
logg (cgs)	$4.55^{+0.02}_{-0.01}$	4.53±0.01	This work

Table 4.3: Table 4.2 continued

Parameter	HD 166620	HD 144579	Reference
Density ($\rho_\odot$)	$1.67^{+0.093}_{-0.059}$	$1.63^{+0.057}_{-0.038}$	This work
Luminosity ($L_\odot$)	$0.36^{+0.02}_{-0.01}$	0.47±0.01	This work
Distance (pc)	11.09±0.002	14.36±0.002	This work
Age (Gyr)	$10.09^{+2.73}_{-3.76}$	$11.90^{+1.57}_{-2.50}$	This work
log(R' HK) (median)	-5.143	-4.976	This work
P_{rot}	42.4 d	–	Vaughan et al. (1981)
P_{cyc}	17 years	–	Baum et al. (2022)

the iron abundance, were then used to calculate the alpha-over-iron abundance, $[\alpha/\mathrm{Fe}]$. We find that HD 166620 has an alpha element overabundance of 0.20 and HD 144579 of 0.26, calculated with respect to iron as compared to the Sun. That makes these stars as alpha-enhanced as the planet hosts K2-111 (Mortier et al., 2020) or TOI-561 (Lacedelli et al., 2021) whose small planets have densities that are significantly below the densities of small planets around non-alpha-enhanced stars.

Stellar mass, radius, density, luminosity, distance, and age were further obtained from an isochrones and evolutionary tracks analysis, as described in Mortier et al. (2020). We used the three individual values for effective temperature and metallicity, the Gaia DR3 parallax, an apparent magnitude in 8 photometric filters (listed in Table 4.2), and two different sets of isochrones (Dartmouth and MIST). For each parameter, six sets of posterior distributions were obtained using the code `isochrones` (Morton, 2015) and the nested sampling algorithm `MultiNest` (Feroz et al., 2019b). All six scenarios yielded consistent results. The final adopted parameters, as listed in Table 4.2, were calculated from the median and 16th and 84th percentile of the merged posterior distributions. A more precise value for surface gravity was then calculated using the stellar mass and radius.

4.2 Analyses and results

We employed TWEAKS as an RV analysis pipeline to search for planetary signals below the 1 m s^{-1} RV barrier imposed by stellar activity (Figure 1.8) in HD 144579 and HD 166620. As we worked out for CoRoT-7 we analysed the RVs for the activity component induced by the spectral line shape changes (SCALPELS **U**-vectors). We found that the first three of the five leading principal components are the strongest contributors to the shape-driven velocity

Table 4.4: Prior distributions used in the radial velocity model in the KIMA analysis. $\mathscr{LU}$ - Log Uniform ; $\mathscr{MLU}$ - Modified Log Uniform; $\mathscr{K}$ - Kumaraswamy ; $\mathscr{U}$ - Uniform ; $\mathscr{G}$ - Gaussian.

Parameter	Notation	HD 166620	HD 144579
Number of Keplerians to fit	N_p	$\mathscr{U}(0, 5)$	$\mathscr{U}(0, 5)$
Orbital period (days)	P	$\mathscr{U}(1.1, 3492)$	$\mathscr{U}(1.1, 2700)$
RV Semi-amplitude (m s^{-1})	K	$\mathscr{MLU}(0.2, 15)$	$\mathscr{MLU}(0.2, 4)$
Eccentricity	e	$\mathscr{K}(0.867, 3.03)$	$\mathscr{K}(0.867, 3.03)$
Orbital phase	ϕ	$\mathscr{U}(0, 2\pi)$	$\mathscr{U}(0, 2\pi)$
Longitude of line of sight	ω	$\mathscr{U}(0, 2\pi)$	$\mathscr{U}(0, 2\pi)$
Decorrelation vectors (m s^{-1})	β	$\mathscr{G}(0, 10)$	$\mathscr{G}(0, 10)$
Systemic velocity (m s^{-1})	γ	$\mathscr{U}$(-2.89,5.37)	$\mathscr{U}$(-4.84,4.96)
White noise (m s^{-1})	σ_{jit}	$\mathscr{LU}$(0.01, 10 x rms)	$\mathscr{LU}$(0.01, 10 x rms)

signal (see Figures B.1 & B.2 in Appendix). The raw implementation of SCALPELS returns not just the basis vectors, but also the RV response. However, the shape time series in Figure 4.5 are from the SCALPELS-only implementation, not the coefficients of the amplitude obtained from the simultaneous fitting of any Keplerian as we did for CoRoT-7. The basis vectors (**U**-vectors) representing the shape components identified by SCALPELS are used for stellar activity decorrelation using the KIMA nested sampling package (Faria et al., 2018).

4.2.1 Prior distribution

The models we fit for the two stars are defined by the priors given in Table 4.4. These priors are comparable to those described in Faria et al. (2020), but tailored for HD 166620 and HD 144579 independently.

We set the number of Keplerians N_p to be a free parameter with a uniform prior between 0 and 5. i.e., KIMA will try to fit a maximum of 5 Keplerians to the data simultaneously and independently of each other. For this search, we use a uniform distribution for the mean anomaly of the first observation ϕ, the centre of mass RV γ, and the longitude of periastron ω, as there is no justification to favour any specific value within these parameter regimes. We use a Kumaraswamy distribution (α= 0.867 and β= 3.03) (Kumaraswamy, 1980) for planetary eccentricities, which favours lower values but still permits the exploration of higher eccentricities when the data requires it (Standing et al., 2022). The values for shape parameters α and β are justified by the Beta distribution in Kipping (2013).

The stellar reflex orbital semi-amplitude (K) is the one parameter to sample with the highest sensitivity when establishing robust detection limits (Standing et al., 2022). We will describe this analysis in Section 4.2.5. In this regard, we use a modified Log-Uniform (Jeffreys)

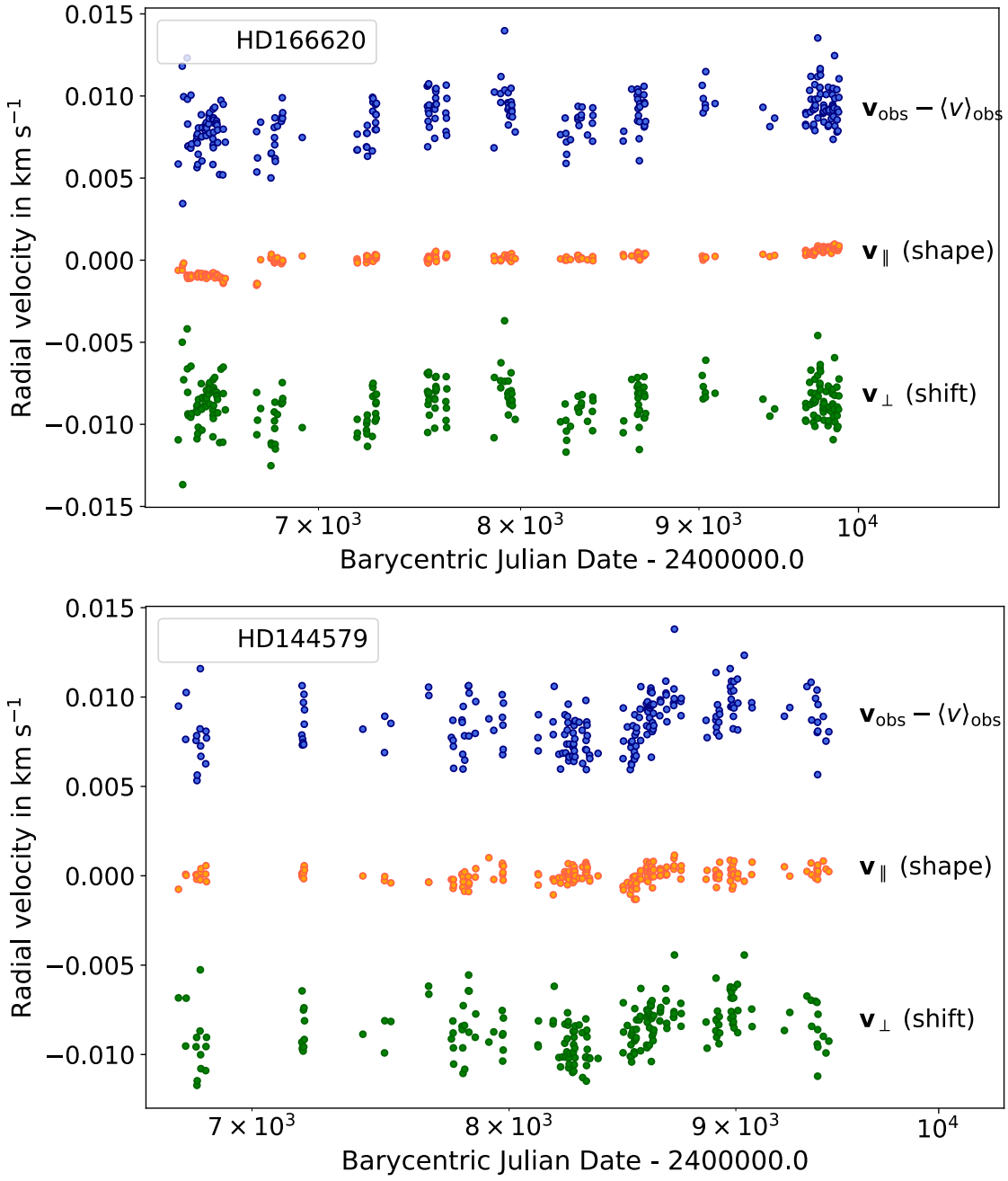


Figure 4.5: In both panels, the blue scatterplot at the top is the barycentric RV with its own mean subtracted. The orange time series in the middle represents the shape-driven component obtained from the SCALPELS projection, while the green one shows the 'cleaned' shift-driven velocities, obtained when the shape-driven variations are subtracted from the observed RVs. Please note that an offset is introduced to each RV component for better illustration. The jump in the shape time series of HD166620, at around BJD= 2456737 corresponds to the focus intervention that occurred in the HARPS-N instrument. Figure from Anna John et al. (2023).

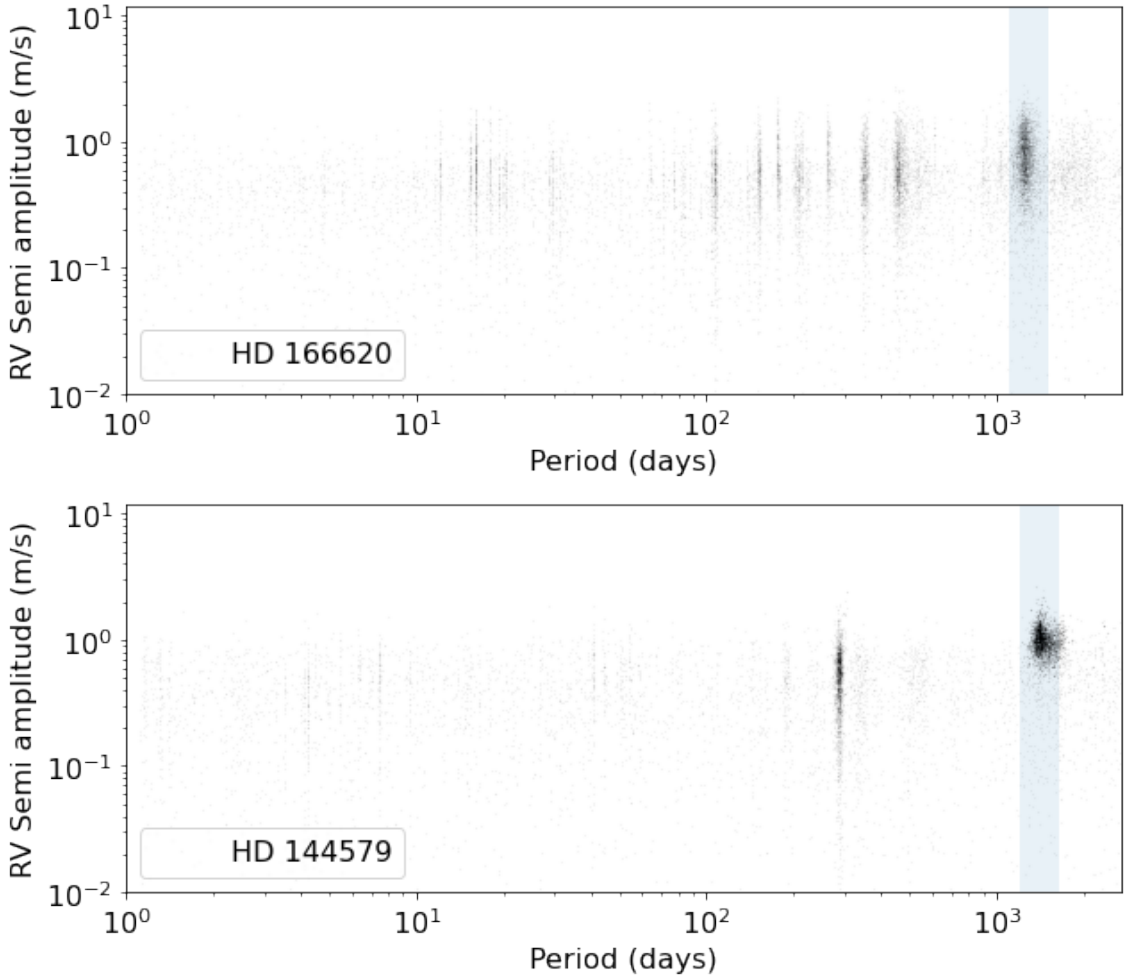


Figure 4.6: The KIMA joint posterior distribution of RV semi-amplitudes (in log scale) for HD 166620 and HD 144579 in log orbital period space are shown. Both stars exhibit an anomalous signal with a similar orbital period of 1400 days. The posterior orbital period space is populated with more over-density patches, which represent signs of additional signals, possibly planets, aliases and/or harmonics. Figure from Anna John et al. (2023).

prior. The prior values given for K in Table 4.4 are obtained by carrying out a number of initial runs with uninformative priors, to establish a plausible upper limit. The model also accounted for the systemic velocity of the centre of mass of the system (γ: which corresponds to an RV offset measured by HARPS-N) and a jitter term ($\sigma_{\rm jit}$) added in quadrature, to effectively represent the uncorrelated noise sources (uncorrelated at the timescale of the planet signals we are looking for) such as granulation and small night-to-night wavelength calibration errors. For $\sigma_{\rm jit}$, we again use a Jeffreys prior. In order to account for the stellar activity, the SCALPELS U-vectors, representing the projection of line-shape changes onto the RV are included as decorrelation vectors (as in Anna John et al. (2022) and Chapter 3), using a Gaussian prior for their amplitude coefficients.

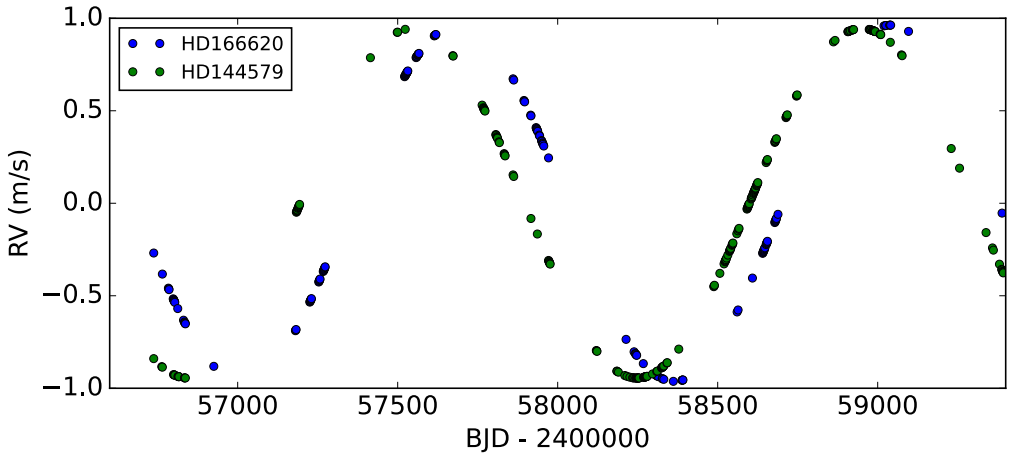


Figure 4.7: The orbits have been simulated using the best-fit parameters for the ∼ 1400d signal, observed in HD 166620 and HD 144579. The signals are nearly in phase (there is a small offset) and exhibit comparable RV variations of 95 cm s^{-1}. The phase offset possibly arises from the slightly different periods (1390 and 1420 days) of these signals. Figure from Anna John et al. (2023).

4.2.2 1400 day periodicity

To identify the candidate planet signals, we performed a similar analysis to that described for CoRoT-7 in the previous chapter. Figure 4.6 shows the posterior distribution for the RV semi-amplitudes for HD 166620 and HD 144759, plotted against the posterior orbital periods. The analyses initially revealed a long-term periodicity of approximately 1400 days in both stars. We obtained meaningful fits for 1400-day sinusoids in both stars while decorrelating against the SCALPELS shape-driven variations. We found that these signals are also more or less in phase and share very similar sub-m s^{-1} (95 cm s^{-1}) semi-amplitudes (Figure 4.7). This suggested a systematic, rather than a dynamical origin. Therefore, we investigated further the nature and possible origin of these common periodic signals.

Instrumental zero-point subtraction

We searched for a common signal with a similar orbital period (∼ 1400 d), phase, and semi-amplitude in 12 of the most intensively observed RPS targets to gain clarity on whether or not this signal is common to all the RPS targets. From parallel analyses using YARARA (Cretignier et al., 2021) and SCALPELS (Collier Cameron et al., 2021), we discovered sporadic sub-m s^{-1} jumps in the observed RV data of most of the RPS targets, in a pattern that mimics a long-term periodicity. Further investigation revealed that these discontinuities in RV occurred at the times of maintenance interventions performed on the HARPS-N instrument. These were the periodic warm-ups performed on the detector to eliminate frost build-up from water con-

tamination arising from a small leak in the detector cryostat (Dumusque et al., 2021). The detector cryostat was changed in October 2021. This solved the leak problem, but the jumps remain in observations made prior to this intervention.

Table 4.5: BJD boundaries and RV zero-point offsets calculated from the full sample including both HD 166620 and HD 144579. The 'nan' values for offsets represent the intervals where no observations were obtained.

Date of intervention	RV zeropoints
2456738.50000	-0.2805
2456826.50000	-0.9587
2456947.50000	-0.7384
2457056.50000	-1.6168
2457072.50000	nan
2457076.50000	nan
2457161.50000	-0.2981
2457308.50000	nan
2457478.50000	-0.1927
2457687.50000	0.4363
2457854.50000	-0.0994
2458071.50000	0.1848
2458231.50000	0.4039
2458412.50000	-0.2475
2458554.50000	-0.4048
2458683.50000	-0.5337
2458839.50000	0.2337
2458997.50000	0.0001
2459049.50000	0.6544
2459180.50000	0.7807
2459316.50000	1.0554
2459390.50000	-0.0803

We experimented with establishing a zero-point offset between the interventions and subtracting it from the target star RVs in order to account for these substantial instrumental interventions (cryostat warmups). A sample set of stars with long observation baselines, good seasonal sampling, and photon uncertainties of less than 1 m s^{-1} was selected for this purpose. Table B.1 in Appendix displays the list of stars (including the Sun) chosen using these criteria, the majority of which shared a comparable long-term periodicity. This enabled better quantification of the RV offsets between major interventions performed on the instrument. To prevent removing genuine signals from the data of HD 166620 and HD 144579 analysed in this paper, we removed those stars from the sample used to calculate the zero point offsets induced by detector warmups. i.e., we omitted HD 144579 from the sample set to calculate the zero points for HD 144579 and did the same for HD 166620.

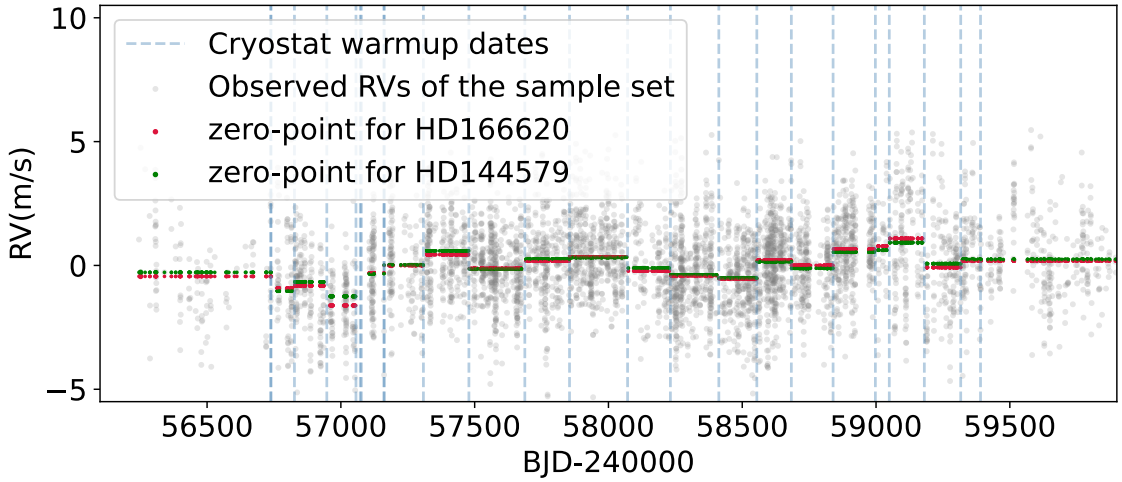


Figure 4.8: The estimated zero points in RVs between the epochs of cryostat warmups (Table 4.5) for HD 166620 and HD 144579 are shown by red and green points respectively. It is evident that when we analyse HD 166620 and HD 144579 individually, we find very similar values for the offsets. The grey points in the background represent the median-subtracted shift RVs of 11 other HARPS-N targets, corrected for activity variations using SCALPELS. The dashed grey vertical lines indicate the dates of instrumental interventions considered for the zero-point estimations. Figure from Anna John et al. (2023).

The zero-point estimation was performed as follows. The observed RVs of each star were first corrected for activity cycles by removing the SCALPELS shape component. We then removed the extreme outliers after subtracting the median RV from the activity-corrected shift RVs. We divided the data set into segments between the known dates of interventions and estimated the zero point as the median of the combined corrected RV data set in each segment. The floating chunk offset (FCO) technique, used to correct for night-to-night zero-point velocities by Hatzes (2014); Hatzes et al. (2011, 2010), served as the basis for chunk-by-chunk offset subtraction used in this study. The dates of the cryostat warmups that define the segment boundaries while estimating the zero points are represented by the vertical lines in Figure 4.8. Then, we simply subtract the estimated chunk-by-chunk zero-point (illustrated in red and green in Figure 4.8) from the observed RV data of the target stars for further investigation. The hence obtained BJD boundaries and zero-point offsets are given in Table 4.5. This finding serves as a powerful demonstration of the ability of HARPS-N to monitor such a sub-m s^{-1} level signal over ten years.

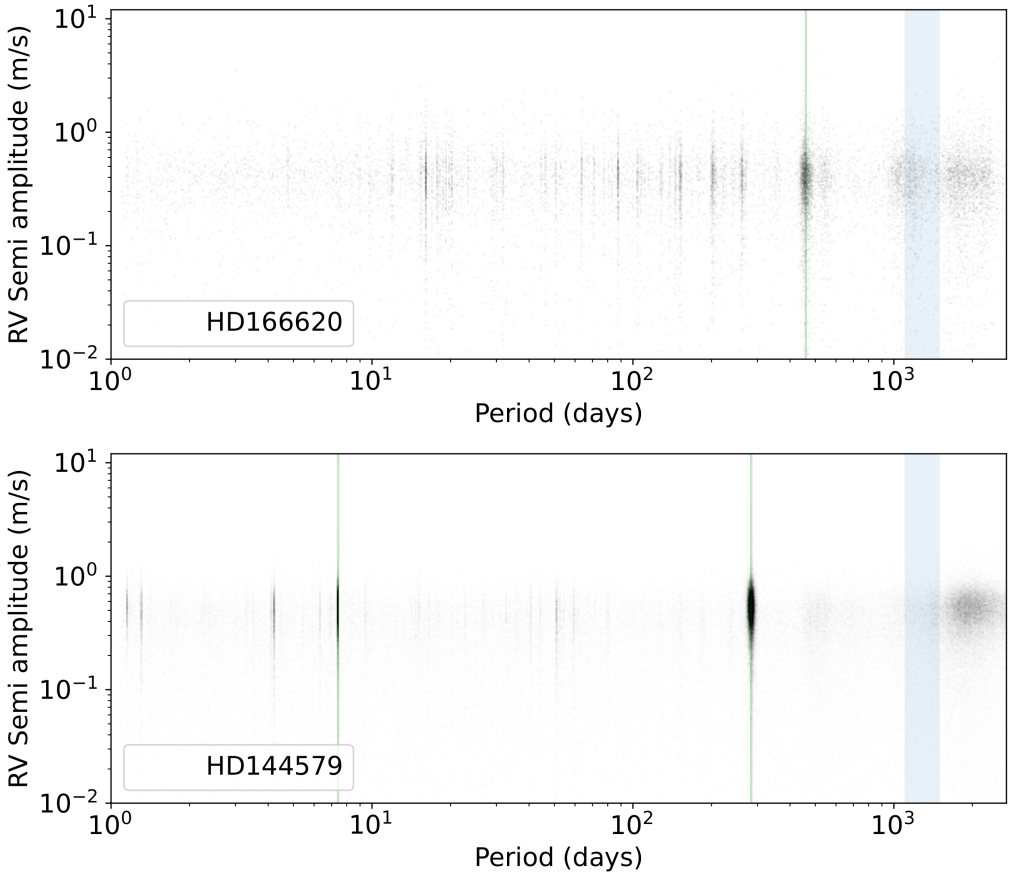


Figure 4.9: The joint posterior distributions from KIMA after subtracting the zero points estimated for the chunks of observations between major cryostat interventions, shown in Figure 4.8. The common long-term instrumental signal is shown in Figure 4.6 with orbital period ∼1400 d is hence modelled out. The displayed posteriors are obtained from more than 100000 independent trial models for HD 166620 (Top) and HD 144579 (Bottom). Figure from Anna John et al. (2023).

4.2.3 Posterior analysis

HD 166620: The tale of a 'quiet' star

Figure 4.9 displays the RV semi-amplitude posteriors from KIMA, after they have been adjusted for the long-term instrumental systematics (using the zero-point subtraction method outlined in 4.2.2). The prevalent 1400 d signal that was caused by occasional cryostat warmups is no longer dominant (compare Figures 4.6 and 4.9).

We then directed our search for planet candidates in the offset-subtracted RVs using TWEAKS (described in section 4.2.2) that takes into consideration the instrumental shifts and stellar activity cycles in both the wavelength and time domains. Using the KIMA nested-sampling package (Faria et al., 2018) and a model with up to five unknown Keplerian signals, we performed a blind search of the radial velocities. With a sum of Keplerian functions from N_p

orbiting planets, the algorithm models the RV time series while estimating the posterior distributions for all of the orbital parameters. With SCALPELS (Collier Cameron et al., 2021), which uses principal-component analysis of the autocorrelation function of the CCF, the time-domain activity-decorrelation vectors were computed. These vectors are then used as independent activity indicators for linear decorrelation in KIMA .

The top panel of Figure 4.9 displays the resultant joint posteriors for HD 166620 after the removal of the floating-chunk zero-point estimates between cryostat warm-ups and decorrelation against the SCALPELS decorrelation vectors. The posterior samples below the detection threshold appear as a background 'fog' following the priors for K and log P. A relatively well-defined signal appears at an orbital period of 460.59 d. Using a Gaussian mixture model based on Feroz et al. (2019a), we computed the probabilities of posterior samples belonging to Gaussian foreground and background populations in $\log K$ and $\log P$, to exclude the posterior points coming from the uniform prior 'fog'. This way, we isolated the foreground probability of the cluster and obtained a reliable estimation for the RV semi-amplitude for this signal as 0.41 ± 0.10 m s^{-1}. If this signal is planetary, this would correspond to a planet with an upper $M \sin i$ mass limit of 3.34 ± 0.85 $M_\oplus$, when taking into account a stellar mass of 0.76 $M_\odot$. This method increases the precision and accuracy of K compared to the standard measurement of the posterior in KIMA.

We conducted a False Inclusion Probability (FIP) analysis (Hara et al., 2022b) in frequency space, with the bin size set to the Nyquist frequency resolution over the entire data duration. The Gaussian mixture model allowed us to calculate the probability of posteriors in the foreground, independently of the choice of frequency bin width. We first computed the True Inclusion Probability (TIP), as the fraction of all KIMA trial models (N_s) that contain planets whose periods fall in the frequency interval of interest, belonging to the aforeground. In other words, TIP is the number of models for which the orbital period (P) and RV semi-amplitude (K) of a sampled planet fall within their respective foreground populations.

$$\mathbf{TIP} = \frac{\sum_{i=1}^{N_s} \mathrm{Pr_i(foreground)}}{N_s} \tag{4.1}$$

The FIP was then obtained as:

$$\mathbf{FIP} = 1 - \mathbf{TIP} \tag{4.2}$$

Hence, we simultaneously searched for multiple planets using a frequency window sliding

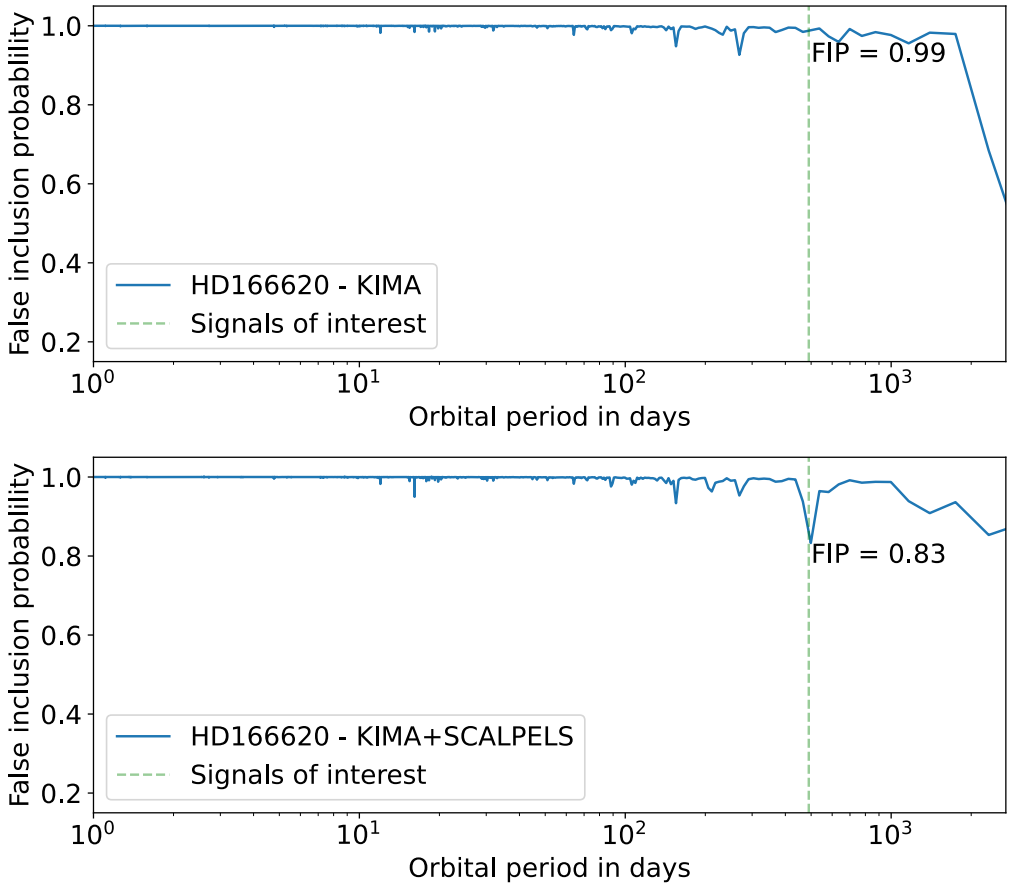


Figure 4.10: False Inclusion Probability (FIP) periodogram of HD 166620 showing the minimal values at orbital periods of potential detections. Top : FIP when no SCALPELS U-vectors are used for activity decorrelation in KIMA. Bottom: A more significant FIP value is observed at the same orbital periods when SCALPELS is included. Figure from Anna John et al. (2023).

across the full frequency range covered by the posterior distribution. A FIP value of 0.83 (Figure 4.10) was found at an orbital period of 460.59 d. This is not what we anticipate from a strong planetary signal detection, particularly for a star that is magnetically quiet. A FIP value of 0.83 indicates that only 17% of the models tested favoured a detection at this orbital period interval. To validate this, we performed several analyses by shuffling the observing seasons between years and found that the long-term signal at 460.59 d is comparable in its amplitude and its FIP to window-function artefacts appearing at other periods in the permuted datasets. We conclude that it too is most likely to be a window-function artefact.

HD 144579: A moderately active 'lone' star

Unlike in HD 166620, the joint posteriors from the HD 144579 RVs adjusted for long-term systematics show strong signals at 7.39 and 284.13 d (bottom panel of Figure 4.9) with RV semi-amplitudes of 0.50 ± 0.11 and 0.60 ± 0.07 m s^{-1}. If these signals are planetary, they

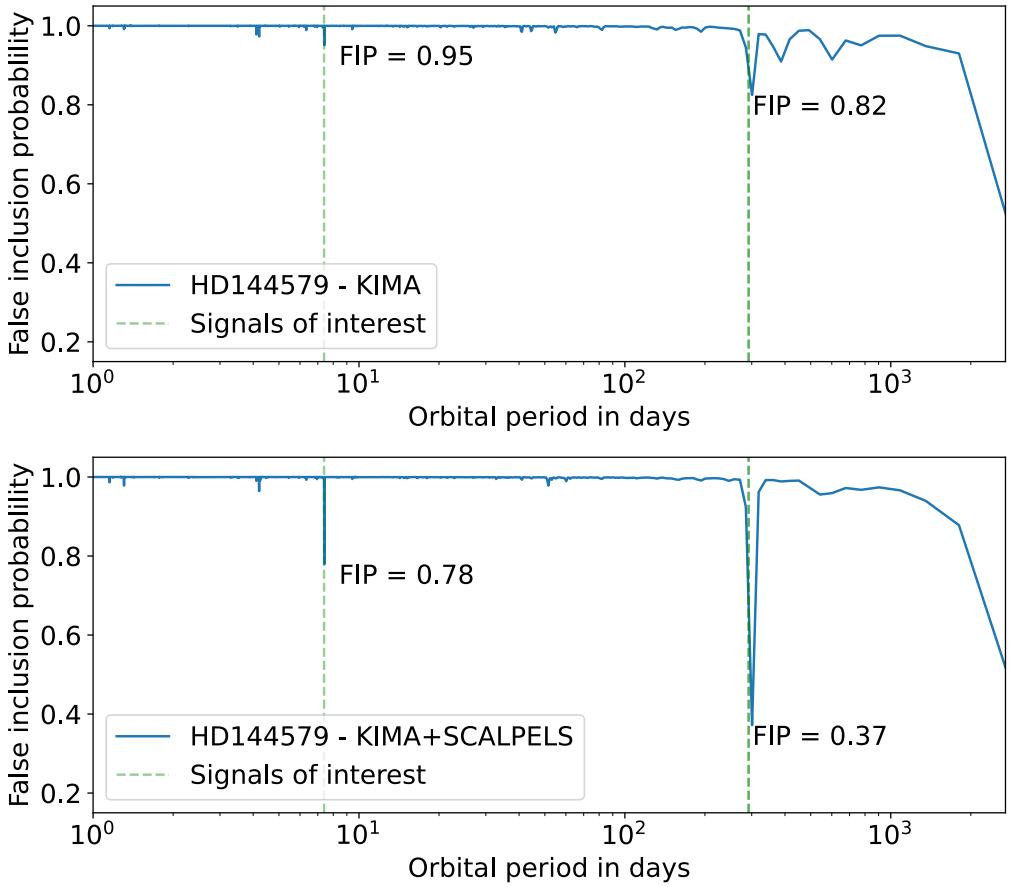


Figure 4.11: False Inclusion Probability (FIP) for HD 144579 before and after stellar activity decorrelation. The top and bottom panels are as in Figure 4.10. Here, the FIP periodogram of HD 144579 shows clearer detection of the signals of interest, when decorrelated against the SCALPELS U-vectors. Figure from Anna John et al. (2023).

would correspond to planets with $M \sin i$ of 1.28 ± 0.29 $M_\oplus$ and 4.14 ± 0.52 $M_\oplus$, respectively, from the foreground posterior probability contained using a Gaussian mixture model. The obtained results favoured $\geq 4.5\sigma$ detections for both signals. However, based on the ratio of probabilities of consecutive values of the number of planets iterated in the KIMA model comparison, the sampler favoured zero planets (based on the 150 Bayes Factor threshold), as shown in Figure 4.13.

The False Inclusion Probabilities from the posteriors of HD 144579 are shown in Figure 4.11. In contrast to HD 166620, substantially lower FIPs are observed in HD 144579 at orbital periods of 7.39 and 284.13 d. Although there is a notable decrease in the FIP when SCALPELS decorrelation vectors are included in the model, the values are still not conclusive. Using the 284.13 d signal as an example, reporting a detection based on a FIP value of 0.37 (Figure 4.11) is analogous to boarding an aeroplane with only a 63% chance of making it to the destination.

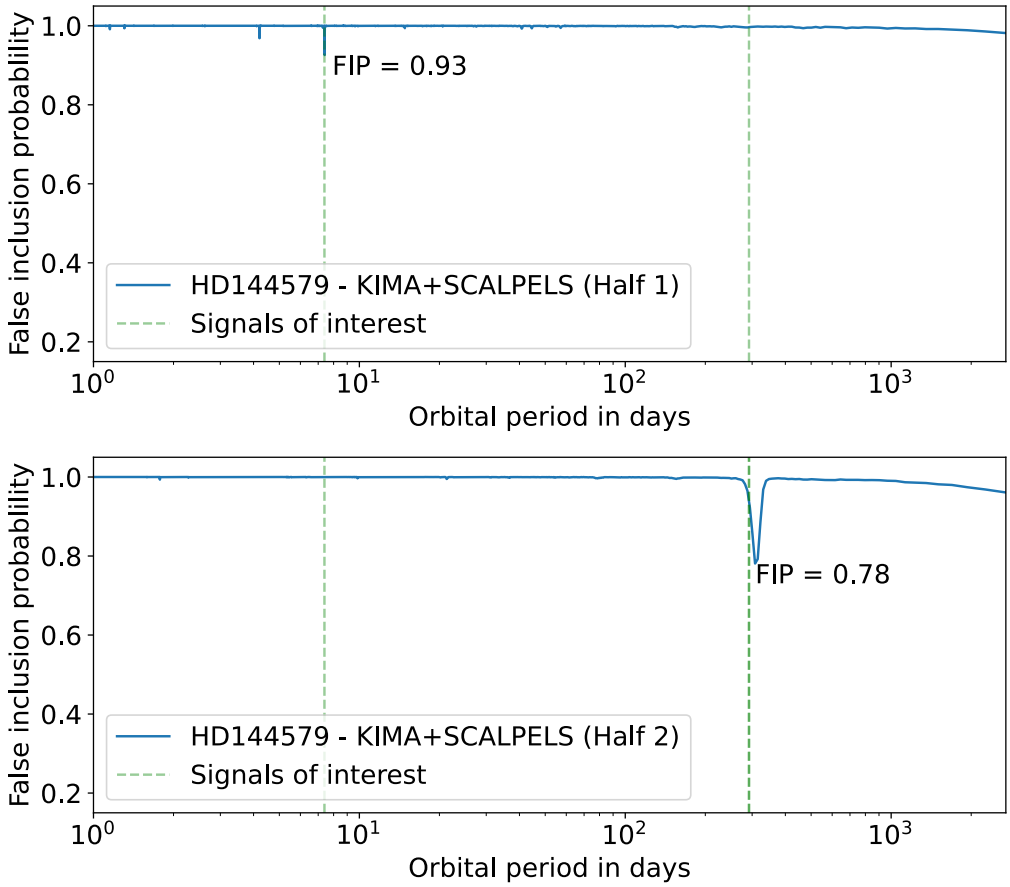


Figure 4.12: False Inclusion Probability (FIP) for HD 144579 before and after stellar activity decorrelation. The observations are now divided into two subsets to investigate the coherency of the signals of interest that appeared in the bottom panel of Figure 4.11. The FIP diagrams obtained for individual half subsets of the HD 144579 data are shown in the two panels. Each data subset covered an observation baseline of > 1500 days. Unlike any coherent signal, none of the signals of our interest was commonly detected in both data subsets. Hence, each of these signals could be an artefact of the sampling pattern arising from individual seasons. Figure from Anna John et al. (2023).

Moreover, several ambiguous detections can be triggered at spurious periods because of the sampling pattern and cross-talk between various aliases (e.g., Anna John et al., 2022). To guard against this possibility, we analysed the system further by splitting the data train. The primary goal was to examine the individual data halves for any candidate signals. If the baseline covered by any individual subset is appropriate for sampling the signals of interest, then any coherent signal should be detectable independently in n number of data subsets, while the significance of detection depends on the size of the subset. There is a caveat, though. The stellar activity can probably manifest differently in the subsets and the ability of TWEAKS to completely clean the time series could vary, leading to stronger signals in some cases, and weaker in others.

Figure 4.12 shows the FIPs obtained from the posteriors of individual half subsets. Con-

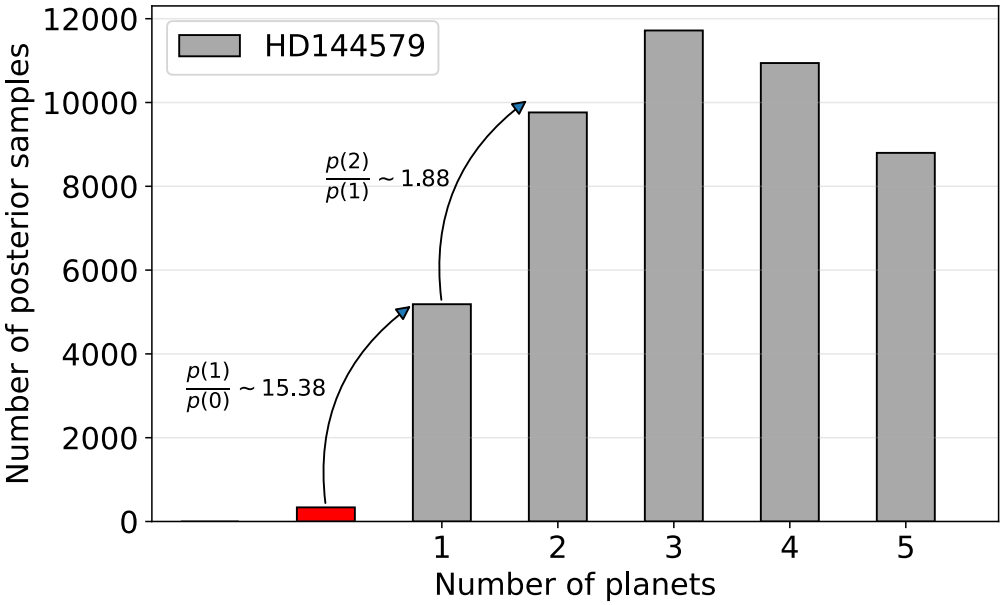


Figure 4.13: The posterior distribution for the number of planets N_p. The counts are the number of posterior samples in trial models with a specific number of planets. The probability ratios between models with 0,1 and 2 planets are shown. The posterior distribution suggests (Figure 4.9) that there are likely to be planets present. The 'confident detection' criterion adapted by KIMA, however, takes into account the ratio of probabilities for successive values of N_p, which favours N_p=0, rather than depending solely on the probability values. The preferred model (N_p=0) appears underrepresented, because of the significantly larger parameter space for models with successively greater N_p. On the other hand, when a Keplerian was injected, KIMA favoured N_p=1 (Figure B.4). A similar representation for HD 166620 is shown in Fig:B.3.. Figure from Anna John et al. (2023).

trary to what would be expected from any coherent planetary signal, neither of the signals of interest were consistently found in both of the two data subsets. Each of these signals could therefore be an artefact of the sampling pattern resulting from a particular localised subset of the observations or the existing data may not be adequate to confirm these detections.

To validate this finding, we also performed some injection and recovery tests by injecting a Keplerian with an orbital period of 210.28 d and an RV semi-amplitude of 60 cm s^{-1} (comparable to the semi-amplitude of the 284 d signal). We then repeated the analysis and successfully recovered the injected Keplerian as independent marginal detections in individual half-data subsets (Figure B.5). An injected signal with a slightly larger semi-amplitude (80 cm s^{-1}) was consistently better detected in the individual subsets as shown in Figure B.6. We also experimented with injecting a short-period signal with P = 5.12 d which is closer to the 7.39 d signal and with an RV semi-amplitude K = 60 cm s^{-1}. Unlike the long-period injected signal, this signal was only recovered in one of the half sets (Figure B.7). On the other hand an 80 cm s^{-1} injected Keplerian was successfully recovered in individual subsets (Figure B.8). This suggests that K = 60 cm s^{-1} is below the detection threshold for the data partitioning test at P = 5.12

Table 4.6: Derived posterior values for Signals of Interest (SOIs) from the KIMA+SCALPELS RV analysis detailed in Section 4.2.3. The FIP values given are obtained from the entire data. $M_p \sin i$ values are estimated considering if the SOIs are planetary.

Parameter (unit)	HD 166620 SOI.1	HD 144579 SOI.1	HD 144579 SOI.2
Derived parameters from KIMA *posteriors*			
P (d)	459.88 ±11.05	7.39 ±0.002	284.13 ±4.16
K (m s^{-1})	0.41 ±0.10	0.50 ±0.11	0.60 ±0.07
$M_p \sin i$ ($M_\oplus$)	3.34 ±0.85	1.28 ±0.29	4.14 ±0.52
e	0.22	0.14	0.18
ω (deg)	2.97	3.48	3.01
T_0 (BJD-2400000)	57491.32641	58514.54291	58243.86417
FIP	0.83	0.78	0.37

d. By the same token, it also suggests that the 7.39 d signal might not be entirely spurious. However, without confirmation in independent subsets of the data, we cannot be certain.

This provides a pragmatic estimate of the detection limit in the HARPS-N data to be somewhere between 60 and 80 cm s^{-1}, which is consistent with the formal estimation of detection threshold presented in Section 4.2.5. In addition, these simulations show that the data in hand are not sufficient enough to conclude the origin of the 7.39-day and 280-day signals.

4.2.4 Impact of stellar activity mitigation using SCALPELS, in the model selection

The difference in evidence (a.k.a Bayes factor) between a model with decorrelation performed against stellar activity using SCALPELS U-vectors and a model uncorrected for stellar activity is given in Table 4.7. The shape component comprising the corresponding U-vectors for individual stars are shown in orange colour in both panels of Figure 4.5. As anticipated for a star in a Maunder-minimum state with no identifiable magnetic activity cycles, the shape component obtained for HD 166620 looks essentially flat. However, SCALPELS effectively deals with the focus intervention that occurred in the HARPS-N instrument, which is seen as an evident jump in the shape time series, at around BJD= 2456737. This intervention was planned as the FWHM in all stars observed by HARPS-N was showing a significant drift, likely due to a change of instrument focus with time. A component was removed that was supposed to stabilize the focus over time but was not performing as expected. This fixed the FWHM drift issue, however, a significant jump in FWHM, likely due to a PSF change, was observed after this intervention. This jump has been efficiently tracked by some of the SCALPELS basis vectors (Figures B.1 & B.2 in Appendix).

Table 4.7: Bayesian evidence for models with and without SCALPELS vectors used for stellar activity decorrelation.

Star ID	KIMA ($\log\mathscr{Z}_1$)	KIMA + SCALPELS ($\log\mathscr{Z}_2$)	Bayes Factor $\mathscr{Z}_2/\mathscr{Z}_1$
HD 166620	-457.29	-453.65	38.09
HD 144579	-362.94	-357.56	217.02

As is evident from the shape component in the lower panel of Figure 4.5, SCALPELS identified a greater amplitude of shape-driven RV variation in HD 144579 than in HD 166620, owing to its greater intrinsic activity level. If we now take a look at Table 4.7, we can see how the Bayes Factor reflects the impact of shape corrections on each star. The Bayes factor provides a measure of how strongly one model is supported over the other by the data. We use the same Jeffreys' scale as a benchmark for our model evaluation. In accordance with Trotta (2008), we defined the threshold for "moderate" evidence at Bayes Factor = 12 and for "strong" evidence at Bayes Factor = 150. While SCALPELS decorrelation improves the model by a factor of 38 for HD 166620, HD 144579 shows a significant improvement of 217 times. In both cases, this shows that the use of CCF-shape decorrelation improves, rather than degrades, the marginal likelihood of the fit.

4.2.5 Detection limits

A distinct characteristic of a trans-dimensional nested sampler, such as KIMA, is that it will create a map of all signals that are consistent with the data when forced to explore higher N_p than is explicitly detected. Since those suggested signals are still technically undetected, this posterior can be used to generate a detection threshold (Standing et al., 2022). According to Standing et al. (2022), the detection limits obtained with KIMA marginalize over a greater range of orbital parameters than can be computationally achieved with the standard injection-recovery tests, and KIMA assumes Keplerian signals, rather than sinusoidals.

To establish a detection limit on additional undetected planetary signals in the radial-velocity data of HD 166620 and HD 144579 using KIMA, we adhere to the process described in Standing et al. (2022), where they use the Bayesian approach with KIMA to produce detection limits to answer the question "what is compatible with the data?". A robust detection limit is usually obtained after the removal of any candidate Keplerians present in the data. Here, we

can directly compute the detection thresholds from the posteriors, since no valid detection can be established using the data presented. We performed KIMA using the priors shown in Table 4.4 to establish the detection limits, with the exception that N_p was fixed to 1. To exclude Keplerians that would have otherwise been discovered, KIMA thereby marginalises over the whole permitted parameter space while looking for any potential Keplerian signals that would have produced a reasonable likelihood. In principle, this approach is analogous to forcing the sampler to fit an additional signal to the data (Standing et al., 2022, 2023; Sairam & Triaud, 2022).

As a well-sampled posterior is necessary for generating a reliable detection limit, the number of saves in KIMA was set to 200000 to obtain a minimum of 30000 effective posterior models. The hence-obtained posterior is divided into log-spaced bins in the orbital period (P) to calculate the detection limit. Within each bin, the 95th percentile of the semi-amplitude (K) distribution is then evaluated. The top envelope of the resultant posterior's density will then represent the detection limit. The findings of our analysis of the detection limits for the HD 166620 and HD 144579 are shown in the top and bottom panels of Figure 4.14. The grey scale hexbins display the posterior sample density from the KIMA + SCALPELS trials on each target. The red lines enveloping the entire posterior space represent the calculated detection limit from a single run with decorrelation performed against the SCALPELS U-vectors. The no-SCALPELS version is used to compute the faded blue detection limit line.

Figures 4.10 and 4.11 depict the effect of SCALPELS correction in HD 166620 and HD 144579. While the window function artefact at P∼460 d in HD 166620 appears well below the detection threshold, the corresponding signals in HD 144579 at 7.39 and 284 d obstruct the improvement in the detection limit despite being decorrelated against SCALPELS U-vectors.

4.2.6 Probing RV signals in the sub-m/s regime

Doing the stellar activity decorrelation in the wavelength domain using SCALPELS also has an evident impact in bringing down the detection limit more towards the sub m s^{-1} regime. In Figure 4.14, the detection limits calculated with and without SCALPELS decorrelation are shown. Although the impact varies slightly across various period regimes, SCALPELS decorrelation allows for better probing in the low-mass regime in both stars. This is in line with a recent study by Standing et al. (2022), in which they show how the activity cycle affects the detection efficiency of planets in the solar case. They found that the likelihood of finding a

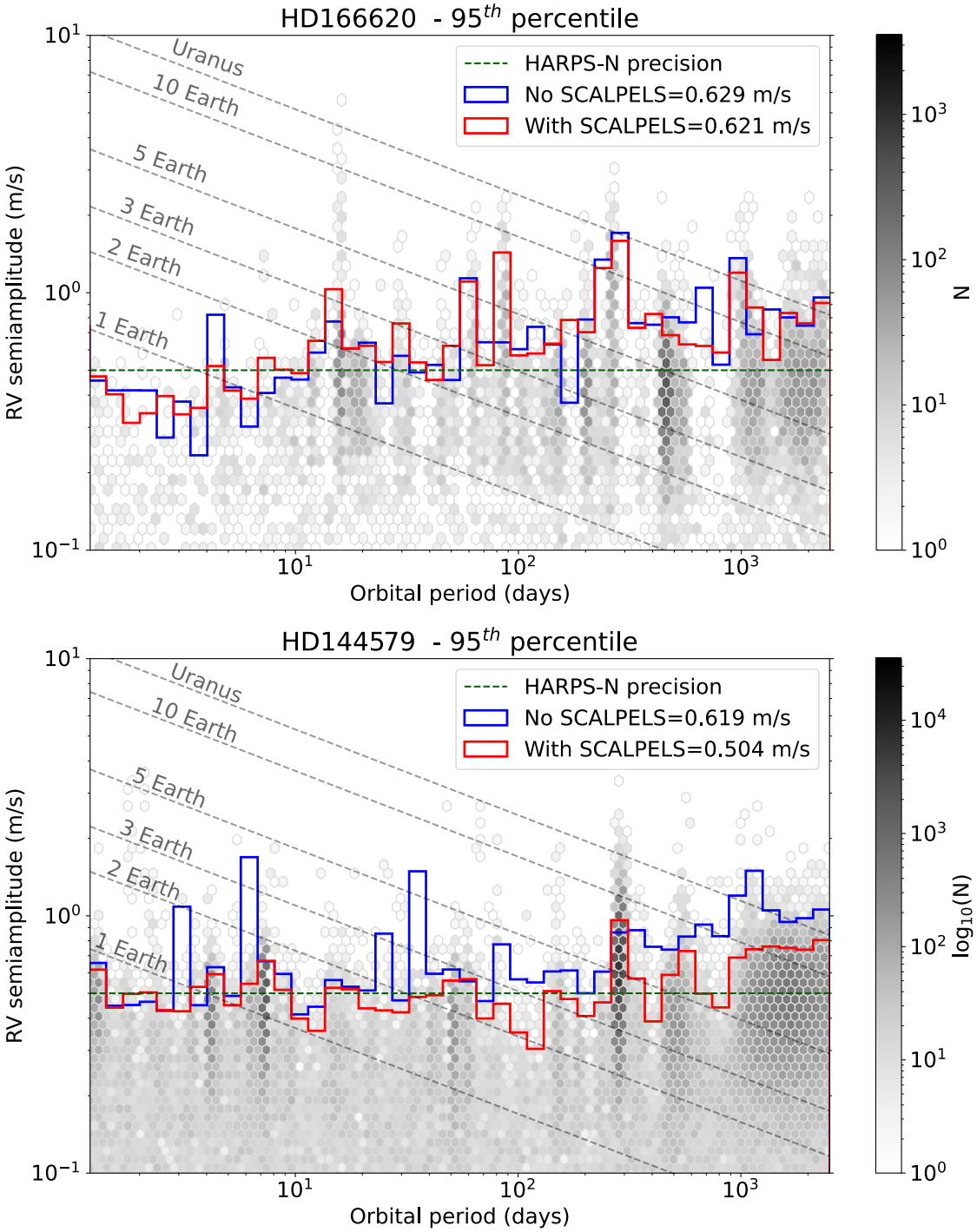


Figure 4.14: A greyscale hexbin plot representing the density of posterior samples derived from the KIMA+SCALPELS runs. Faded blue lines represent detection limits derived from a KIMA run without any decorrelation against the scalpels basis vectors. The solid red line shows the detection limit computed from a KIMA run corrected for stellar activity with SCALPELS. The grey dashed lines are contours of constant planet mass ($M \sin i$) for comparison. Figure from Anna John et al. (2023).

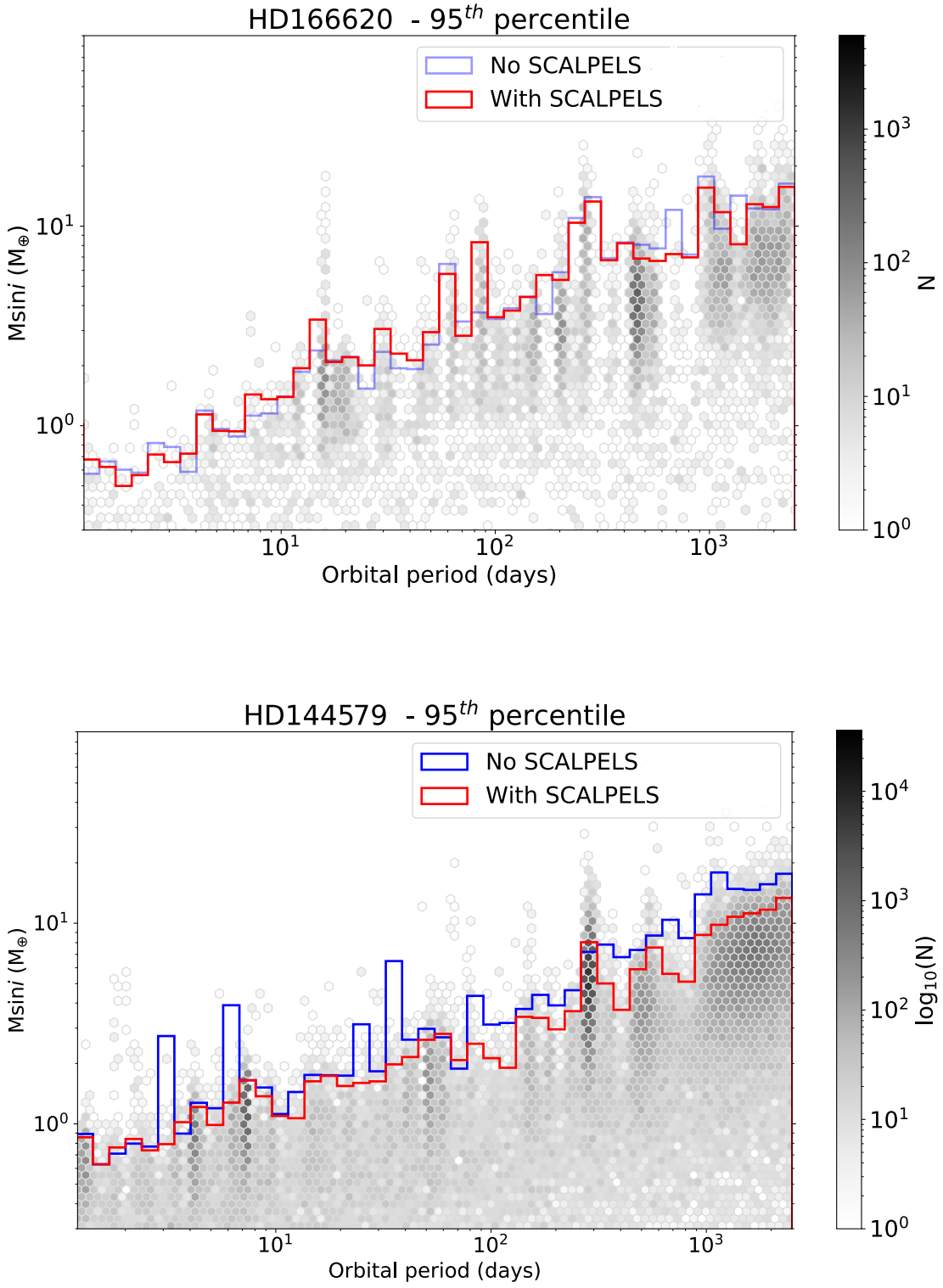


Figure 4.15: Upper detection limits in the $M \sin i$ versus orbital period space. The blue and red stepped histograms respectively show the 95$^{\text{th}}$ percentile detection limit before and after decorrelating against SCALPELS **U**-vectors. Figure from Anna John et al. (2023).

low-mass planet falls by an order of magnitude during the solar activity maximum. Therefore, they suggest a more efficient observational strategy based on a forecast of stellar activity that would enhance the identification of Earth counterparts and their analysis. Our study shows that we could detect smaller planets, when we correct for the stellar activity and instrumental shifts using SCALPELS, despite the activity state of the target star. However, the level of magnetic activity in the individual stars is reflected in the improvement in the ability to detect low-mass planets on the application of adequate stellar activity correction.

While the detection limit does not show significant enhancement in HD 166620, the mean detectability in HD 144579 improves from 0.62 to 0.54 m s^{-1} after correcting for the spectral-line shape changes. The most striking improvement is observed in the long-period regime, where the detection threshold consistently improves at periods longer than 300 days, reducing the detection threshold from $\sim$ 1.2 m s^{-1} to $\sim$ 0.6 m s^{-1} at periods around 1000 days. In the era of upcoming missions like PLATO and HARPS3, striving for Earth-like planets in an Earth-like orbit around a Sun-like star, these findings are of high importance. It is evident from Figure 4.14 that we are sensitive to a wide variety of Super-Earths (5-10$M_\oplus$) in the entire orbital period range ($P < 2800$ d) spanned by the posterior space. Moreover, we can detect planets with masses as low as 1-3$M_\oplus$($P < 300$ d), and more massive planets with 5$M_\oplus$up to $P = 1000$ d. The mean detection threshold computed after correcting for stellar activity falls around 60 cm s^{-1} in HD 166620 and 54 cm s^{-1} in HD 144579, approaching the HARPS-N precision limit imposed by wavelength solution (Dumusque et al., 2021), which, without considering instrument intervention, should provide a similar long-term calibration precision for HARPS-N.

Figure 4.15 displays the same detection limit figure translated into a $M \sin i$-orbital period space. It is now easily readable that under $P = 10$ d, we are able to detect significantly lower-mass planets in HD 166620 and HD 144579 with masses smaller than Earth.

4.3 Summary

We performed a thorough investigation of the HARPS-N spectroscopic data of bright K and G dwarfs HD 166620 and HD 144579, neither of which is known to host a planetary companion. We examined these stellar systems using TWEAKS. The major objective was to search for planetary reflex-motion and investigate the impact of stellar activity mitigation on the capability

of detecting low-mass planets, in stars with different levels of magnetic activity. We found no significant detections in either of the stars. We also ruled out the possibility of being misled by erroneous signals from sampling patterns by performing data-splitting exercises and validating using injection-recovery tests. The injection recovery tests also showed that the data in hand might be sufficient to conclude the origin of the 7.39-day and 280-day signals. The data partitioning test is something we would advise as a standard procedure to validate the coherency of a signal before making any detections public. More rigorous coherency tests have also been proposed by Gregory (2016) and Hara et al. (2022b).

Additionally, we provide an estimate for the detection limits on the HARPS-N radial velocity data using posteriors from the KIMA diffusive nested sampler. We demonstrate the varied impact of stellar activity on the detection efficiency of planets in HD 166620 which is in the Maunder minimum state and HD 144579 which is moderately active. The likelihood of finding a low-mass planet increases noticeably across a wide period range when the inherent star variability is de-trended against SCALPELS **U**-vectors. The 54 cm s^{-1} detection threshold achieved based on the aforementioned decorrelations brings us closer to the known calibration precision value provided by the HARPS-N instrument (50 cm s^{-1}) (Dumusque et al., 2021). These detection limits were further verified using injection-recovery tests.

The precise detection and mass determination made possible by TWEAKS has greatly aided in the precise characterization and bulk density measurements of multiple TESS, HARPS, HARPS-N, RPS, and K2 targets also. These contributions will be discussed in the next chapter.

5

Application of TWEAKS to *TESS*, *Kepler* and RPS targets

Improving the precision of RV measurements is of paramount importance for enhancing the scientific reliability of follow-up efforts for planetary systems discovered by transit surveys such as TESS, K2 and upcoming missions like HWO and PLATO. This chapter[1] investigates TWEAKS' ability to improve RV detections in a regime between the two extremes represented by the activity-dominated CoRoT-7 system (Chapter 3) and the low-activity stars discussed in Chapter 4. The systems discussed here are in an intermediate regime where the RV amplitudes of super-Earths and mini-Neptunes, and the measurement bias caused by activity-driven CCF-shape changes, are comparable and above the detection threshold. As in the case of CoRoT-7, I found that activity de-trending and allowance for other planet signals are important for minimising measurement bias and evaluating realistic uncertainties on planet mass.

[1]This Chapter uses material from, and is based on my contributions to Palethorpe et al. (2024), Wilson et al in review, Mantovan et al. (2023), Dalal et al in prep and Stalport et al. (2023).

In this chapter, I present my contributions to several recent collaborative publications characterising multiple TESS, *Kepler* and HARPS-N RPS targets using TWEAKS. This includes planetary systems: TOI 1778, TOI 1730, TOI 5398, HD 48948 and HD 99492. All the TOIs and K2 systems were first discovered via the transit method, by the *TESS* and/or Kepler2 missions. They were then followed up with HARPS-N to obtain their precise masses. HD 48948 and HD 99492 were observed by HARPS-N as part of the Rocky Planet Search program; neither has any known transiting planets. I used the basic TWEAKS pipeline to mitigate the stellar activity impact in all the systems to detect the planetary radial velocity signals and measure their masses.

5.1 TOI 1778

TOI 1778 or HD 77946 or TIC39699648 is an active star with a mini-Neptune-sized planet crossing its disc every 6.52 days. An RV validation study led by Larissa Palethorpe was ongoing when I finished setting up the TWEAKS workflow. The RV analysis of TOI 1778 was made complicated by the intrinsic stellar activity, as it imposed uncertainty in the detection and precise mass measurement of the planet whose orbital period is comparable to the first harmonic of the stellar rotation period. So, I participated in the analysis of the radial velocity data, aiming to provide an independent improvement of the detection significance and mass estimation by mitigating the effects of stellar activity. The final paper, Palethorpe et al. (2024), was recently accepted for publication in MNRAS.

5.1.1 History and observations

TOI 1778 was first observed in Transiting Exoplanet Survey Satellite (TESS) sector 21 and a transiting planetary signal was detected with an orbital period of 6.52 days. According to Palethorpe et al. (2024), this candidate planetary signal was initially detected by the SPOC pipeline (Science Processing Operations Center; Jenkins et al. (2016)), which then got a TOI (*TESS* Object of Interest) title after being verified and vetted by TEC (TESS-ExoClass [2]) and *TESS* team, respectively. The detection was then confirmed in photometry from *TESS* sector 47. *CHEOPS* also visited the system twice and observed the planet transiting. The whole log of observations obtained as part of this *CHEOPS* AO-2 Guest Observers (GO) programme (ID:07, PI: Mortier) is given in Table 1 of Palethorpe et al. (2024). According to Palethorpe et al.

[2]https://github.com/christopherburke/TESS-ExoClass

(2024), this sub-Neptune's estimated radius of 2.7 ± 0.08 $R_\oplus$places it above the radius-valley (Fulton et al., 2017), which is the gap (1.5 - 2 $R_\oplus$) in the orbital period - radius space that divides sub-Neptunes from super-Earths.

For measuring the planetary mass, HARPS-N followed up on this target between December 2020 and January 2023 (BJD 2459190.74 and BJD 2459954.68), as part of the Guaranteed Time Observations (GTO) programme, and collected 102 spectra. I analysed the 102 CCFs associated with these spectra using TWEAKS, to confirm the transit detection, measure the planet candidate's precise mass and look for other planetary signals in the system.

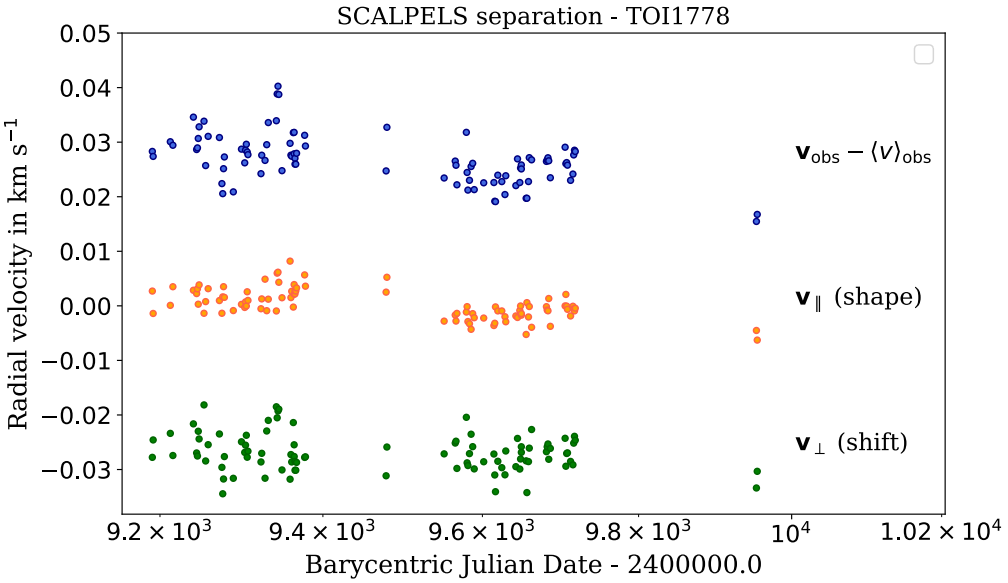


Figure 5.1: Blue scatterplot at the top is the barycentric RV with its own mean subtracted. The orange time series in the middle represents the shape-driven component obtained from the SCALPELS projection, while the green one shows the 'cleaned' shift-driven velocities, obtained when the shape-driven variations are subtracted from the observed RVs. Please note that an offset is introduced to each RV component for better illustration.

5.1.2 Results and discussion

I used the TWEAKS to perform the stellar activity mitigation in the wavelength and time domain simultaneously. This is achieved by incorporating the SCALPELS (Collier Cameron et al., 2021) basis vectors, representing the shape components (stellar activity component), in the nested sampling package KIMA (Faria et al., 2018), for stellar activity decorrelation. Figure 5.1 shows the SCALPELS separation and the obtained shift RVs (planet component, shown in green), corrected for stellar activity. Along with the stellar variability, the SCALPELS shape component (orange) also tracks the RV offsets between two seasons that occurred due to the

cryostat replacement (BJD=2459502). As we saw in the CoRoT-7 case (Section 3.2.2), some shift-like signals elude SCALPELS analysis, so any remaining rotationally-modulated signals were modelled with GP regression applied to the RVs.

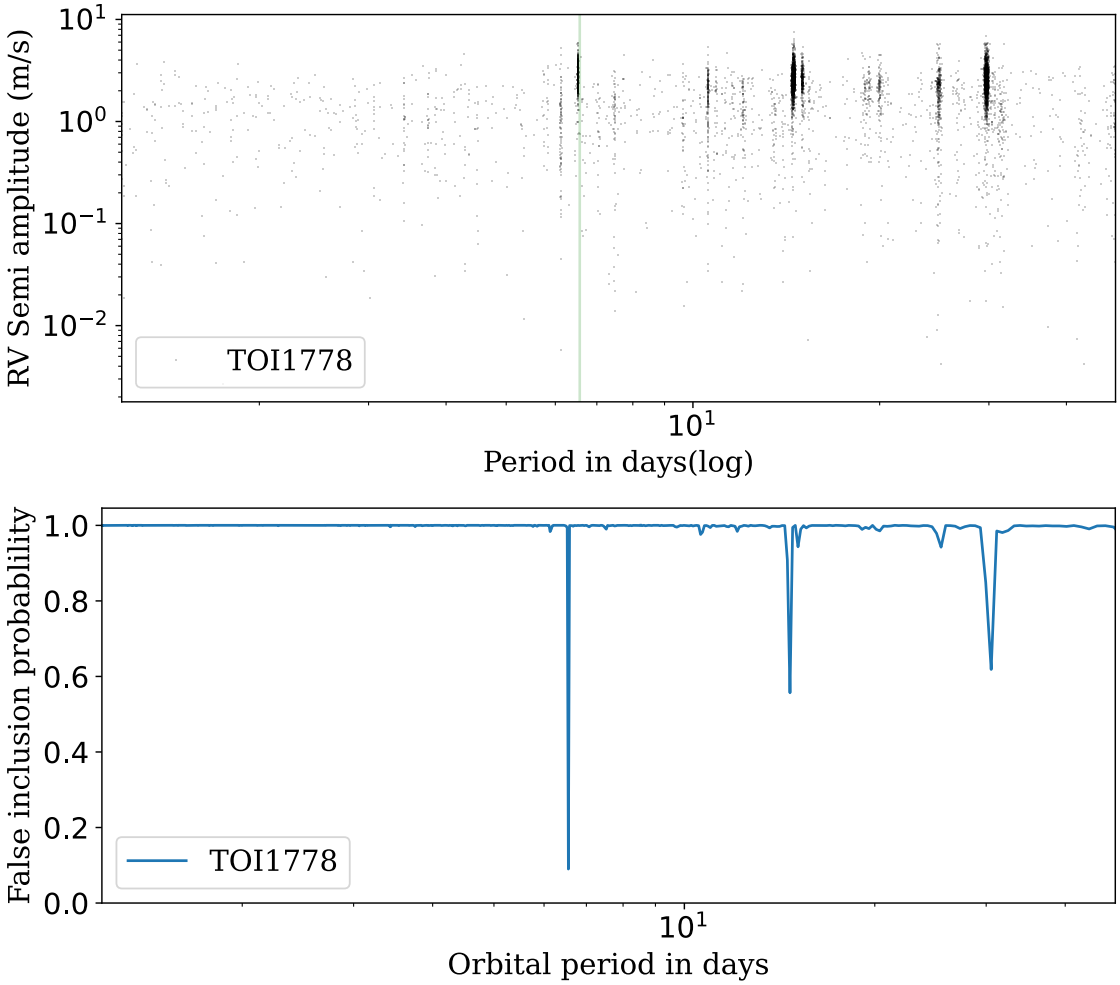


Figure 5.2: Top: The joint posteriors from a KIMA analysis that simultaneously decorrelates against the SCALPELS **U**-vectors and GP used to constrain the stellar rotation and searches blindly for planets in TOI 1778 RVs. The light green vertical lines indicate the orbital period of the known transiting signal. **Bottom:** The FIP diagram for the posteriors showing strong (close to zero) FIP at the orbital period of the known transiting planet.

Similar to CoRoT-7, we used a quasi-periodic GP kernel described by Equation 2.8, where t and t' are two times of observation, and the other terms are, the stellar variability amplitude (η_1), the coherence timescale (associated with the growth and decay of spot regions and associated faculae on the surface of the star, η_2), the variability (or rotation) period of the star (η_3) and the harmonic complexity (number of bumps in the RV curve per rotation cycle, which depends on the number and longitude distribution of active regions, as well as the

smoothing effect of limb darkening and foreshortening), η_4. The GP used log-uniform priors on the amplitude ($0 < \ln(\eta_1) < 2.7$) and active-region evolution timescale ($24 < \eta_2 < 40$ d). These boundaries were selected based on the stellar type of TOI-1778, a G dwarf, meaning its active regions evolve on approximately the same timescale as the Sun. The $v \sin i = 3.5 \pm 0.5$ km s^{-1} , and the stellar radius of 1.31 R$_\odot$ limits the stellar rotation period to a value of below 22 days. Based on this estimate and a few initial runs with uninformative priors, the GP priors on the stellar rotation period were set to be uniform with $12 < \eta_3 < 22$ d. The log harmonic complexity was controlled with a uniform prior -1 $< \ln(\eta_4) < 0$. In short, we used a model with up to five unknown Keplerian signals, stellar activity decorrelation performed against the SCALPELS basis vectors, incorporated with a quasi-periodic GP regression for a blind RV search.

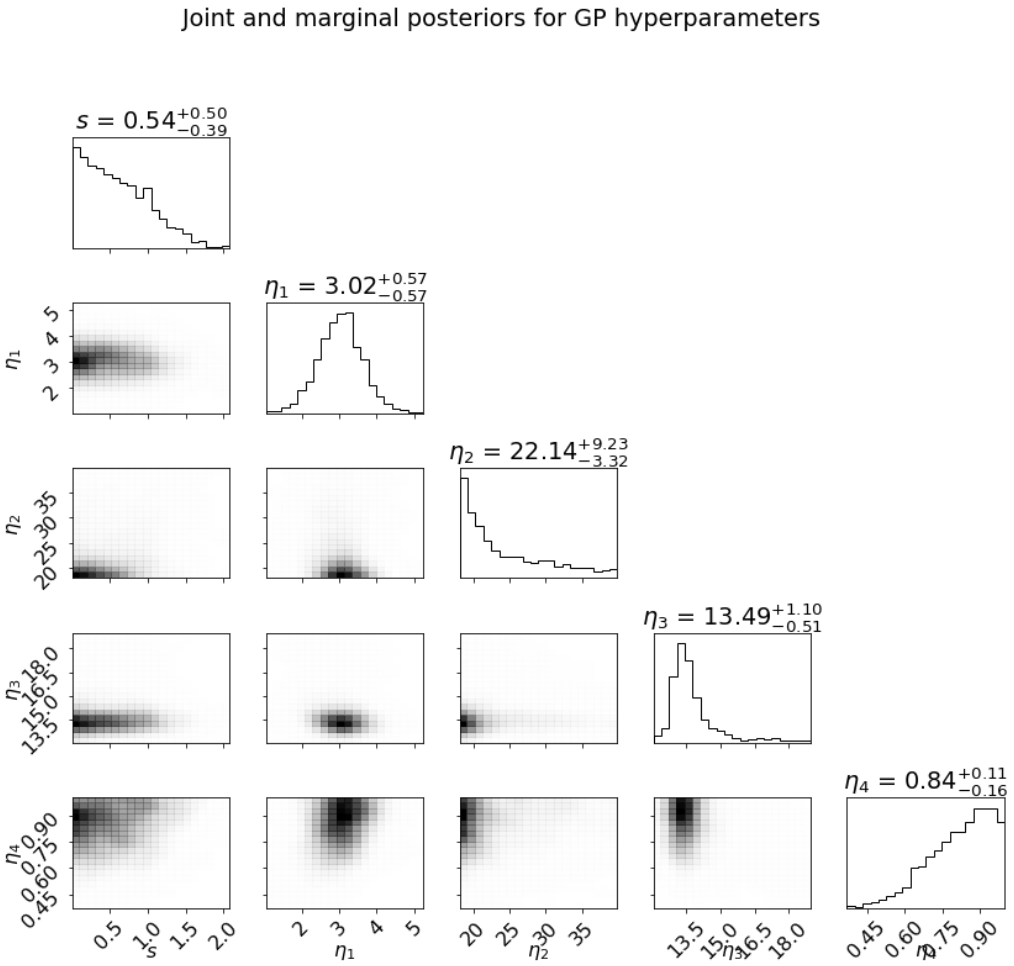


Figure 5.3: The posterior distribution of the GP hyperparameters for TOI 1778. The stellar rotation period (η_3) converges around 13.49 days.

The joint posteriors from this blind search showed clear detection of a Keplerian signal at orbital period $\sim$ 6.52 days, as well as identified a stellar rotation of $P_{\rm rot} = 13.49^{+1.10}_{-0.51}$ days (See

Figure 5.3). Next, using priors on the orbital ephemeris of the transit candidate, I performed a second analysis treating it as a 'known' signal, much like what I had done for CoRoT-7b in Chapter 3. By integrating the resulting RV semi-amplitude of 2.64 ± 0.50 m s^{-1} (Figure 5.2) with the stellar mass and orbital inclination (1.17 $M_{\odot}$ and 87.10 degrees respectively), I measured the mass of TOI 1778 b to be 8.69 ± 1.62 $M_{\oplus}$, with a precision of 21%. This result was consistent with the results from other analyses used in the study such as joint GP, photometric, and the spectroscopic fit determined using `PyORBIT`. The precision in the mass measurement enabled us to constrain its interior composition to be made of an iron-core fraction smaller than that of the Earth and an H/He atmosphere that is about ∼1% of the total mass (Palethorpe et al., 2024).

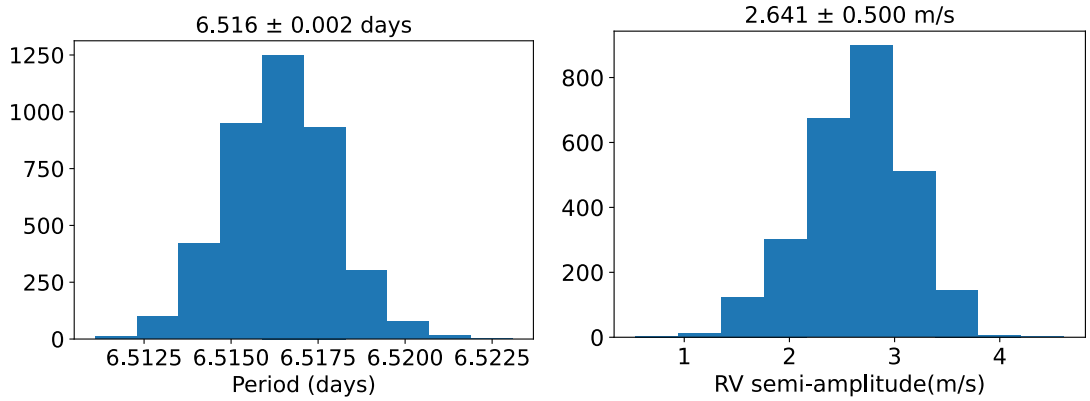


Figure 5.4: The histogram showing the posterior distribution for the orbital period and radial velocity semi-amplitude of the transiting planet signal.

5.2 TOI 1730

TOI 1730, a.k.a LHS 1903, which has been thoroughly studied by Wilson et al (in review), is the second target that benefited from the rigorous planet search and mass determination offered by TWEAKS.

5.2.1 History and observations

TOI 1730 is a multi-planetary system initially detected by *TESS* (Ricker et al., 2015). Following its initial identification by the SPOC and the subsequent transit validation tests, this system was designated as the *TESS* Object of Interest (Guerrero et al., 2021). Through a preliminary search[3], three transiting planets, with radii ∼ 1.5, 2.0, and 2.5 $R_{\oplus}$, were discovered to orbit

[3]https://tess.mit.edu/followup/

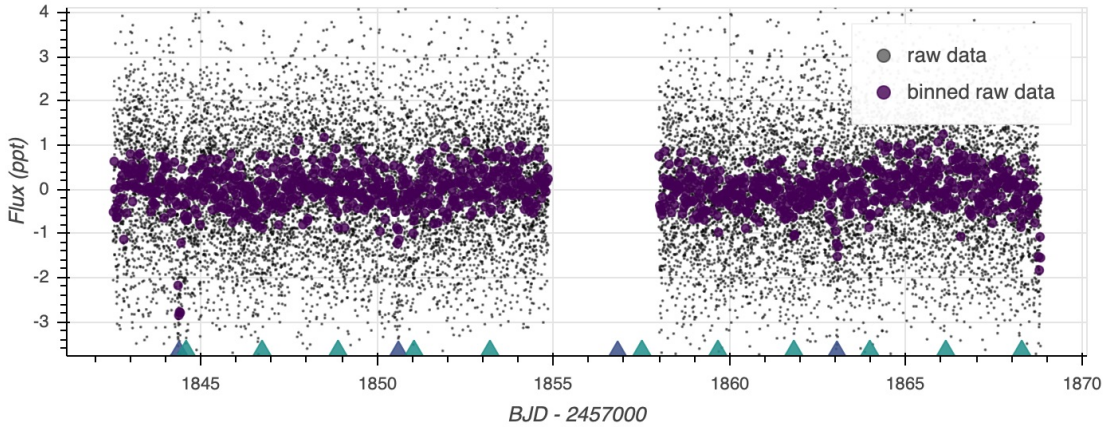


Figure 5.5: *TESS* lightcurve of TOI 1730 from 2 sectors. Triangles on the x-axis indicate the epochs of transits. Figure credits: https://hposborn.github.io/toi_plots/TOI1730_plot.html.

this M-dwarf host star every 2.16, 6.23, and 12.57 days (Figure 5.5). This suggested that the TOI 1730 planetary system seemingly traverses the radius valley (Fulton et al., 2017). Additional photometric observations were obtained from *CHEOPS* as part of the Guaranteed Time Observations (GTO) programme 54 ('X-Gal') to confirm and further photometrically characterise the TOI 1730 planetary system (Wilson et al in review).

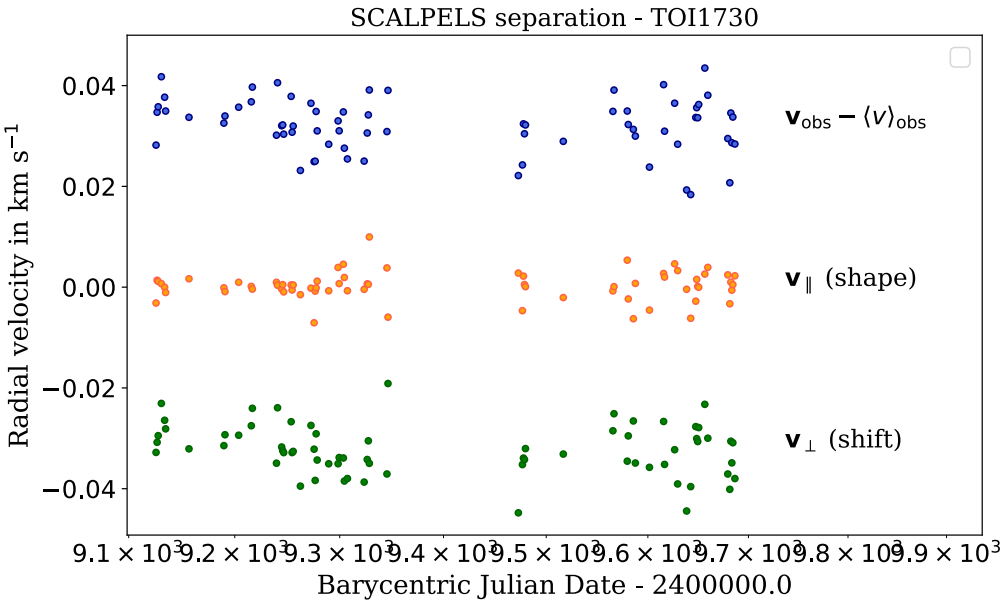


Figure 5.6: SCALPELS signal separation for TOI 1730. The colours have the same meaning as in Figure 5.1.

Later, an RV follow-up campaign within the GTO was spurred by the discovery of these multiple transiting planets around an M-dwarf. Spanning across three observing seasons and a 769-day baseline, HARPS-N monitored TOI 1730 from 2020 October 3 to 2022 November

10, with a brief hiatus caused by the Cumbre Vieja volcano outburst in late 2021. The hence obtained 65 spectra were used for the TWEAKS analysis, to validate the transit detections, find their masses and search for additional planets in the system.

5.2.2 Results and discussion

I used a model with up to five unknown Keplerian signals and a quasi-periodic stellar activity model to do a blind search of the radial velocities using TWEAKS. Through principal-component analysis of the spectral CCF's autocorrelation function, time-domain activity-decorrelation vectors were computed using SCALPELS (See Figure 5.6). As we saw in the cases of CoRoT-7 and TOI 1778, some shift-like stellar-activity signals can bypass SCALPELS analysis. To guard against this, a quasi-periodic GP regression was used to model any rotationally-modulated signals that remained.

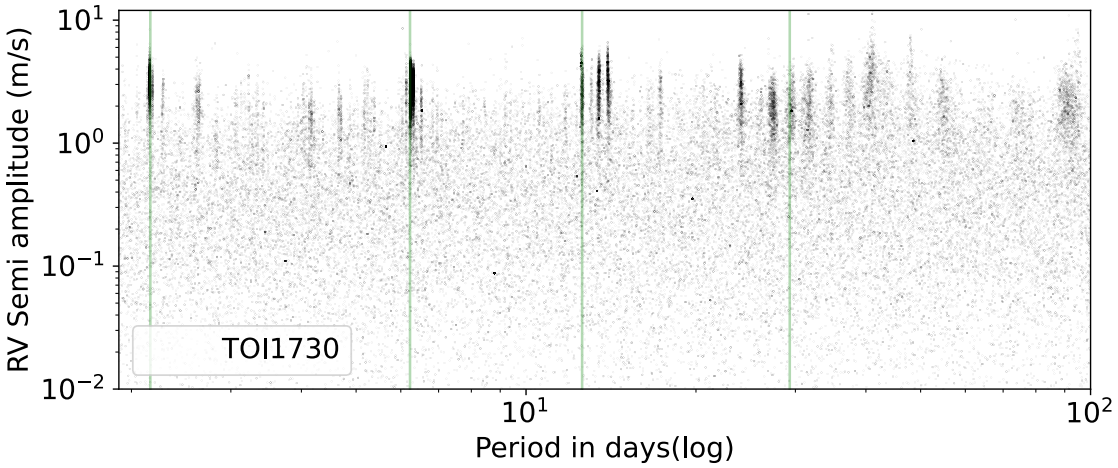


Figure 5.7: The joint posteriors from a KIMA analysis that simultaneously decorrelates against the SCALPELS **U**-vectors and GP used to constrain the stellar rotation and searches blindly for planets. The light green vertical lines indicate the orbital periods of the known transiting signals.

The quasi-periodic kernel adopted log-uniform priors on stellar variability amplitude ($0 < \ln(\eta_1) < 2.7$) and active-region evolution timescale ($60 < \eta_2 < 120$ d), for this preliminary investigation. A uniform distribution with $35 < \eta_3 < 50$ d and $-1.6 < \ln(\eta_4) < 0$ was used for the rotation period and log harmonic complexity priors, respectively. As in the previous cases, the priors were carefully tailored after a few initial runs with uninformative priors.

With η_3, or the stellar rotation period, strongly detected at 41.8 d, the joint posteriors revealed strong detections at 2.15 and 6.22 d, and marginal detections at 12.6 and 29.3 d (Figure 5.7). Using the same SCALPELS decorrelation vectors and GP regression model, a sec-

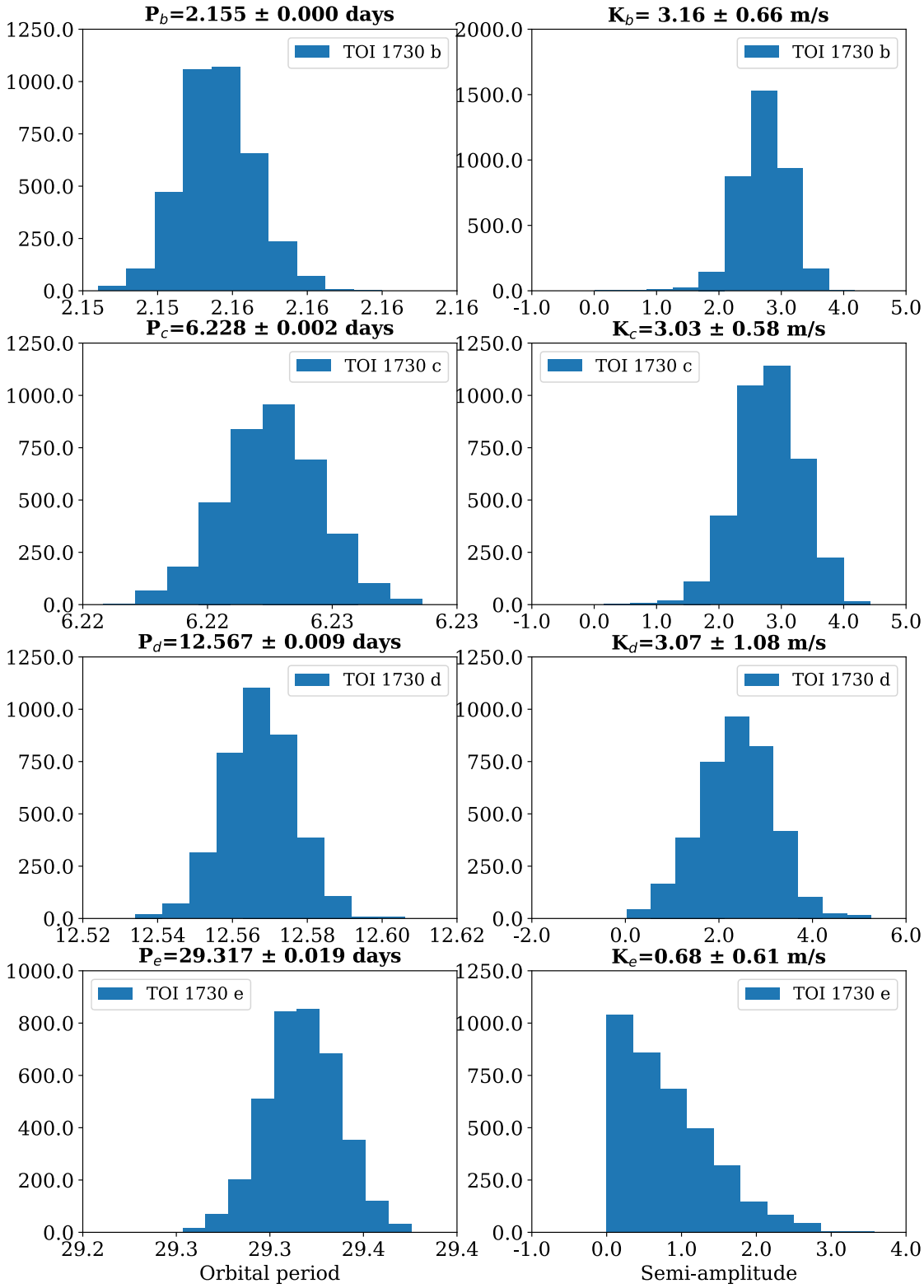


Figure 5.8: The histograms showing the posterior distribution for the orbital period (Left column) and radial velocity semi-amplitude (Right column) of the transiting and non-transiting planet signals around TOI 1730. These results are from a TWEAKS run where the four signals were considered as 'known' planets while a simultaneous SCALPELS decorrelation and GP regression was performed.

ond analysis treating the four transiting planets as known signals produced orbital parameters (Figure 5.8) that concurred with the results of the other analysis used in the paper which used photometry and RV data jointly.

The Gaussian mixture modelling employed in TWEAKS was then used to validate the outcomes of the above-described initial runs and to evaluate the system architecture further with HARPS-N RVs. The obtained results favoured a marginal detection of the 29.3 d planetary signal and favoured $>3\sigma$ detections for the three inner planets (Figure 5.8). We anticipated that the orbital period's (29.3 d) closeness to the typical lunar synodic period of 29.52 days may be the cause of the detection's lower significance. The total duration of observation (769 days) is much shorter than the beat period ($\frac{1}{\frac{1}{29.3}-\frac{1}{29.52}} \sim 4000$ days). Therefore, it is highly likely that these two signals might overlap. These might be made worse by the observational epochs, as two or three blocks of observations were obtained each month, with one run usually centred on the full Moon each cycle due to HARPS-N's scheduling availability- limited dark-sky period. Due to the high absolute stellar RV and distance to the ecliptic plane (about 25^o), Moon contamination is anticipated to be negligible; but, stronger contamination may alter the CCF shape, which would then get corrected using SCALPELS. We conducted an additional analysis employing a GP alone (and no SCALPELS) to verify this claim, which revealed strong detections ($\geq 3\sigma$) for all four transiting planets. The obtained radial velocity semi-amplitudes are 3.16±0.66, 3.03±0.58, 3.07±1.08 and 1.92±0.31 m s^{-1} , respectively (Figure 5.7).

These strong detections yielded refined mass measurements, which enabled the authors to determine the internal structures of the LHS 1903 planets (Wilson et. al in review).

5.3 HD 48948

HD 48948 is another interesting active star to which we applied TWEAKS to search for planetary signals. Dalal et.al (in review), analysed this system thoroughly by co-ordinating analyses using several time-domain and wavelength-domain approaches. The investigation includes the usage of stacked BGLS (Mortier & Collier Cameron, 2017), $\ell 1$ periodogram (Hara et al., 2017), GP regression with `MAGPy_RV` (Rescigno et al., 2023), pyOrbit (Malavolta et al., 2017) and YARARA (Cretignier et al., 2021) were used. Except YARARA, these are all time-domain approaches. TWEAKS applies its time-domain and wavelength-domain integrated approach to produce independent yet consistent results. In this study, TWEAKS was used as a standalone

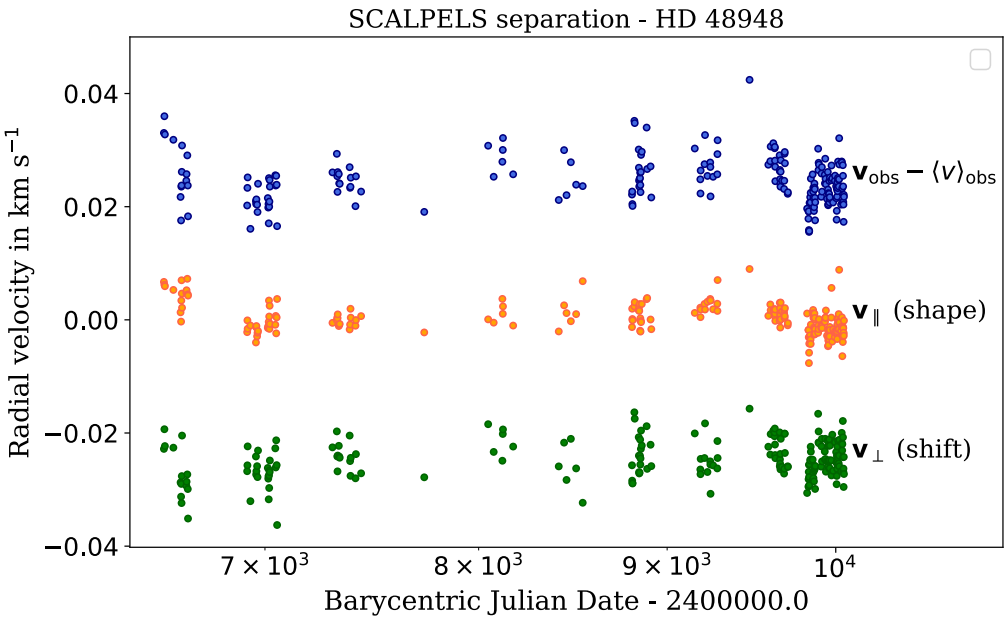


Figure 5.9: SCALPELS signal separation for HD 48948.

and simultaneous planet search tool, which allowed us to compare its performance with other contemporary techniques that handle stellar variability and instrumental systematics simultaneously while searching for planets.

5.3.1 History and observations

HD48948 is a K3-dwarf star 52.18 light-years (Gaia Collaboration et al., 2023) away from Earth, of which HARPS-N GTO collaboration have collected observations from October 2013 to April 2023, as part of the RPS program. I used 189 binned spectra for the TWEAKS analysis which were observed strategically to minimise the effect of stellar activity phenomena like granulation, which spans from minutes to hours.

5.3.2 Results and discussion

Using TWEAKS, the CCFs derived from the HARPS-N DRS were independently analysed for planetary signals. I started with a blind search of radial velocities, using a model with up to five undetected Keplerian signals using KIMA. To perform stellar activity decorrelation, I integrated the spectral line-shape decorrelation vectors produced by SCALPELS into KIMA.

The resultant joint posteriors favoured the detection of two Keplerian signals at orbital periods 7.3401 ± 0.0004 days and 38.01 ± 0.06 days. However, spurious signals due to instrumental systematics were observed at periods greater than 300 days. Therefore, in the

subsequent analysis, we limited the orbital period for the search for Keplerians to 270 days, to avoid any long-term systematics potentially impacting and interfering with any other valid planetary signals. As seen in Figure 5.10, this analysis reveals the existence of a third candidate on an orbit spanning 150.98 ± 0.77 days, without affecting the accuracy of the two inner planets' detection. The False Inclusion Probability (FIP) analysis (Hara et al., 2022a) done in frequency space, also supported these detections. I also verified the coherency of these signals across different data subsets to validate these findings. In contrast to signals caused by sampling patterns, stellar activity, or aliases, the 7.34 and 37.98-day signals were identified in all subsets with $\sigma \geq 5$.

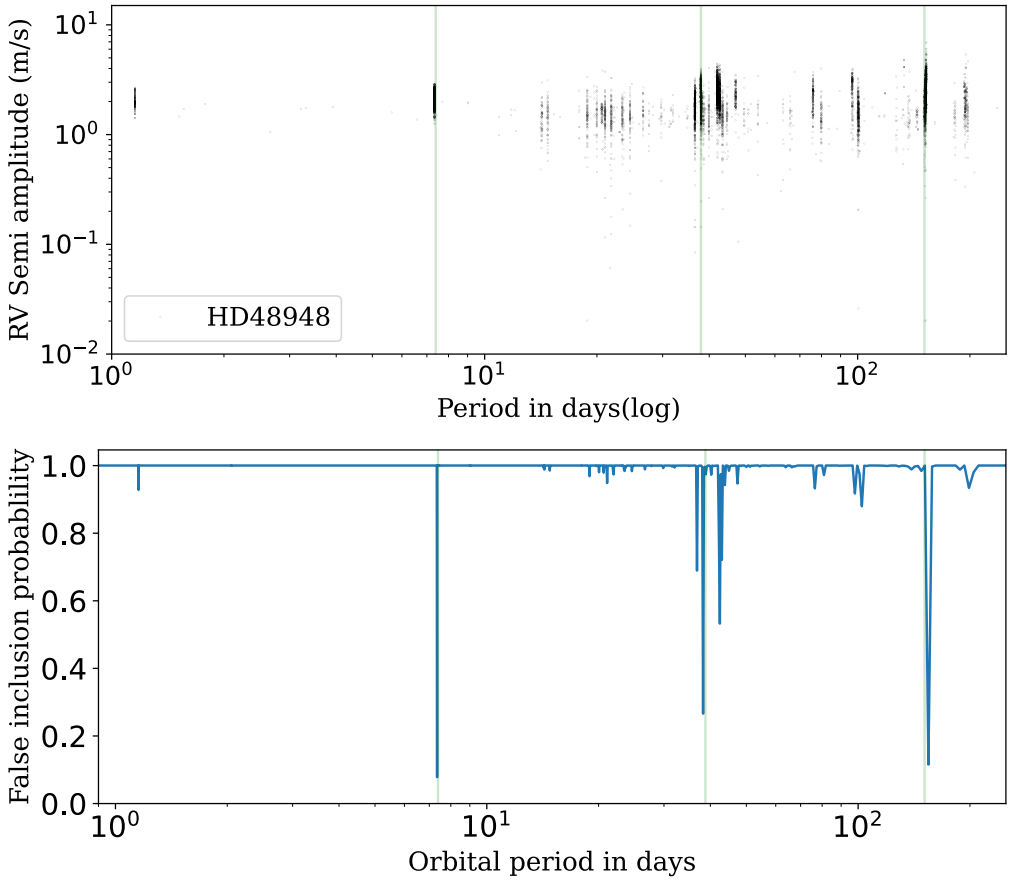


Figure 5.10: Top: posterior distribution. Bottom: False Inclusion Probability (FIP) periodogram. The vertical plain lines (light green) mark the three candidate planet detections with significant FIPs, at 7.34 days (HD 48948 b), 37.98 days (HD 48948 c), and 150.92 days (HD 48948 d), respectively. The clustering around the 37.98-day planet signal can be attributed to the stellar rotation period and its one-year aliases. Figure from Dalal et al. (in review).

I then performed a Gaussian Mixture modelling to the posteriors to obtain the planetary orbital parameters. The minimum masses for the first two planetary signals are calculated as 5.43 ± 0.30 $M_\oplus$and 6.33 ± 1.61 $M_\oplus$, respectively, using the derived RV semi-amplitudes of

2.29 ± 0.13 and 1.48 ± 0.39 m s^{-1} in conjunction with the stellar mass of 0.686 $M_{\odot}$. The RV semi-amplitude of the third signal is 1.55 ± 0.34 m s^{-1} , corresponding to an estimated mass of 10.05 ± 2.27 $M_{\oplus}$. This planet has an effective temperature of 349 K, with a minimal angular separation of 0.029 arc seconds from the star (according to Dalal et al. submitted, Section 8). This along with its proximity (16.8 parsecs) to our solar system makes HD 48948 a very promising candidate for NASA's planned Habitable Worlds Observatory (HWO) mission. Moreover, HD 48948 d will be the second closest super-Earth known within the habitable zone (Dalal et al. in review, Section 11.2).

5.4 HD 99492

Stalport et al. (2023) analysed the HARPS-N RVs for HD 99492 for planets by using different stellar activity mitigation tools. This includes using YARARA (Cretignier et al., 2021) to eliminate telluric absorption, and instrumental systematics at the spectral level, and using SPLEAF (Delisle et al., 2020) to apply a multidimensional correlated noise model to further reduce the stellar noise. TWEAKS functioned as an autonomous planet finder tool in this comprehensive analysis, that mitigates the effects of instrumental systematics and stellar activity while simultaneously offering independent confirmation of planetary signals. I would like to emphasise that all other techniques mentioned here and in Section 5.3 are pure-time domain techniques, which mostly use GP regression to address activity. These techniques only involve wavelength-domain mitigation when multivariate GP regression constrained against activity indicators is carried out. This work along with the analysis of HD 48948 helped us to benchmark TWEAKS with existing state-of-the-art planet detection tools that correct for stellar variability.

5.4.1 History and observations

HD 99492 (also known as 83 Leo B) is a K2V spectral type star. This star was monitored with HARPS-N between January 15, 2014, and June 17, 2022, yielding a total of 202 nightly binned spectra. The DRS measured radial velocities from the CCFs obtained by cross-correlating the extracted 1-D spectra with a K2 mask.

5.4.2 Results and discussion

The HARPS-N CCFs were independently analysed for planetary signals using TWEAKS. As the standard procedure, I started with a blind search of the radial velocities, using a model with

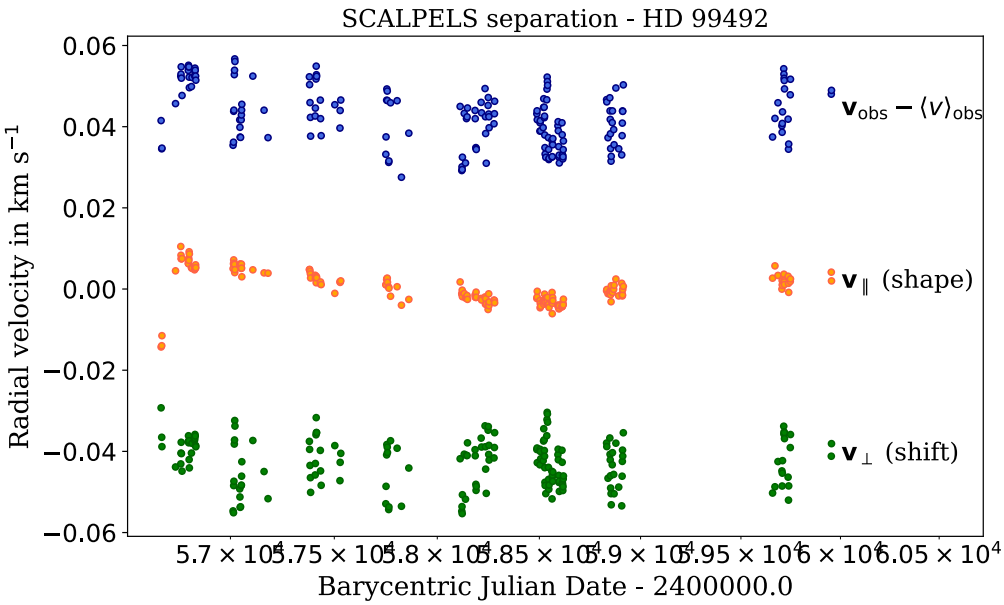


Figure 5.11: SCALPELS signal separation for HD 99492. The shape component (orange) is in good agreement with the simultaneous S-index measurements, showing the star's magnetic activity cycle.

up to five unidentified Keplerian signals. Similar to the analyses done on other stars, the CCF line-shape decorrelation vectors from SCALPELS were used for stellar activity decorrelation in KIMA.

At orbital periods of 17.05 days and 95.2 days, two Keplerian signals were detected strongly from the joint posteriors. The 17.05 days signal showed a strong FIP of 0.13, while the 95.2 days signal was detected even more strongly with a FIP of 0.04; that is, 96% of the mutually independent KIMA posterior solutions favoured a planet detection with an orbital period of 95.2 days. After combining the RV semi-amplitudes derived from this analysis with the stellar mass of 0.85 $M_\odot$, I determined the minimum masses of these planets to be 27.13 ± 1.18 $M_\oplus$and 19.85 ± 2.19 $M_\oplus$, respectively. The results of the combined analysis using YARARA and SPLEAF are compatible with the orbital parameters, and planetary masses obtained from TWEAKS, and are also within formal uncertainties. As we did for HD 166620 and HD 144579, we further verified these detections by verifying the coherency of the signals across independent data subsets (Figure 5.12). The 17.05 and 95-day signals are both identified in all the subgroups with sigma $\geq 9\sigma$, in contrast to signals arising from sampling patterns, stellar activity, or aliases. We found that the signal at 14d was only present in the second half of the data and is not coherent across time, in contrast to the other two planetary signals. So we discarded the possibility of this signal being an additional planet, similar to how we dealt with the 284-day signal found in HD 144579 data (Section 4.2.3).

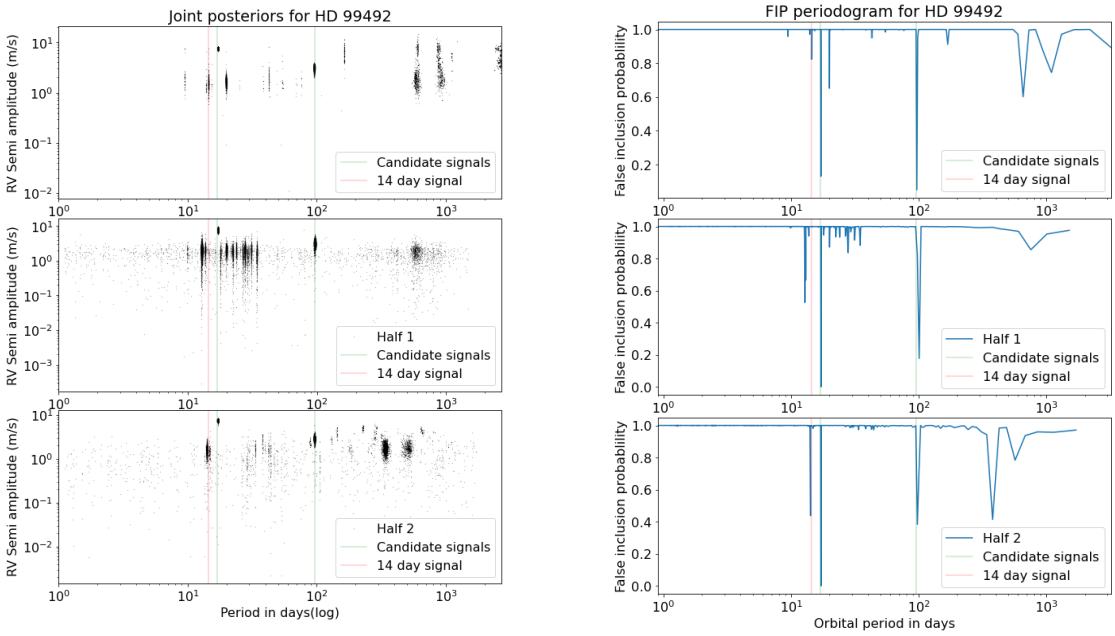


Figure 5.12: Left column: Joint posterior distribution for HD 99492. Right column: FIP periodogram. From top to bottom rows: Analyses on the entire RV time series, on the first half and on the second half of the time series, respectively. The vertical plain lines (light green) depict the two planet detections, at 17.05 and 95.2 days, respectively. The red dashed lines help to guide the eye on the 14d signal, which is inconsistently present in the data. Figure from Stalport et al. (2023).

5.5 TOI 5398

TOI 5398 or BD+37 2118 is a moderately young ($\sim$ 680 Myr, Mantovan et al. (2023)) solar-analogue star (G spectral type). After some TOIs[4] were identified in the system, this star was followed up with the HARPS-N spectrograph, as part of the GAPS program on planets orbiting young stars. The Global Architecture of Planetary System (GAPS) team asked for the possibility of applying TWEAKS on the star's HARPS-N RVs, to independently confirm the presence of the planets they find in *TESS* photometry and obtain their masses.

I employed the same fundamental TWEAKS approach that was used for every other star, with minimal tailoring. In this section, I will describe how I measured the masses of the planets and thereby contributed towards constraining their bulk composition.

5.5.1 History and observations

Mantovan et al. (2023) aimed to characterise the compact multi-planet system orbiting TOI 5398 using a combination of *TESS* and ground-based photometry, and RVs collected with HARPS-N

[4]https://tess.mit.edu/followup/

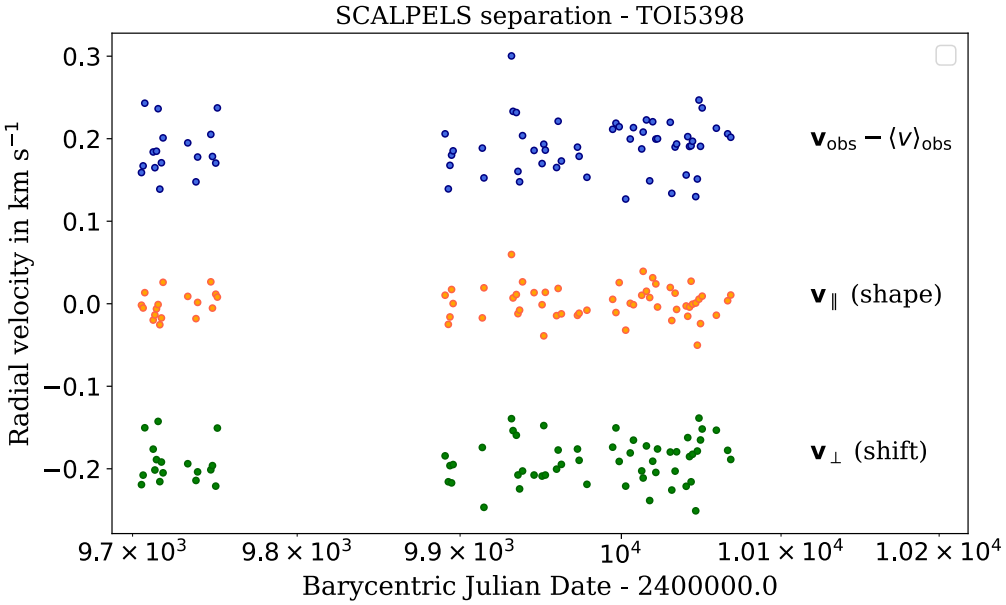


Figure 5.13: SCALPELS signal separation for TOI 5398.

spectrograph. Two candidate exoplanets were found by the *TESS* pipeline: a giant planet (P $\sim$ 10.59 d) that has been confirmed by Mantovan et al. (2022) and designated as TOI 5398 b, and a transiting sub-Neptune (P $\sim$ 4.77 d). The GAPS team observed this star with HARPS-N (A46TAC/32, PI: G. Mantovan) across a time span of 439 days for RV follow-up. I used the hence obtained 86 CCFs for the TWEAKS analysis.

5.5.2 Results and discussions

These 86 HARPS-N CCFs were independently analysed for planetary signals using TWEAKS. To identify the RV component induced by stellar variability, I obtained the time-domain stellar activity-decorrelation vectors using SCALPELS (Collier Cameron et al., 2021). These vectors were then used for spectral line-shape decorrelation in the trans-dimensional nested sampler KIMA (Faria et al., 2018).

A model with up to five unidentified Keplerian signals was used in our first blind search of the RVs using the KIMA. As some planet-like signals are likely to elude SCALPELS analysis, any remaining rotationally-modulated signals were modelled with GP regression applied to the RVs, using a quasi-periodic GP kernel. The kernel used for this GP analysis was similar to the ones used for TOI 1778 and TOI 1730, but the hyperparameters were carefully tailored for TOI 5398. The joint posteriors showed clear detection of TOI 5398 b, at orbital period 10.5905 $\pm$ 0.0002 days with an RV semi-amplitude of 14.51 $\pm$ 3.74 m s^{-1} . Combined with the stellar

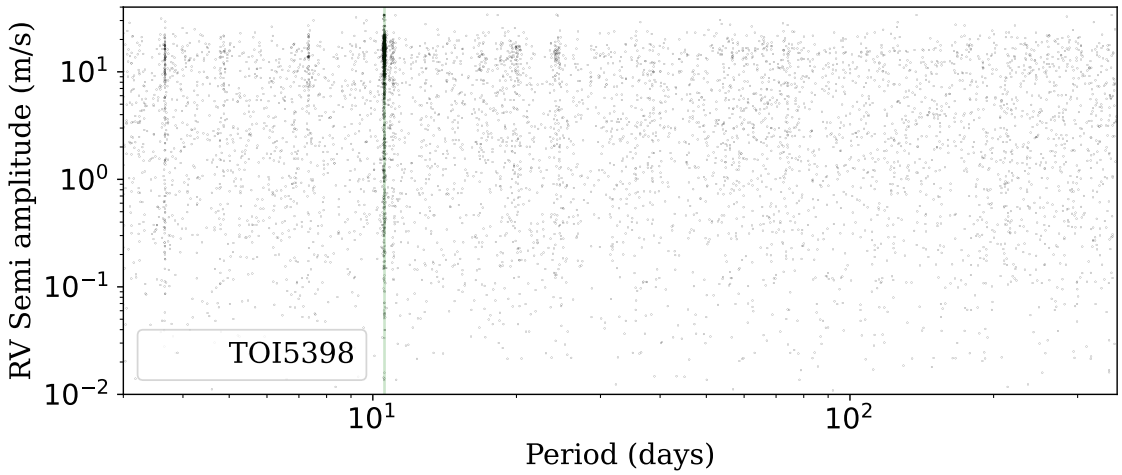


Figure 5.14: Kima posterior distribution for the TOI 5398 RV time series. The vertical line (light green) marks the planet detection, at 10.59 days, the transit period of planet b.

mass of 1.14 $M_\odot$, and an orbital inclination of 89.21^o, the planetary mass was measured to be 57.3 ± 4.77 $M_\oplus$. Additionally, the quasi-periodic GP strongly constrained the stellar rotation to P = 7.36 ± 0.11 days, which is consistent with the stellar radius and v sin i as reported by Mantovan et al. (2023).

TWEAKS analysis served as an independent detection of the planetary signal. My analysis aligned seamlessly with the RV reference model solution and the stellar parameters presented in Mantovan et al. (2023) (See Table 5.1). By providing a precise mass measurement for TOI 5398 b, TWEAKS helped Mantovan et al. (2023) to constrain the bulk density of the gas-giant planet to be 0.29 ± 0.05 g cm^{-3}. This intriguing study revealed that the gas giant TOI 5398 b has a radius similar to Jupiter and a mass about two-thirds that of Saturn, the smallest mass among giants in compact systems.

5.6 Summary

This chapter investigated the capacity of TWEAKS as an integrated time-domain and wavelength-domain method to generate independent but consistent results in various collaborative studies. TWEAKS was used as a standalone and simultaneous planet search tool, which allowed us to compare its performance with other contemporary techniques that handle stellar variability and instrumental systematics simultaneously while searching for planets. Table 5.1 gives a quantitative comparison of RV semi-amplitudes and planetary masses obtained with TWEAKS and via other methods. In the majority (78%) of the cases, TWEAKS yields mass measurements that are more precise or comparable to those obtained using the 'conventional' time-

Table 5.1: Summary table of the RV amplitudes and masses for all the planets discussed in this chapter. The mass measurements from TWEAKS and other methods are given in separate columns to allow a direct comparison of both values and fractional uncertainties given in brackets.

Stars	Planets	Period (days)	RV semi-amplitude ($m\ s^{-1}$) TWEAKS	Masses ($M_\oplus$) TWEAKS	Masses ($M_\oplus$) Other techniques
TOI 1778	TOI 1778 b	6.52±1e-5	2.68±0.54	8.69±1.62 (18%)	8.48±1.46[d] (17%), 8.38±1.32[f] (16%)
TOI 1730	TOI 1730b	2.15±5e-6	3.16±0.66	3.65±0.53 (14%)	3.00±0.48[g] (16%)
	TOI 1730c	6.22±2e-3	3.03±0.58	5.32±1.002 (18%)	4.44±0.89[g] (20%)
	TOI 1730d	12.56±1e-2	3.07±1.08	5.79±1.90 (32%)	5.90±1.47[g] (25%)
	TOI 1730e	29.52±1e-2	1.92±0.31	6.16±1.01 (16%)	5.66±1.91[g] (34%)
TOI 5398	TOI 5398b	10.59±2e-4	14.51±3.4	57.3±4.77 (8%)	58.8±7.55[d] (13%), 58.7±5.65[e] (10%)
HD 48948	HD 48948b	7.34±4e-4	2.28±0.15	5.43±0.30 (6%)	4.86±0.37[a] (8%), 4.96±0.32[b] (7%)
	HD 48948c	38.06±6e-2	1.48±0.39	6.33±1.61 (18%)	3.94±2.90[a] (74%), 6.90±1.01[b] (15%)
	HD 48948d	$151.9 \pm 4e-1$	1.55 ± 0.34	10.05±2.27 (22%)	9.72±2.41[a] (24%), 11.01±1.55[b] (14%)
HD 99492	HD 99492b	17.05±2e-3	7.52±0.33	27.13±1.18 (4%)	25.5±0.6[c] (2%)
	HD 99492c	95.23±9e-2	3.09±0.34	19.85±2.19 (10%)	17.9±1.3[c] (7%)
CoRoT-7	CoRoT-7b	0.85±5e-7	4.29±0.46	6.05±0.65 (10%)	4.73±0.95[h] (20%), 5.53±0.83 (15%)
	CoRoT-7c	3.69±5e-3	5.76±0.29	13.29±0.69 (5%)	13.56±1.08[h] (8%), 12.62±0.75 (6%)
	CoRoT-7d	8.97±1.5	5.53±0.79	17.14±2.55 (15%)	16.7±0.42[k] (3%)

The 'other techniques' listed in the table are covered in the first paragraph of Section 5.6.

domain approaches. **Other techniques** used in the combined analyses of various stellar targets mentioned in this chapter and summarised in Table 5.1 are: [a] A 1D GP analysis on YARARA (Cretignier et al., 2021) treated RVs, [b] A Multi-dimensional GP analysis on YARARA (Cretignier et al., 2021) treated RVs, [c] SPLEAF (Delisle et al., 2020) + YARARA (Cretignier et al., 2021), [d] A 1D GP analysis (https://github.com/LucaMalavolta/PyORBIT) using DRS RVs, [e] A Multi-dimensional GP analysis (Rajpaul et al., 2015; Barragán et al., 2022), [f] A 1D GP analysis (https://github.com/LucaMalavolta/PyORBIT) using s-BART RVs (Silva et al., 2022), [g] `JULIET` (Espinoza et al., 2019) + 1D GP (Foreman-Mackey et al., 2017) on s-BART RVs (Silva et al., 2022), [h] 1D GP (Haywood et al., 2014), [j] 1D GP + nested sampling (Faria et al., 2016), [k] this study (Hatzes et al., 2011) considered the 9.02 d beat period for the analyses.

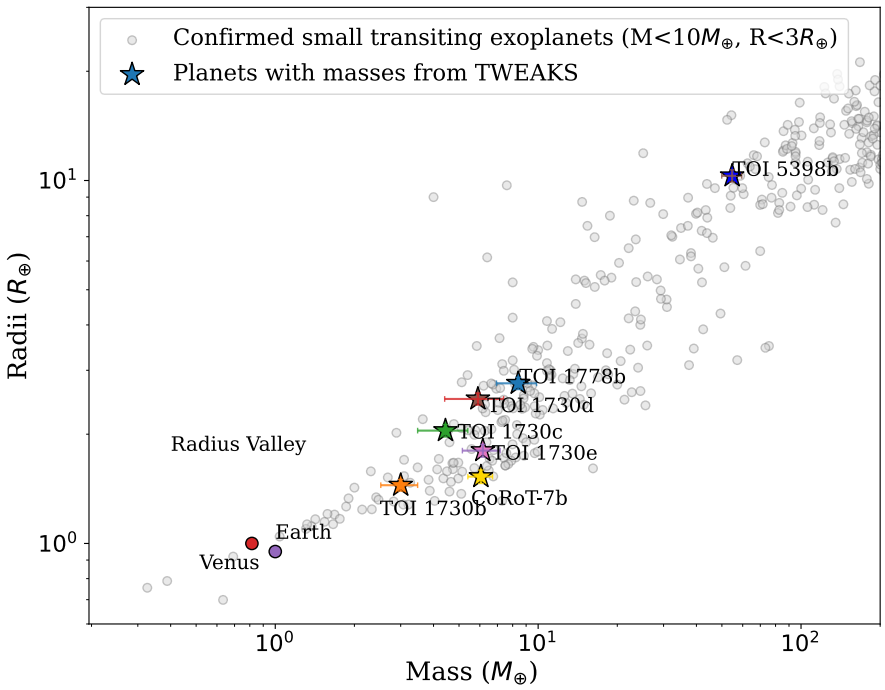


Figure 5.15: Top: The transiting planets TOI 1778b, TOI 5398b, TOI 1730b, TOI 1730c, TOI 1730d, TOI 1730e and CoRoT-7b on a Mass-radius diagram. Earth and Venus are shown as circle-shaped symbols for comparison. An approximate location of the radius valley is also shown for reference. Other exoplanets for which the radius and mass are known are also represented along with the measured uncertainties.

By efficiently constraining planetary mass measurements, TWEAKS contributed significantly towards constraining the bulk densities of transiting planets; TOI-1778b as 0.426 ± 0.090 g cm^{-3}, TOI-5398b to be 0.29 ± 0.05 g cm^{-3}, TOI 1730b as 5.38 ± 1.21 g cm^{-3}, TOI 1730c as

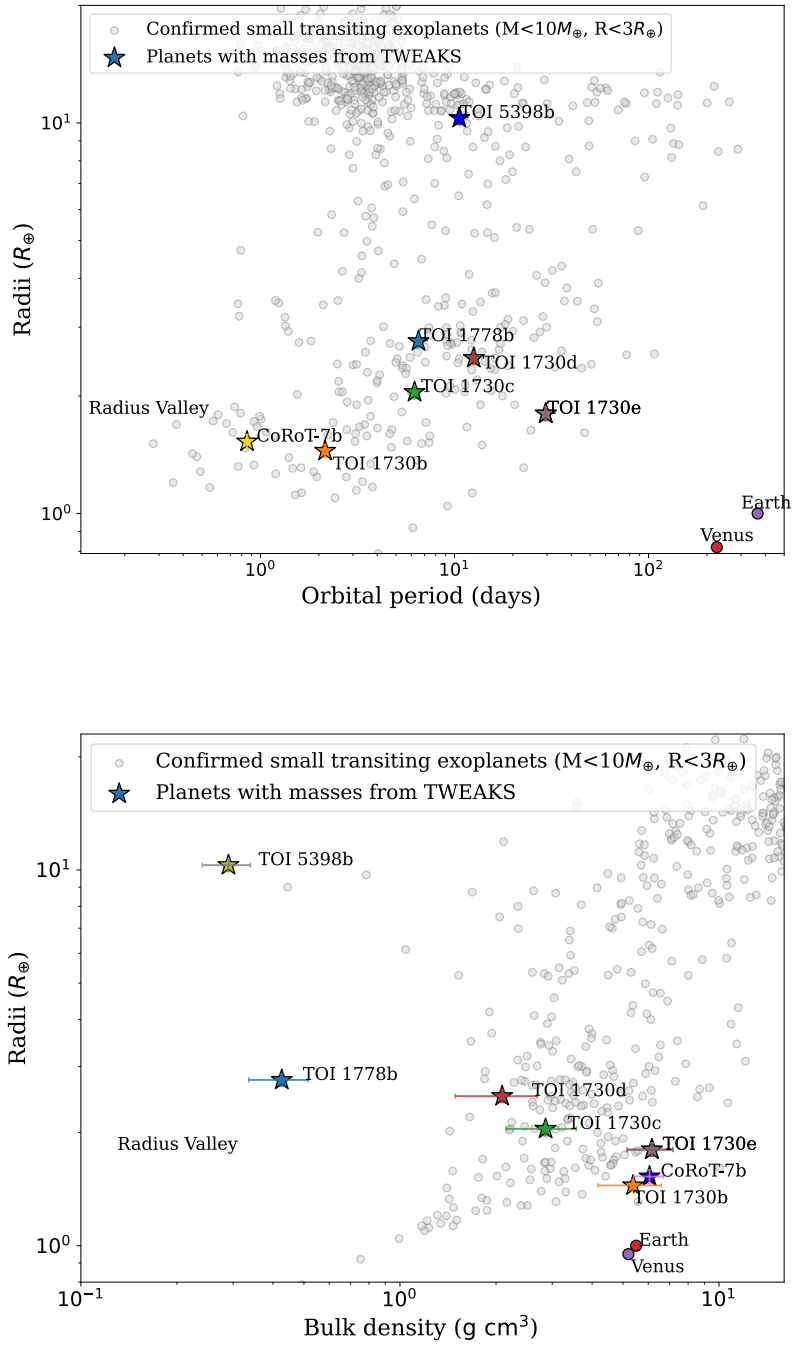


Figure 5.16: Top: The transiting planets TOI 1778b, TOI 5398b, TOI 1730b, TOI 1730c, TOI 1730d, TOI 1730e and CoRoT-7b on a period-radius (top) and density-radius (bottom) space. Earth and Venus are shown as circle-shaped symbols for comparison. An approximate location of the radius valley is also shown for reference.

2.86 ± 0.71 g cm^{-3}, TOI 1730d as 2.09 ± 0.60 g cm^{-3} and TOI 1730e as 5.22 ± 1.76 g cm^{-3} (Figure 5.16). In line with the theory that planets above the radius-gap (Fulton et al., 2017)

are volatile-rich, TOI 1730b and CoRoT-7b appear to have smaller radii (than TOI 1730 e) and Earth-like densities, while TOI 1730c and TOI 1730d have larger radii (than TOI 1730 e) and lower densities (Figure 5.16). This makes TOI 1730e the only long-period planet with a refined density that lies above the radius valley model predictions, and TOI 1730 system the first to contain four small, well-characterized planets that span the M-dwarf radius valley (Wilson et al in review). See the orbital period-radius figure from Wilson et al (in review) for a more accurate depiction of radius valley models.

For non-transiting planets, we could only obtain a minimum mass (Msin$i$), since we lack information about the orbital inclination i (eq: 1.5). Therefore, I obtained minimum masses for the non-transiting planets; HD 48948b as 5.43 ± 0.30 $M_{\oplus}$, HD 48948c as 6.33 ± 1.61 $M_{\oplus}$, HD 48948d as 10.05 ± 2.27 $M_{\oplus}$, HD 99492b as 27.13 ± 1.18 $M_{\oplus}$and HD 99492c as 19.85 ± 2.19 $M_{\oplus}$. See Figure 6.1 for all the planets with masses measured using TWEAKS in the period-mass space.

6

Conclusion & Future prospects

In our relentless pursuit of habitable worlds, we are at the turn of the tide in the realm of ground-based radial velocities (RV), as the cutting-edge extreme precision planet-hunting spectrographs (eg; EXPRES, NEID, ESPRESSO) are locked onto a remarkable goal of achieving $\leq$0.3 m s^{-1} RV precision in the optical domain. This propels us another stride closer to discovering 'Earth-twins orbiting in the habitable zone (HZ) of other Suns' (exo-Earths), by inching towards the requisite precision of $\sim$0.1 m s^{-1} . Although we are in possession of these state-of-the-art instruments, we have not yet harnessed their full potential, mainly owing to the RV barrier imposed by the host star's intrinsic activity at $\sim$1 m s^{-1} . This hinders our ability to definitively enter the realm of terrestrial planet discovery, determine their precise masses, and characterize them. This Thesis centres around understanding and mitigating the stellar activity impacts in RVs and advancing towards the targeted sub m s^{-1} precision.

6.1 What is TWEAKS good for?

Over the course of my PhD, I formulated a workflow called TWEAKS to obtain reliable detections of planets whose RV signals were previously below the detection threshold. The pipeline integrates stellar activity mitigation techniques in the time (KIMA; Faria et al. (2018)) and wavelength domains (SCALPELS; Collier Cameron et al. (2021)). The wavelength domain part of the pipeline separates the apparent RV variations caused by spectral line-**shape** changes (stellar activity component) from the Doppler **shifts** (planet component). The time-domain counterpart of the pipeline performs a GP regression, if stellar activity decorrelation against the SCALPELS **shape** component fails to address any shift-like shape signals while simultaneously searching for pure Doppler shifts using KIMA trans-dimensional nested sampler.

Thanks to this statistically rigorous approach, I can now confidently detect planets and reliably quantify their masses, allowing the precise determination of their bulk densities. My work definitively illustrated how the detection, precise mass measurement, and characterization of exoplanetary systems are made challenging by stellar variability, and the consequently limited knowledge of the presence of additional planets in the system, particularly for the low-mass planets whose RV amplitudes are near the detection threshold.

6.1.1 Reliable bulk densities and minimum masses for transiting and non-transiting exoplanets

A planet's mass is arguably its most fundamental physical characteristic, which together with its radius determines the bulk density, interior composition, and structure. However, the precision of the mass determination is impeded by the imprint of the host star's stellar activity in the RV signal, particularly for Earth-sized planets. Achieving a precision of at least 10% in mass (and 5% in radius) is imperative for distinguishing between rocky planets with iron cores and those primarily composed of water (Zeng et al., 2016). Accumulating reliable mass measurements for a substantial number of exoplanets holds the potential to offer crucial insights into the processes that govern the formation of these planetary systems.

Improving mass measurements of known transiting planets is the major goal of an efficient stellar activity mitigation strategy. Deciphering the atmospheric composition calculations for scale height and surface gravitational accelerations also depends on precise mass measurements. Current stellar activity mitigation methods, encompassing time and wavelength domain approaches, have enabled the extraction of mass from RV signals with amplitudes of 1-2 m s^{-1} , amid stellar variability (eg: Haywood et al. (2014); Rajpaul et al. (2015); Faria et al. (2018); Collier Cameron et al. (2021)). However, for most Earth-sized planets detected by TESS, more advanced methods are required to convert upper limits into precise mass measurements. The challenge that awaits PLATO (set to launch in 2026) will be to measure RV amplitudes of temperate terrestrial planets, which will be an order of magnitude less than RV amplitudes of *TESS* temperate planets orbiting low-mass stars (Orbital semi-amplitude, $K_* \sim (T_{eq}.M_p)/{M_*}^{3/2}$).

TWEAKS provides a strong foundation to expand in this direction. As discussed in Chapters 3 and 5, TWEAKS was used to obtain robust detections and precise mass determinations for several Super-Earths and mini-Neptunes in multiple planetary systems including CoRoT-7; Anna John et al. (2022), HD 99492; Stalport et al. (2023), TOI 5398; (Mantovan et al., 2023), TOI 1778; (Palethorpe et al., 2024), TOI 1730; (Wilson et al in review) and HD 48948; (Dalal et al in accepted). Figure 6.1 shows all these published (or close to being published) planets with mass measurements from TWEAKS. Unlike the non-transiting planets, transiting ones benefit from having their radius measured directly through photometry during transits. Therefore, their masses combined with their radii will provide an insight into what the planet

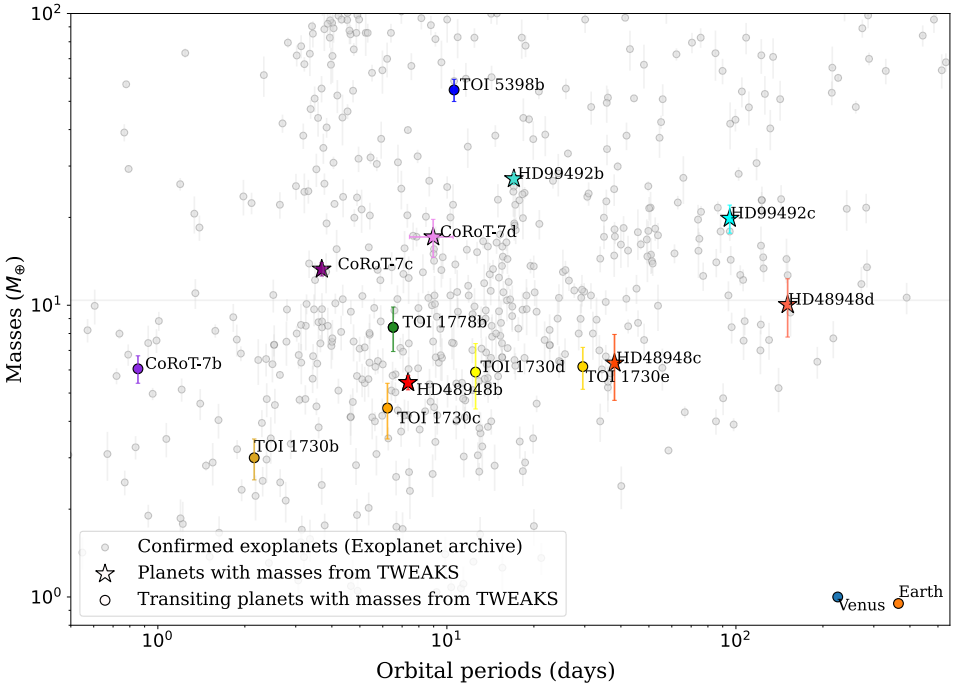


Figure 6.1: The planets with masses determined by TWEAKS are shown in period versus mass space. is made of, and what its structure might be like. The contributions of TWEAKS in constraining the bulk densities of several transiting planets are shown in Figure 5.16.

6.1.2 Pushing down the 1 m/s stellar activity barrier

The new generation of Extreme Precision RV (EPRV) spectrographs can measure RVs at unprecedented precision and stability. However, despite these instrumental advancements, the declining trend in the detectable RV semi-amplitude has hit a 1 m s^{-1} saturation limit, at an amplitude comparable to RV variations produced by the intrinsic variability of the host star. This hampers our capacity to detect Earth-sized low-mass exoplanets orbiting in the HZs of Sun-like stars, which require an RV precision of 0.1 m s^{-1} . In short, we are living in a 'stellar activity limited era' rather than an 'instrument limited era'.

As discussed in Chapter 4, the pipeline I put together for mitigating the stellar activity imprints, in turn, allowed me to push down the 1 m s^{-1} stellar activity barrier in RV measurements. TWEAKS has demonstrated its capability to reach a detection limit of 0.54 m s^{-1} in the decade-long HARPS-N RVs, where the calibration precision of the instrument is 0.50 m s^{-1} . Recently, a few studies have demonstrated how Convolutional Neural Networks (CNN) can be

used to attain a 0.20 m s^{-1} detection limit in the HARPS-N Solar data.

6.2 What next?

The field of exoplanet detection and characterization is highly dynamic and rapidly evolving. The stellar activity barrier, which I made an effort to address during my PhD, is one of the major challenges faced by the exoplanet community today. Breaking through the stellar activity barrier is necessary to reliably detect Earths orbiting other stars in wider orbits. Although the approach put forth by this Thesis, which is complemented by a handful of contemporary techniques offers promise to address the issue impressively, there are still some gaps in our understanding that must be filled to provide an effectively workable solution.

6.2.1 Address the time lag between RV and line-shape proxies

The main caveat in the existing TWEAKS pipeline is the leakage of time-varying (activity-induced) Doppler shifts caused by local changes in the convective gas flow (Collier Cameron et al., 2021; Anna John et al., 2022)), which limits the 'perfect' isolation of the RV Doppler **shift** component. The anticipated reason for this is the dramatic increase in the observed RVs, two to three days before active-region complexes cross the core of the solar disc. This causes temporal offsets between the SCALPELS **shape** component and RVs, which limit the extent of linear decorrelation (See Figure 3.4).

Numerous magnetic activity proxies like Full-Width Half Maximum (FWHM), Bisector Inverse Slope (BIS) and Ca II H&K are also known to share this phase lag (Santos et al., 2014; Queloz et al., 2009). The relationship between the HARPS-N solar RV observations and the two conventional CCF shape indicators, FWHM and BIS, was investigated by Collier Cameron et al. (2019). It was discovered that the apparent variations in RV preceded the BIS by three days and the FWHM by one day. As BIS tracks the faculae (which are longer-lived than spots) and is affected by the magnetic inhibition of convective blueshift, it peaks away from the disc centre. FWHM tracks the brightness deficits or excesses caused by spots or faculae, and peaks at the disc centre. It is also observed that these lags can be a function of the stellar rotation period (Collier Cameron et al., 2019). Jinglin Zhao. et al. (2022) also observed a ∼3-day lag in the ϕESTA metrics (which uses Fourier analysis to trace the CCF line-shape variability), similar to the findings from previous studies on the bisector inverse slope and FWHM.

The time-lag problem affects the activity proxies obtained from line-shape analysis techniques such as SCALPELS and ϕESTA; these proxies exhibit general behaviour similar to the RV but with a time lag. A plot of proxy versus RV thus follows a hysteresis curve.

To tackle this issue, we have to first comprehend the physical origin of magnetic processes inducing these time lags, to build better proxies (resistant to the lags) that can be used for linear RV de-correlation. While a well-crafted multivariate GP regression can be used to model these lags, it also must be grounded in a thorough comprehension of the physical causes underlying the RV shifts. Using CNNs trained to model the time lags between time-domain basis vectors (shape-vectors, which are comparable to the SCALPELS reduced-rank **U**-vectors), Jinglin Zhao recently presented a novel approach involving 'feature matrices' to deal with these lags. As a first step along this line, I worked on obtaining correlation coefficients between **U**-vectors and RVs. I aim to resume this research as soon as I submit this thesis. It is anticipated that the analysis presented in Section 6.2.3 also will open up new avenues for addressing the time lags.

6.2.2 Study the RV signature of granulation phenomena

Novel post-processing tools have demonstrated their efficacy in mitigating non-planetary systematics at the spectrum level by pushing through the '1 m s^{-1} noise floor' imposed by stellar variability. However, these attempts appear to be capped at $\sim$50 cm s^{-1}, a level comparable to granulation-induced RV variations. Recently, granulation phenomena were exposed to be a primary limiting factor in the detection of exo-Earths (Meunier & Lagrange, 2019). The solar observations from HARPS-N show that solar granulation induces RV variations at 0.4 m s^{-1}, whereas active regions passing across the visible solar disc produce signals at 5 m s^{-1} (Collier Cameron et al., 2019). Although granulation has been extensively studied for the Sun and is the easiest to average due to their short associated timescales, supergranulation (which are much larger versions of granules) has a longer lifetime and is potentially more troublesome (Meunier & Lagrange, 2019). Supergranules cover the whole Sun, whose large sizes make them appear to be fewer in number at any given time. This complicates the process of averaging out, and their resultant RV jitter remains substantial (0.28 to 1.12 m s^{-1} ; Meunier & Lagrange (2019)). **Therefore, if we were to examine raw solar RV data, these would prevent Earth from being discovered around the Sun, as the Sun's reflex motion in response to Earth's pull is $\sim$ 0.09 m s^{-1}). If this is where we stand with the data of our closest star**

where we can obtain extreme precision observations, it will be an enormous challenge to detect Earths around other stars of similar spectral type.

Unlike magnetic activity regions traversing the rotating stellar disc, super granulation does not have any counterpart in photometry either (because the temperature contrast is < 1K; Meunier et al. (2007)). This hinders the possibility of using photometric data to reconstruct the RV perturbations arising from these flows as used by several time-domain methods (Aigrain et al., 2012; Haywood et al., 2014). Therefore, finding a direct spectral proxy for granulation is crucial to overcoming the stellar activity barrier in determining precise masses of exo-Earths.

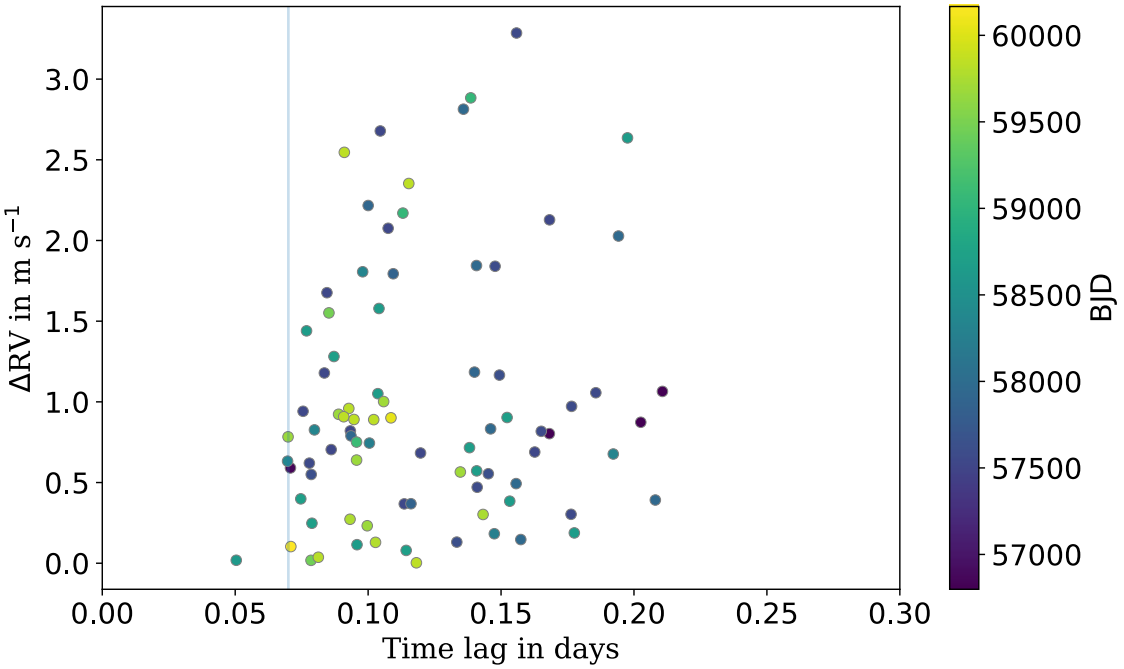


Figure 6.2: 141 intra-night pairs of 900-s HARPS-N observations of HD166620 secured between 100 minutes and 5 hours apart since 2012 show statistically significant RMS velocity scatter, expected to be due to granulation, that increases with the lag between observations. The bifurcation in the RMS distribution is concerning, as we cannot establish any time dependency if it is instrumental, demanding immediate investigation. Figure from the accepted proposal.

I have recently been awarded HARPS-N time to study the RV signature of granulation phenomena (Program: AOT49/10, PI: Anna John) in HD 166620, the K2 star in the Maunder Minimum phase (See Section 4.1.4 for more information about this star). During this phase of exceptionally low magnetic activity, we expect granulation to be the dominant nightly noise source, and supergranulation to manifest over 2 to 3 nights. Existing 12 years of RPS program observations for HD 166620, suggest that there is detectable intra-night RMS scatter in this star (Figure 6.2), indicating the presence of one or more sources of correlated noise, to which

granulation is likely to be a significant contributor. We will be observing HD 166620 densely across 2 back-to-back nights over the summer of 2024, with short exposures. The main goal is to determine how variations in RV correlate with spectral line shapes, or instrument telemetry, or both. While the existing observations were obtained to average out this noise, the proposed observations will enable us to identify the origin of the observed dispersion, whether it is astrophysical or instrumental.

My goal is to establish if one can trace the effect of granulation/super-granulation on stellar lines and achieve precision RVs to better than 0.5 m s^{-1} . This research might shed new light on our understanding of stellar granulation and how it affects RV variations. Studying the impact of stellar surface convection on the RV signal is an area of paramount importance to improve our capability of detecting Earth analogues around solar-like stars. The intensive intra-night cadence programme is essentially a part of a bigger objective, which is to serve as an initial test/proof of concept that can enable some of these interesting granulation questions.

6.2.3 Investigate CCFs for formation temperature

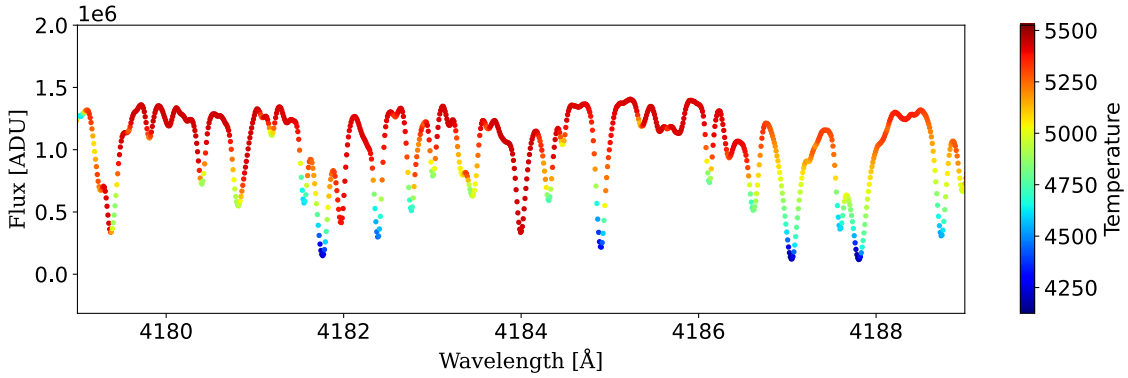


Figure 6.3: A segment of an individual HARPS-N Solar spectrum colour-coded based on the average formation temperature.

Al Moulla et al. (2022) examined how derived RVs depended on the temperature at which spectral line segments formed. Using spectral synthesis (PySME[1]), Al Moulla et al. (2022) determined the stellar temperature below which 50% of the emergent flux originates for each observed wavelength point of unblended spectral lines. Then, using template matching, they created an RV time series for various temperature ranges. One of the major findings is that, as the temperature increases, the effect of convective blueshift progressively grows. That is, the line segments corresponding to the coolest temperatures are the least impacted by the

[1]https://github.com/AWehrhahn/SME

convective blueshift. It was demonstrated that the absolute convective blueshift of line cores is a linear function of formation temperature, instead of a nonlinear function of line depth (Al Moulla et al. (2022),' the third signature of granulation'; Gray (2009)). The continuum and cores of weak lines form in deeper, hotter layers of the photosphere, where the convective upflows in the bright cores of granules are faster than in the cool upper layers where the cores of strong lines are formed. Given that the formation temperature maps more consistently with photospheric depth than with line depth, it seems like an appropriate variable to probe various layers in the photosphere (Al Moulla et al., 2022).

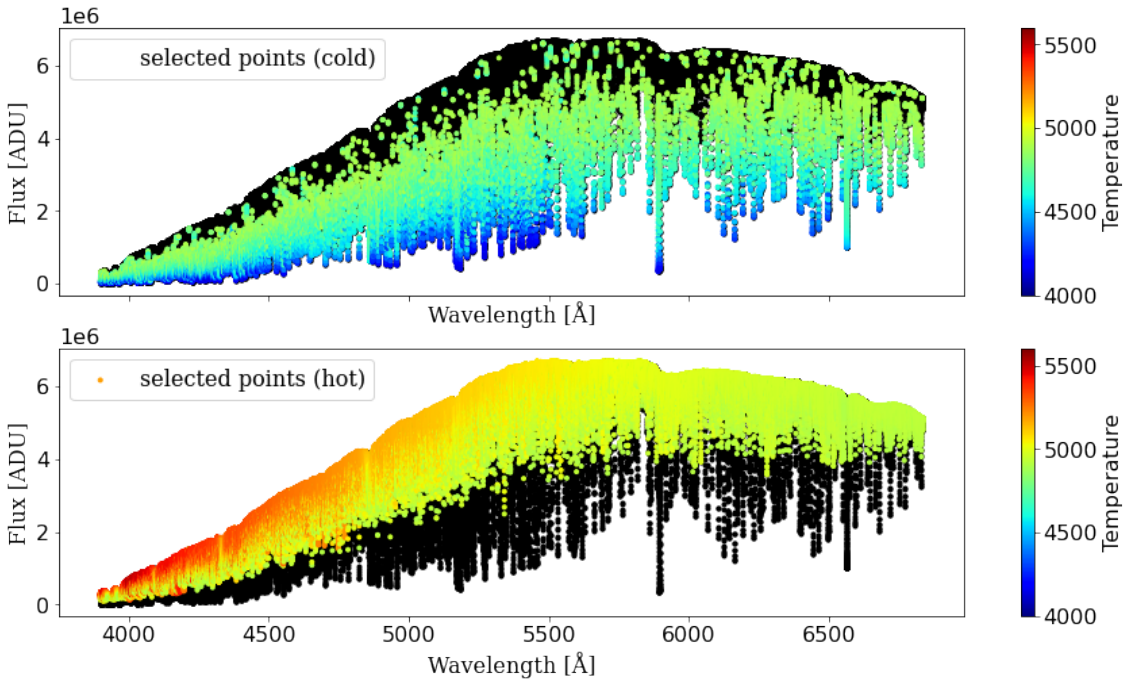


Figure 6.4: A normalised instantaneous spectrum showing the selection of spectral points based on the average formation temperature.

As a preliminary line of investigation, I employed the average formation temperature technique in the most recent HARPS-N Solar spectra in collaboration with AlMoulla. Figure 6.3 shows a segment of the spectrum colour-coded with the formation temperature of each point in the spectrum. Cooler temperatures correspond to the cores of the lines forming at the top of the photosphere, while higher temperatures form the wings of the lines forming deep within the photosphere.

The idea is to extract CCFs from different parts of the spectrum representing the points formed at different temperatures and to investigate the resultant **U**-vectors. To start with, I divided the individual spectrum (continuum colour corrected using `RASSINE`; Cretignier et al.

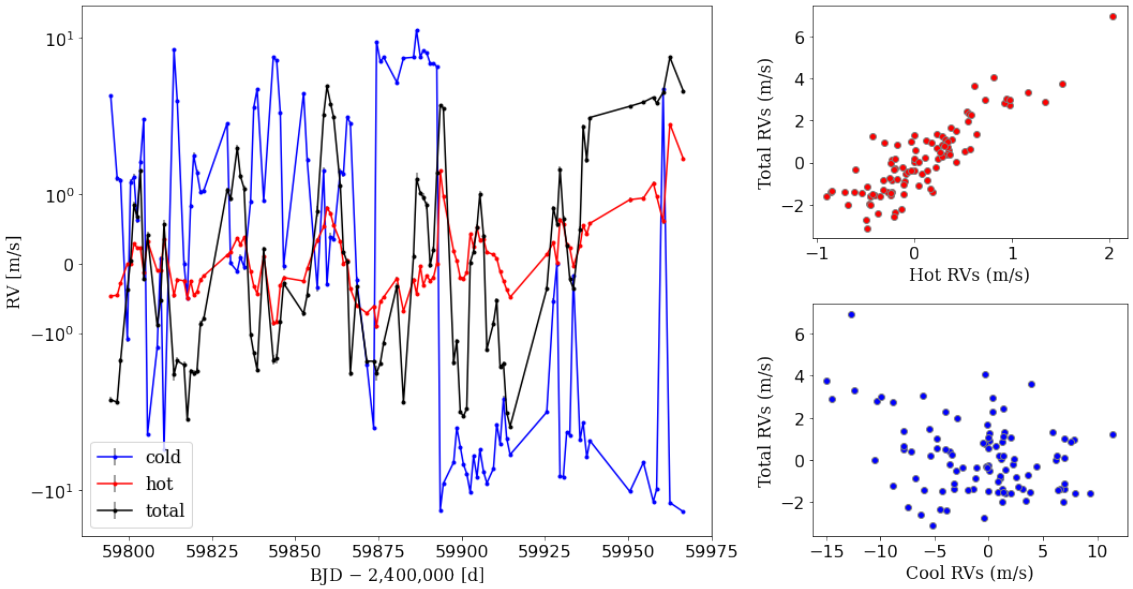


Figure 6.5: The RV time series obtained from the hot (red) and cool (blue) CCFs. RVs from the entire CCFs are shown in black. The right panels show the correlation of hot and cool RVs with the total RVs. A strong correlation is observed between the hot and total RVs.

(2020b) and post-processed using YARARA; Cretignier et al. (2020a)) into two groups; a cool and a hot group as shown in Figure 6.4. I then extracted the CCFs from the cool and hot points and measured the corresponding RVs.

The left panel of figure 6.5 shows that the line segments that formed at cooler temperature ranges have an inverted RV variation compared to the hotter temperature ranges. This is in agreement with the results obtained by Al Moulla et al. (2022), from the RVs using template matching. Therefore, measuring the RV for various photosphere temperatures may be a useful technique for distinguishing between planetary and stellar activity signals. The next step is to obtain different sets of **U**-vectors for the hot and cool CCFs using SCALPELS. I also aim to investigate the capacity of this tool to handle the time lags mentioned in the previous section. It will also be interesting to see if the differences between the red and blue RVs are linearly independent of the SCALPELS basis vectors. They are effectively **shift** RVs, but by subtracting one set of RVs from the other, any true dynamical motion will be eliminated.

A

Appendix

This section includes some algorithms, figures and tables that support the science described in Chapters 2 and 3.

Simultaneous sinusoidal fit: Steps

After computing the singular value decomposition (SVD) of the ACF rows;

1. Compute $\mathbf{F} = \{\cos \omega_1 t_j, \sin \omega_1 t_j, \cdots, \cos \omega_n t_j, \sin \omega_n t_j\}$.

2. Concatenate $\mathbf{A}_{(m\times(2n+l))} = \begin{bmatrix}\mathbf{F}_{m\times 2n} & \mathbf{U}_{A,(m\times l)}\end{bmatrix}$ using reduced rank l.

3. Compute reduced-rank covariance matrix using:

$$\frac{1}{m}\mathbf{R}^T_{k_{\max}} \cdot \mathbf{R}_{k_{\max}} = \frac{1}{m}(\mathbf{P}_{C,k_{\max}} \cdot \mathrm{diag}(\mathbf{S}^2_{C,k_{\max}}) \cdot \mathbf{P}^T_{C,k_{\max}}). \tag{A.1}$$

4. Compute row variances σ_j^2 of $\mathbf{R} - \mathbf{R}_{k_{\max}}$

Table A.1: Top: The periods and semi-amplitudes of the strongest signals in the periodograms from simultaneous modelling of CCF shape changes and planetary motion, made with prior knowledge of the periods (from l1) of 3 planet candidates and a single period representing the stellar rotation are listed. **Bottom** A pair of beat periods corresponding to the stellar rotation is included instead of the single period, considering the varying phase and amplitude.

P(days)	K(m s^{-1})	σ_K(m s^{-1})
0.8535	2.915	0.280
3.6963	5.112	0.302
8.9674	3.223	0.292
22.9425	5.437	0.321
0.8535	3.099	0.284
3.6963	5.343	0.303
8.9674	3.320	0.293
22.9425	5.142	0.326
23.6968	2.923	0.337

5. Compute model of full covariance matrix as:

$$\Sigma_{\xi,\eta}(t_j) \simeq \sigma_j^2 \times \max\left(1 - \frac{|\nu_\xi - \nu_\eta|}{\delta\nu}, 0\right) + \frac{1}{m}\mathbf{R}^T_{k_{\max}} \cdot \mathbf{R}_{k_{\max}}. \tag{A.2}$$

6. Compute $\mathbf{C}'$ and and Var($\mathbf{v}(t_j)$) using:

$$\mathrm{Var}(\mathbf{v}(t_j)) = \frac{1}{\mathbf{C}'(\nu, t_j)^T \cdot \mathbf{\Sigma}^{-1}(t_j) \cdot \mathbf{C}'(\nu, t_j)}. \tag{A.3}$$

7. Construct $\Sigma = \mathrm{Diag}(\mathrm{Var}(\mathbf{v}(t_j)))$.

8. Solve

$$(\mathbf{A}^T \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{A}) \cdot \theta = \mathbf{A}^T \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{v}_{\mathrm{obs}} \tag{A.4}$$

   to obtain θ and $\mathrm{Var}(\theta) = 1/\mathrm{Diag}(\mathbf{A}^T \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{A})$.

9. Partition θ into coefficients of columns of $\mathbf{F}$ and $\hat{\alpha}$.

10. Compute $\mathbf{v}_\parallel = \hat{\alpha} \cdot \mathbf{U}_A$ in reduced-rank basis.

11. Compute $\mathbf{v}_\perp = \mathbf{v}_{\mathrm{obs}} - \mathbf{v}_\parallel$.

Return: RV amplitudes and variances, $\mathbf{v}_\parallel$, $\mathbf{v}_\perp$.

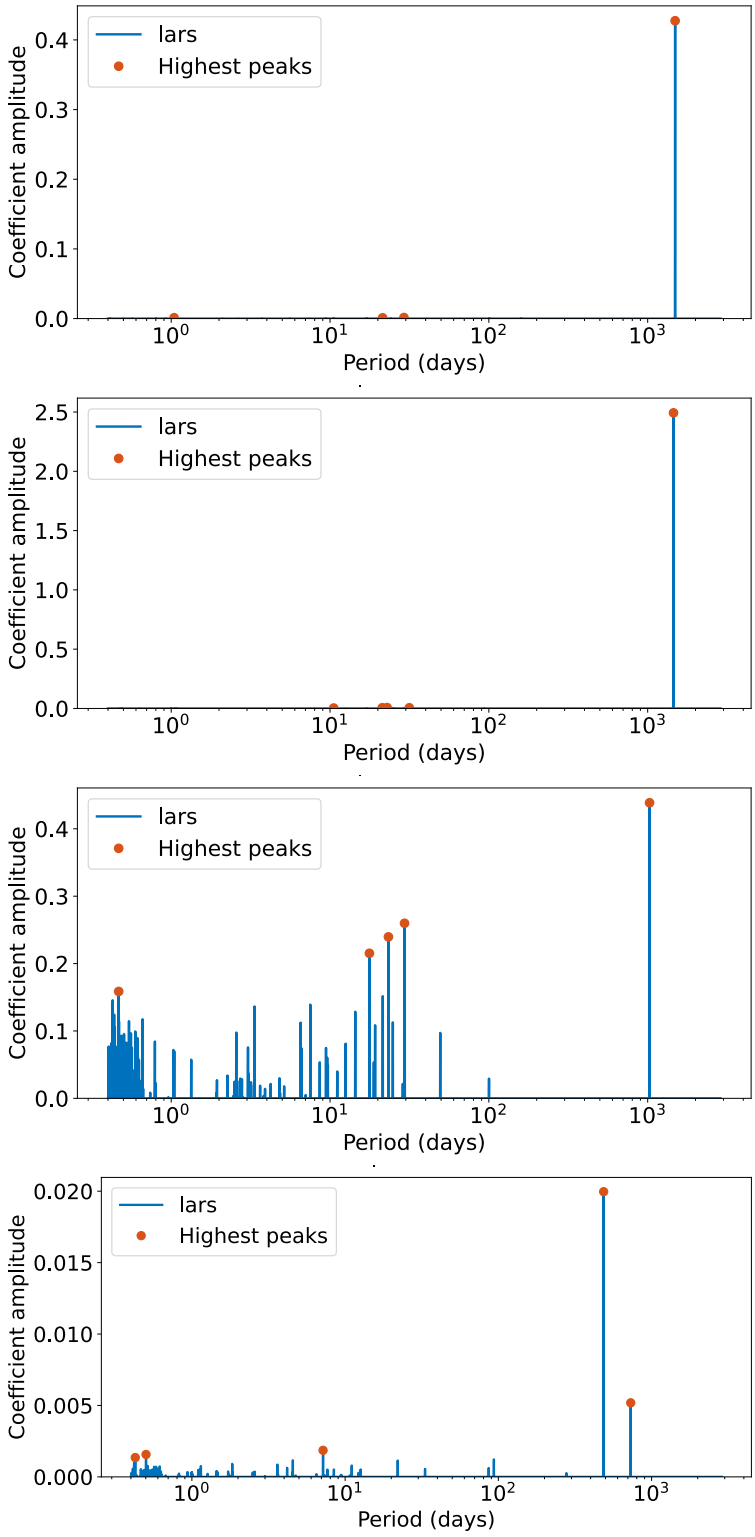


Figure A.1: Top: The ℓ_1- periodogram for the shape-driven RVs and traditional activity indicators like FWHM, Area and the BIS are shown from top to the bottom panels. None of these show significant signals at 0.85 days, 3.69 days and 8.96 days, adding to the case for a dynamical origin. The corresponding False Alarm Probabilities (FAP) are also listed above each panel.

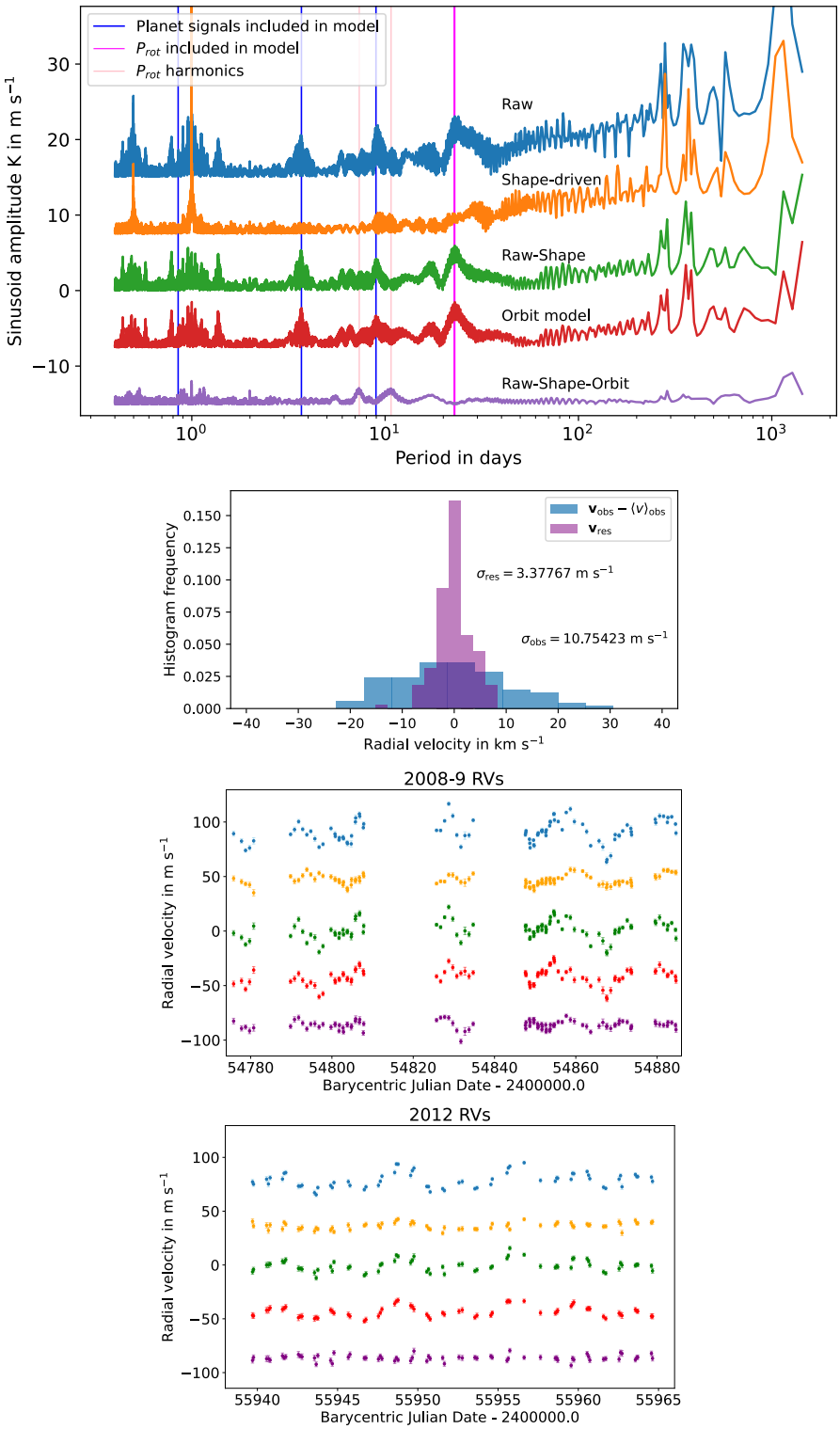


Figure A.2: Left: All the signals found by the ℓ_1- periodogram are fitted including the 3 planet candidate signals (blue vertical lines) and 1 pair of closely spaced beat period associated with the rotation period (magenta vertical lines). The orbital model signal (red) matches with the shift signal (green), except for the harmonics (light pink vertical bars) of the stellar rotation period. **Right**: The resultant histogram of residuals with significantly reduced RMS scatter is shown.

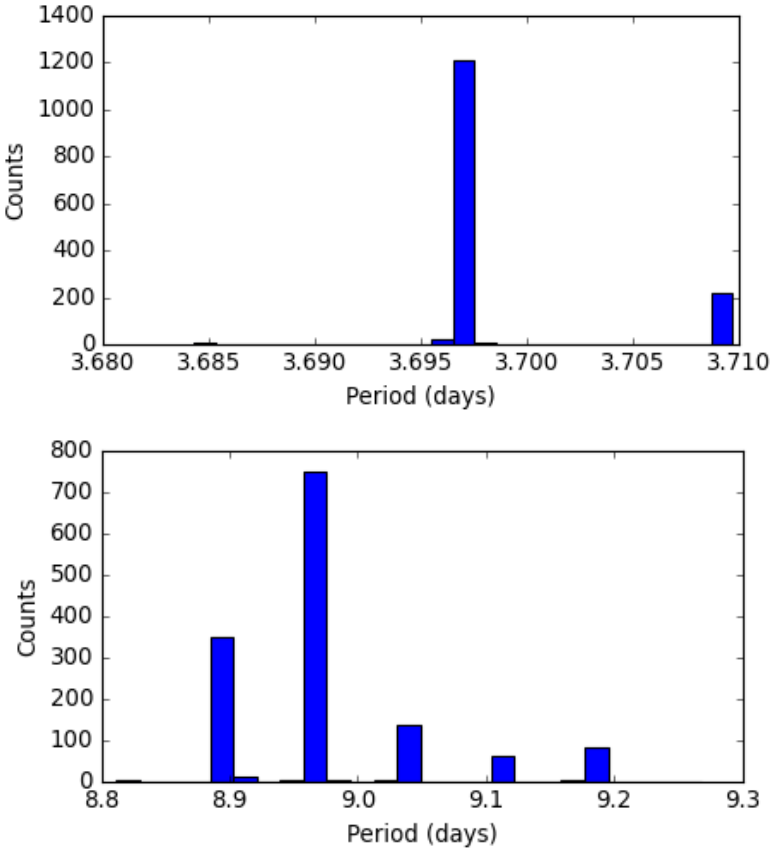


Figure A.3: The FIP histograms show the most probable periods for CoRoT-7c and CoRoT-7d from the corresponding interference patterns.

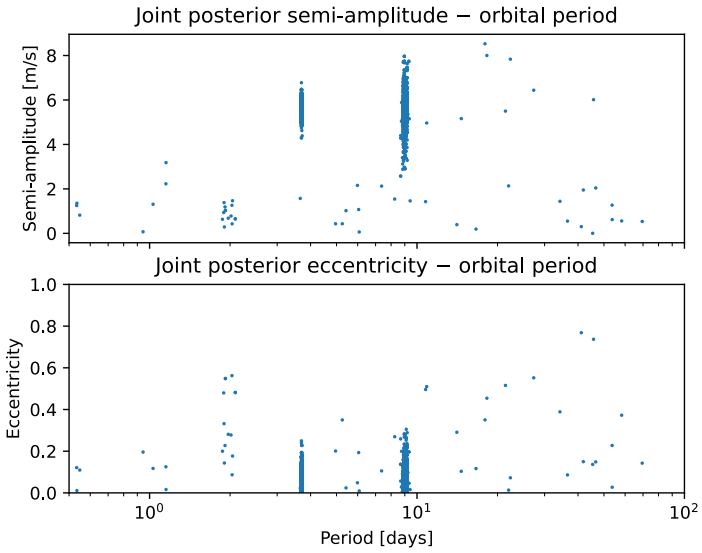


Figure A.4: Joint posterior distribution for the semi-amplitudes and eccentricities along with the orbital periods in the x-axes

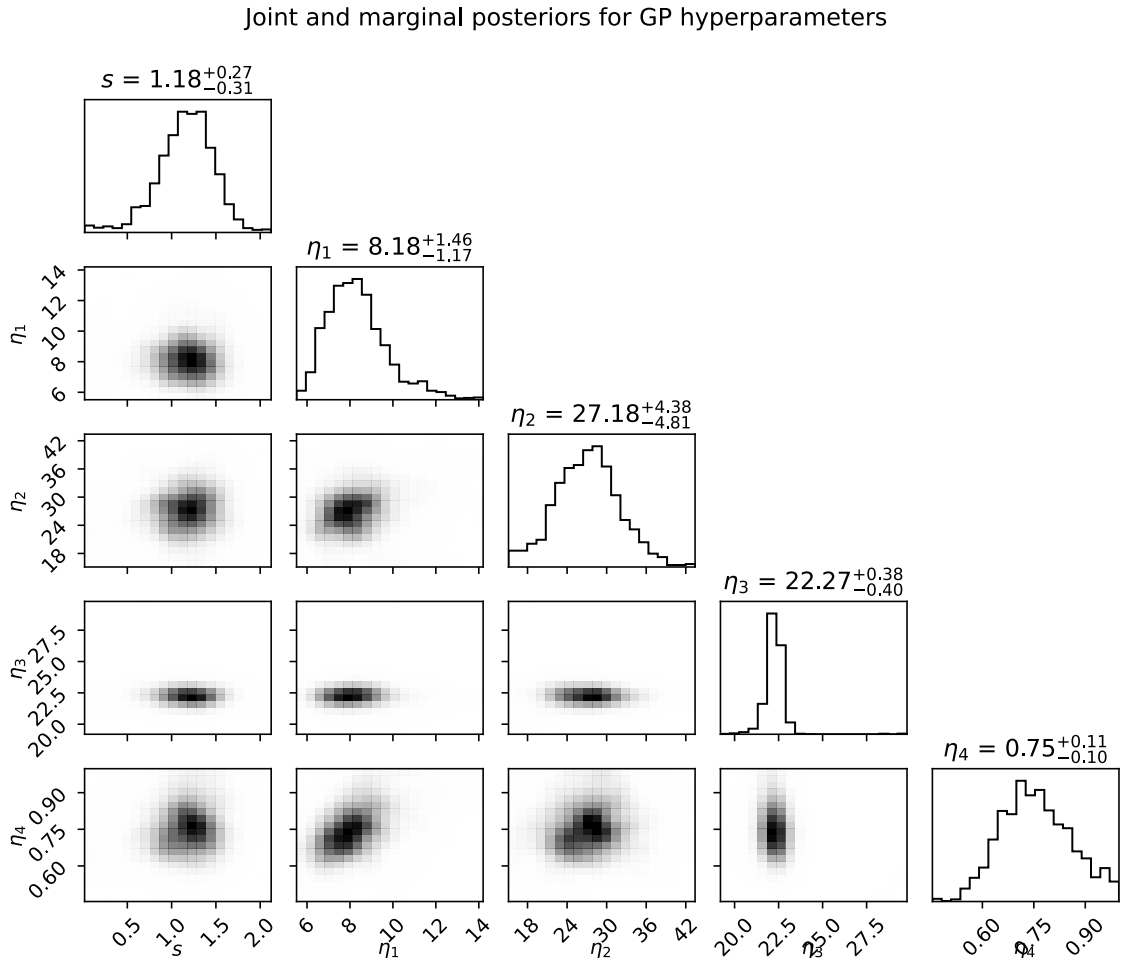


Figure A.5: Posterior distributions for the GP parameters and the extra white noise. The samples for all values of Np were combined. The median of the posterior and the uncertainties calculated from the 16% and 84% quantiles are given in respective titles.

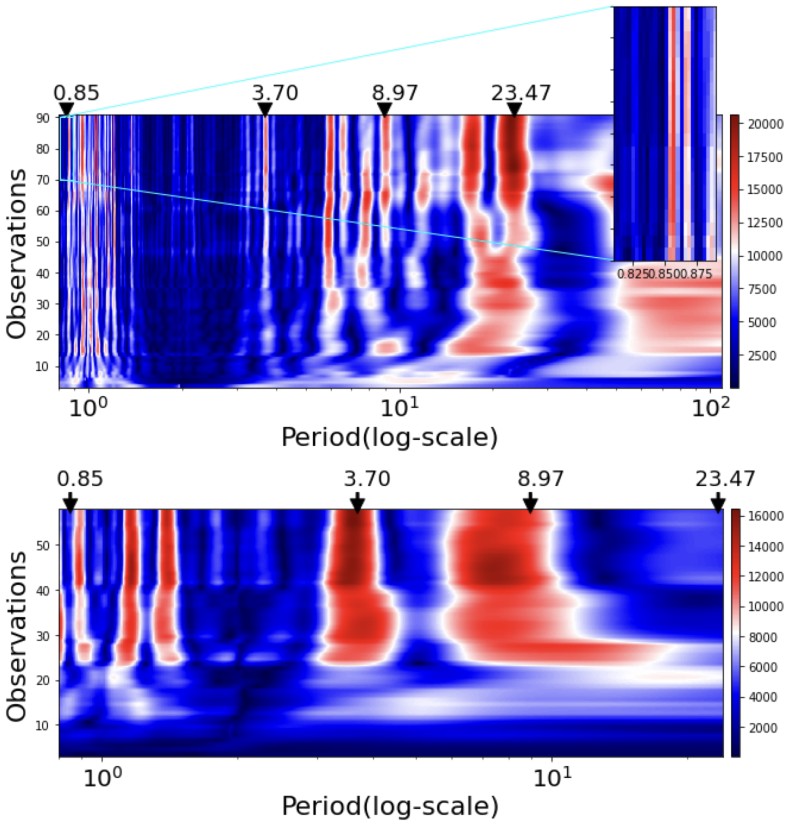


Figure A.6: SBGLS periodograms for the 2008-9 and 2012 data subsets are given in the left and right panels respectively.

B

Appendix B

This section includes some figures and tables that support the science described in Chapter 4. The are some FIP periodograms from the injection recovery tests, performed with injected Keplerians of varied orbital periods and RV semi-amplitudes. The correlation between the SCALPELS U-vectors and traditional activity indicators are also shown in two corner plots for individual stars.

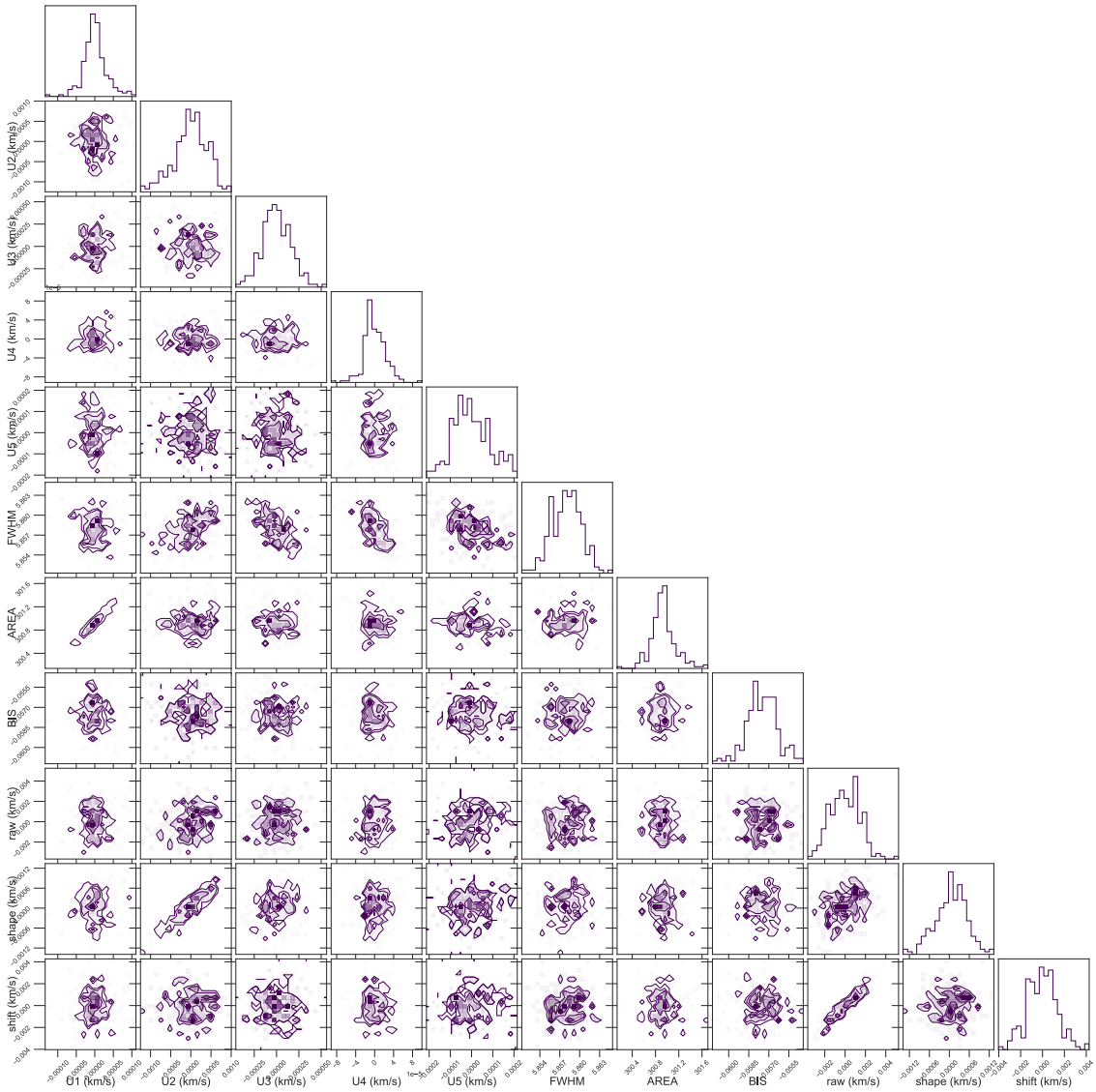


Figure B.1: The correlations plots for HD144579 showing the leading **U**-vector components of the residual CCF of RV time series and the traditional activity indicators such as FWHM, area and Bisector span. $\mathbf{U}_2$ is the prominent contributor to the entire shape component which shows a weak correlation with the FWHM, while the $\mathbf{U}_1$ component is strongly correlated with the area. $\mathbf{U}_3$ shows an anticorrelation with the FWHM.

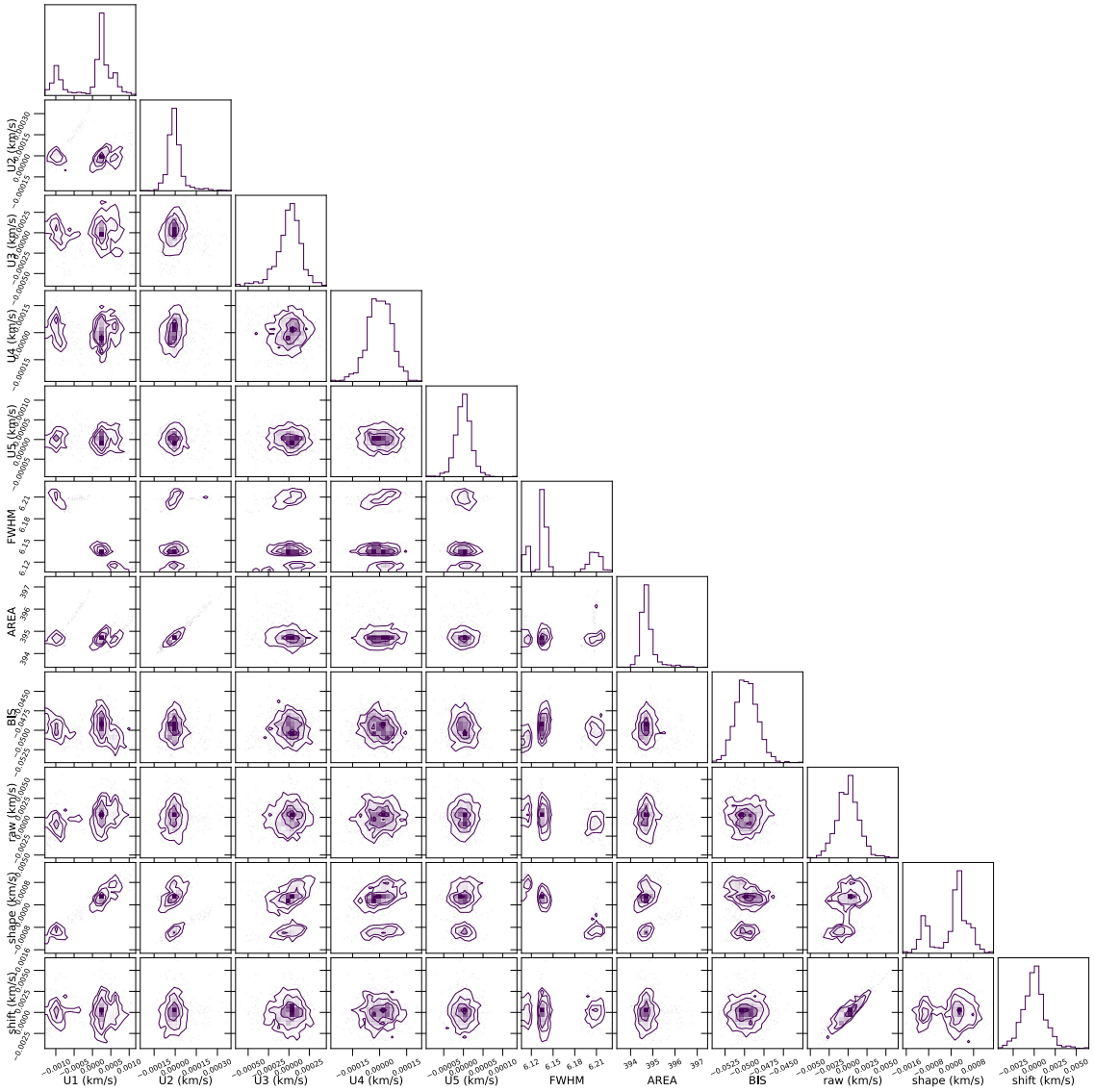


Figure B.2: The correlations plots for HD166620 showing the leading **U**-vector components of the residual CCF of RV time series and the traditional activity indicators such as FWHM, area and Bisector span. $\mathbf{U}_1$ is the prominent contributor to the entire shape component which shows a strong anti-correlation with the FWHM, while the $\mathbf{U}_2$ component is strongly correlated with the FWHM.

Table B.1: RPS targets used in the sample dataset to monitor the 1400-day periodicity and calculate the zero points. Note: HD 166620 is taken out from the list when estimating zero points for itself. Similarly, for HD 144579 as well.

Star	No. of observations	After nightly binning	Time span (days)
HD 10476	1062	257	4108
HD 122064	709	280	4187
HD 127334	1432	388	4023
HD 128165	473	268	3674
HD 144579	888	257	3646
HD 158633	626	167	3301
HD 166620	947	318	4080
HD 32147	505	139	4091
HD 3651	720	154	3558
HD 4628	1540	379	4099
HD 62613	527	157	3976
Sun	857		

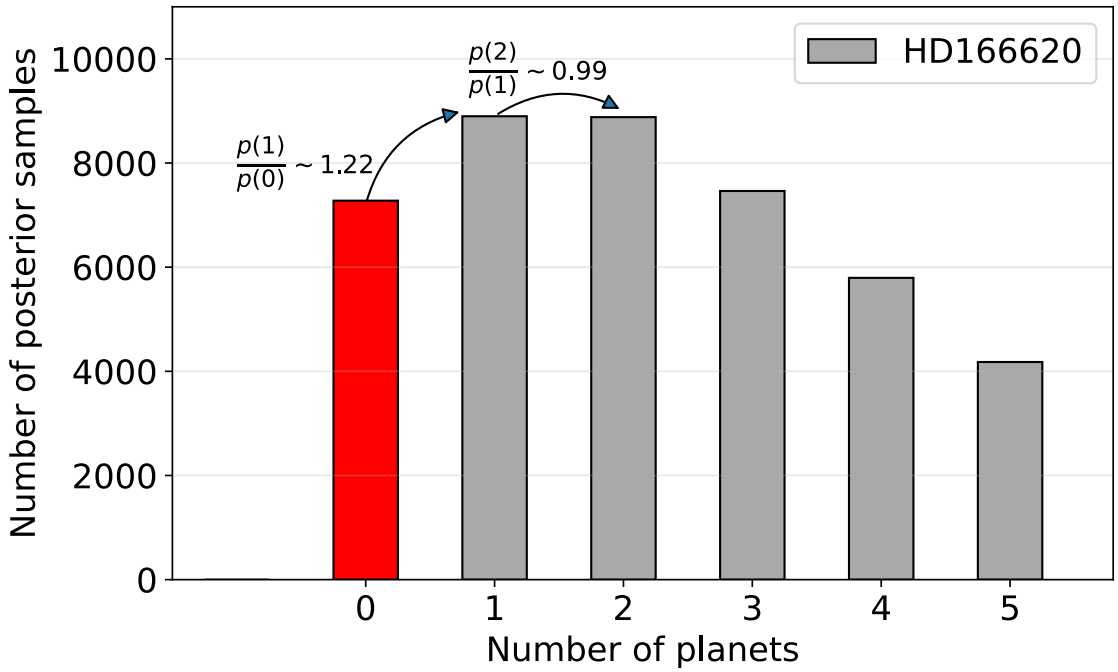


Figure B.3: The posterior distribution for HD 166620, for the number of planets Np. The counts are the number of posterior samples in trial models with a specific number of planets. The probability ratios between models with 0,1 and 2 planets are shown. The posterior distribution suggests that there are likely to be 'zero' planets present.

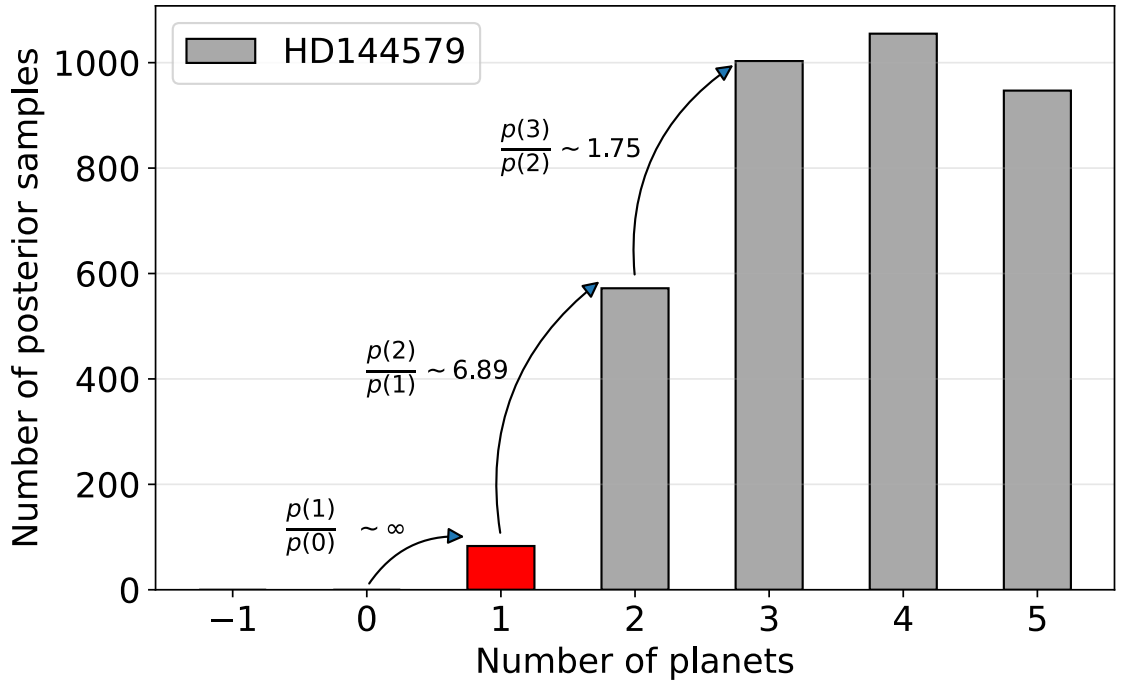


Figure B.4: A Keplerian with P=210.28 d and K=0.8 m s^{-1} is injected to the data. The ratio of probabilities for successive values of N_p now favours N_p=1.

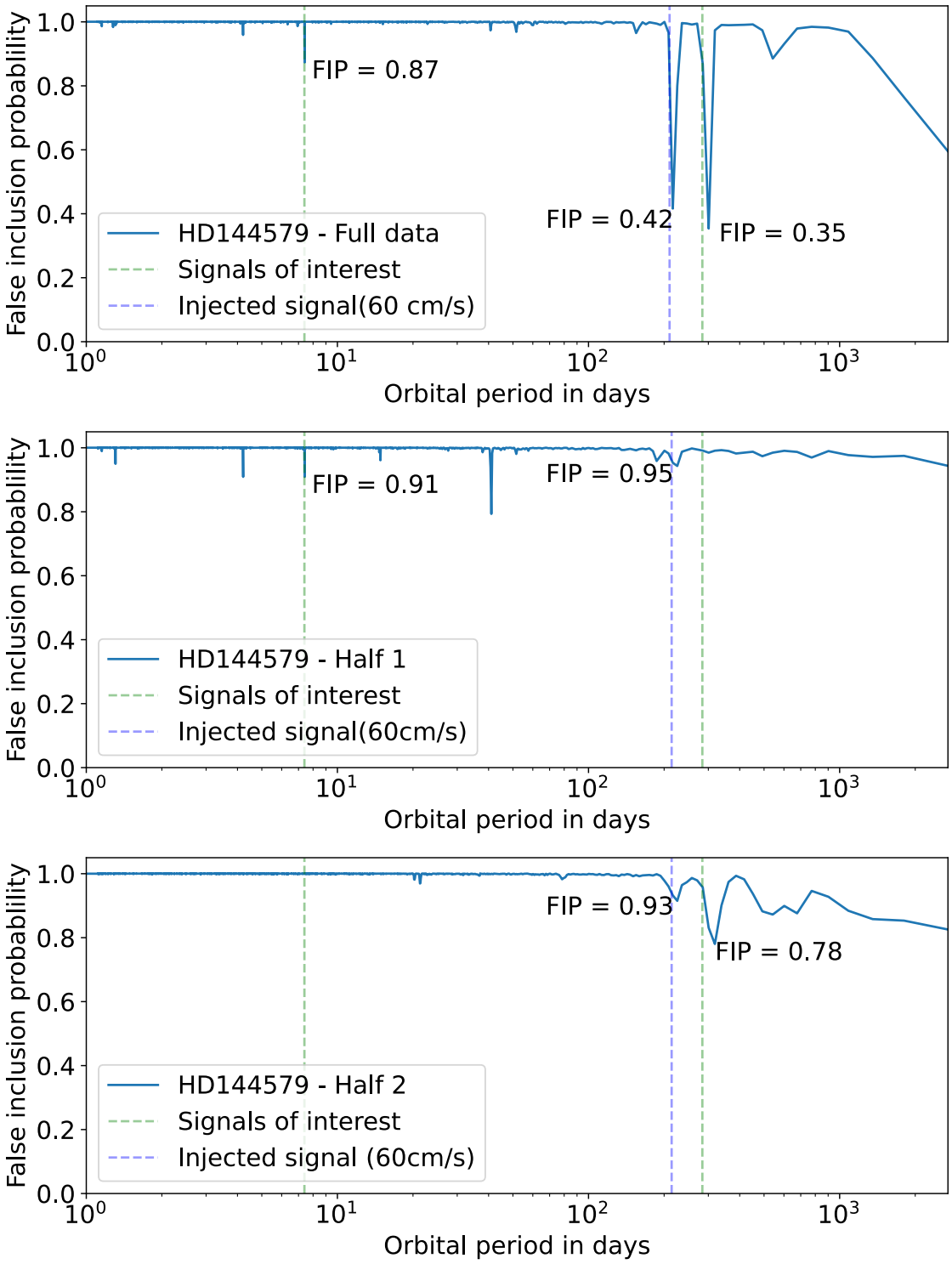


Figure B.5: A Keplerian with P=210.28 d and K=0.6 m s^{-1} is injected to the data. The observations are then divided into two subsets to investigate how a coherent signal appears in the individual halves. The FIP diagrams obtained for individual half subsets of the HD 144579 data are shown in the two panels. Unlike the signals of interest at P= 7.39 and 284 d, the injected Kepelerian is marginally detected in both data sets independently.

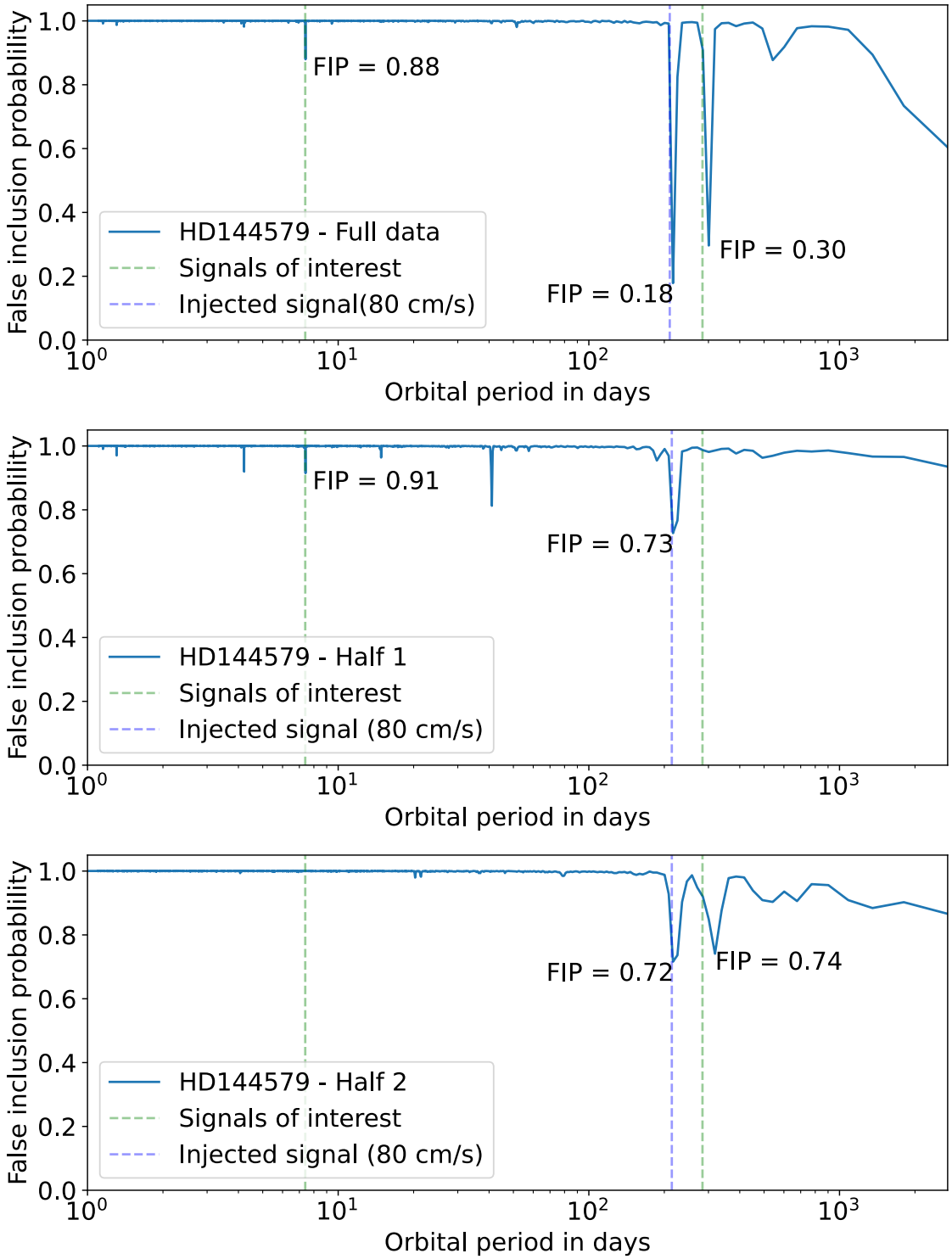


Figure B.6: A Keplerian with P=210.28 d and K=0.8 m s^{-1} is injected to the data. The FIP diagrams obtained for individual half subsets of the HD 144579 data are shown in the two panels. The injected Keplerian is now better detected in both data sets independently.

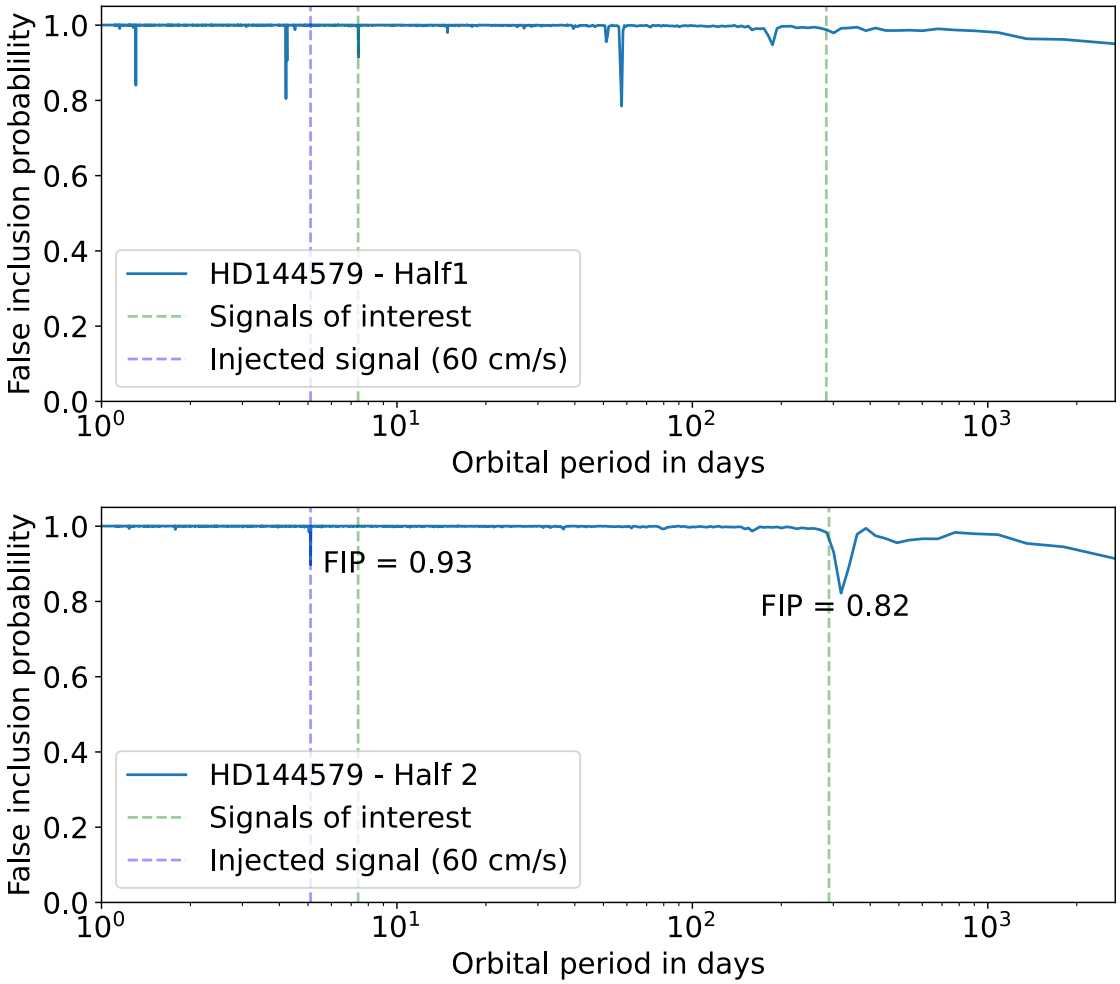


Figure B.7: A Keplerian with P=5.12 d and K=0.6 m s^{-1} is injected into the data. The FIP diagrams obtained for individual half subsets of the HD 144579 data are shown in the two panels. The injected Keplerian is only detected in one of the data sets.

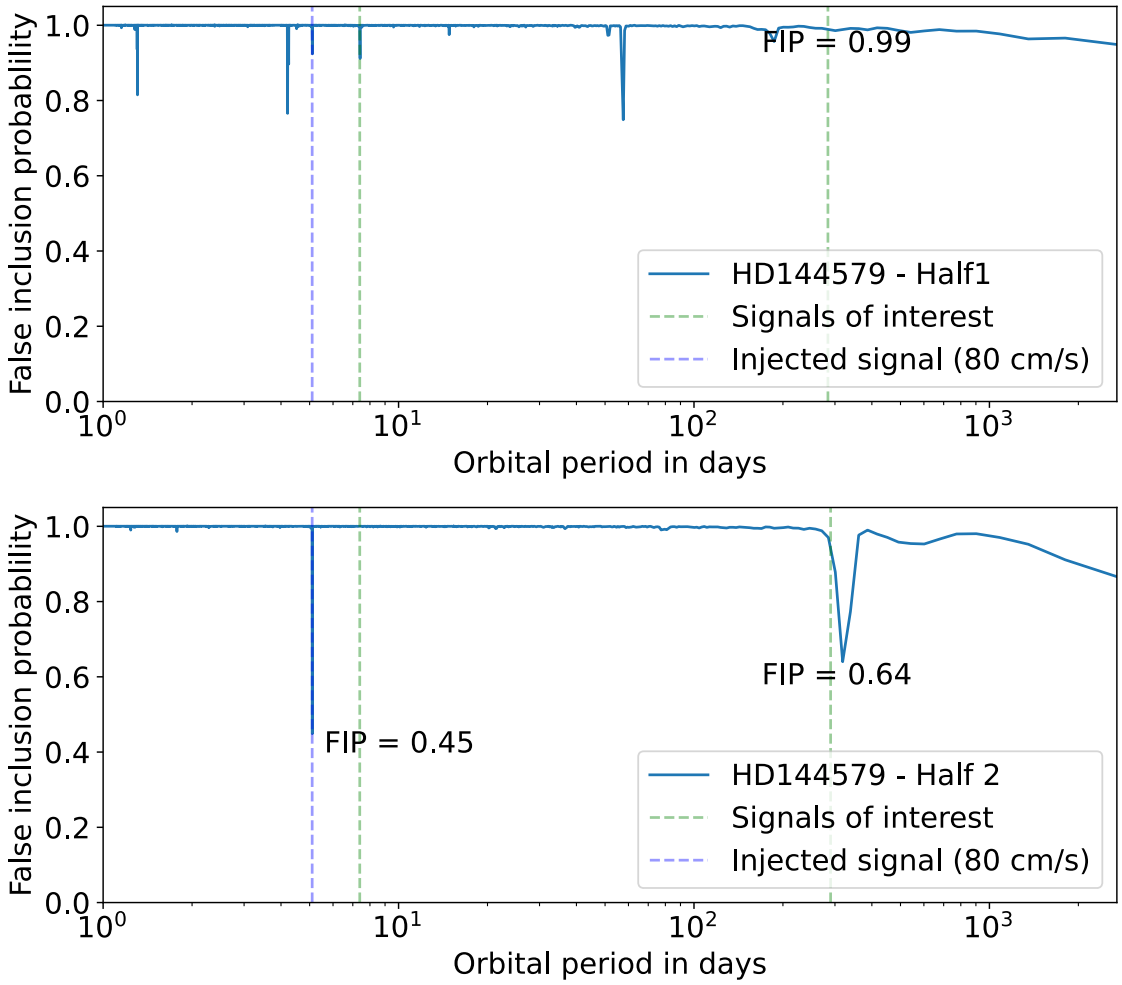


Figure B.8: A Keplerian with P=5.12 d and K=0.8 m s^{-1} is injected into the data. The FIP diagrams obtained for individual half subsets of the HD 144579 data are shown in the two panels. The injected Keplerian is now detected in both data sets independently, strongly in one half and marginally in the other half. This can be due to the cross-talk of the injected signal between the nearby 7.34-day signal and its strong 1-day alias possibly arising from the sampling pattern of the first half data set.

Bibliography

Adler, R. L., & Konheim, A. G. 1962, Proceedings of the American Mathematical Society, 13, 425

Aigrain, S., Pont, F., & Zucker, S. 2012, MNRAS, 419, 3147, 1110.1034

Al Moulla, K., Dumusque, X., Cretignier, M., Zhao, Y., & Valenti, J. A. 2022, A&A, 664, A34, 2205.07047

Anna John, A. et al. 2023, MNRAS, 525, 1687, 2308.01348

Anna John, A., Collier Cameron, A., & Wilson, T. G. 2022, MNRAS, 515, 3975, 2206.14216

Asplund, M., Nordlund, Å., Trampedach, R., Allende Prieto, C., & Stein, R. F. 2000, A&A, 359, 729, astro-ph/0005320

Auvergne, M. et al. 2009, A&A, 506, 411, 0901.2206

Bakos, G., Noyes, R. W., Kovács, G., Stanek, K. Z., Sasselov, D. D., & Domsa, I. 2004, PASP, 116, 266, astro-ph/0401219

Baliunas, S. L. et al. 1995, ApJ, 438, 269

Baranne, A. et al. 1996, A&AS, 119, 373

Barnes, S. A. 2007, ApJ, 669, 1167, 0704.3068

Barragán, O., Aigrain, S., Rajpaul, V. M., & Zicher, N. 2022, MNRAS, 509, 866, 2109.14086

Barros, S. C. C. et al. 2014, A&A, 569, A74, 1407.8099

Basri, G. et al. 2011, AJ, 141, 20, 1008.1092

Baum, A. C., Wright, J. T., Luhn, J. K., & Isaacson, H. 2022, AJ, 163, 183, 2203.13376

Blackman, R. T. et al. 2020, AJ, 159, 238, 2003.08852

Boisse, I., Bouchy, F., Hébrard, G., Bonfils, X., Santos, N., & Vauclair, S. 2011, in The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution, ed. A. Sozzetti, M. G. Lattanzi, & A. P. Boss, Vol. 276, 399–400

Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733

Brewer, B. J. 2014, arXiv e-prints, arXiv:1411.3921, 1411.3921

Brewer, B. J., Pártay, L. B., & Csányi, G. 2010, DNEST: Diffusive Nested Sampling, Astrophysics Source Code Library, record ascl:1010.029, 1010.029

Brown, T. M., Gilliland, R. L., Noyes, R. W., & Ramsey, L. W. 1991, ApJ, 368, 599

Buchhave, L. A. et al. 2012, Nature, 486, 375

Carrión-González, Ó., García Muñoz, A., Cabrera, J., Csizmadia, S., Santos, N. C., & Rauer, H. 2020, A&A, 640, A136, 2006.08784

Chaplin, W. J., Cegla, H. M., Watson, C. A., Davies, G. R., & Ball, W. H. 2019, AJ, 157, 163, 1903.00657

Chu, J.-H., Rogers, A., Ionita-Laza, I., Darvishi, K., Mills, R., Lee, C., & Raby, B. 2013, BMC bioinformatics, 14, 157

Collier Cameron, A. et al. 2021, MNRAS, 505, 1699, 2011.00018

——. 2019, MNRAS, 487, 1082, 1904.12186

Collier Cameron, A., Pollacco, D., Hellier, C., West, R., WASP Consortium, & SOPHIE & CORALIE Planet-Search Teams. 2009, in Transiting Planets, ed. F. Pont, D. Sasselov, & M. J. Holman, Vol. 253, 29–35

Cosentino, R. et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84461V

Crass, J. et al. 2021, arXiv e-prints, arXiv:2107.14291, 2107.14291

Cretignier, M., Dumusque, X., Allart, R., Pepe, F., & Lovis, C. 2020a, A&A, 633, A76, 1912.05192

Cretignier, M., Dumusque, X., Hara, N. C., & Pepe, F. 2021, A&A, 653, A43, 2106.07301

Cretignier, M., Francfort, J., Dumusque, X., Allart, R., & Pepe, F. 2020b, A&A, 640, A42, 2006.13098

Cutri, R. M. et al. 2003, VizieR Online Data Catalog, II/246

——. 2021, VizieR Online Data Catalog, II/328

de Beurs, Z. L., Vanderburg, A., Shallue, C. J., & Harps-N Collaboration. 2021, in American Astronomical Society Meeting Abstracts, Vol. 53, American Astronomical Society Meeting Abstracts, 332.04

Del Moro, D. 2004, A&A, 428, 1007

Delisle, J. B., Hara, N., & Ségransan, D. 2020, A&A, 638, A95, 2004.10678

Di Marcantonio, P. 2019, in Galactic Archaeology in the Gaia Era, 4

Dobbs-Dixon, I., Lin, D. N. C., & Mardling, R. A. 2004, ApJ, 610, 464, astro-ph/0408191

Dorren, J. D. 1987, ApJ, 320, 756

Dravins, D., Lindegren, L., & Nordlund, A. 1981, A&A, 96, 345

Dumusque, X. 2018, A&A, 620, A47, 1809.01548

Dumusque, X., Boisse, I., & Santos, N. C. 2014, ApJ, 796, 132, 1409.3594

Dumusque, X. et al. 2021, A&A, 648, A103, 2009.01945

Dumusque, X., Santos, N. C., Udry, S., Lovis, C., & Bonfils, X. 2011a, A&A, 527, A82, 1101.0954

Dumusque, X., Udry, S., Lovis, C., Santos, N. C., & Monteiro, M. J. P. F. G. 2011b, A&A, 525, A140, 1010.2616

Duncan, D. K. et al. 1991, ApJS, 76, 383

Eddy, J. A. 1976, Science, 192, 1189

Espinoza, N., Kossakowski, D., & Brahm, R. 2019, mnras, 490, 2262, 1812.08549

Faria, J. P. et al. 2020, A&A, 635, A13, 1911.11714

Faria, J. P., Haywood, R. D., Brewer, B. J., Figueira, P., Oshagh, M., Santerne, A., & Santos, N. C. 2016, A&A, 588, A31, 1601.07495

Faria, J. P., Santos, N. C., Figueira, P., & Brewer, B. J. 2018, The Journal of Open Source Software, 3, 487, 1806.08305

Faria, J. P. d. S. 2018, PhD thesis, University of Porto, Portugal

Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019a, The Open Journal of Astrophysics, 2, 10, 1306.2144

——. 2019b, The Open Journal of Astrophysics, 2, 10, 1306.2144

Ferraz, S., Tadeu Dos Santos, M., Beaugé, C., Michtchenko, T. A., & Rodríguez, A. 2011, A&A, 531, A161, 1011.2144

Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220, 1703.09710

Fulton, B. J. et al. 2017, AJ, 154, 109, 1703.10375

Gaia Collaboration et al. 2021, A&A, 650, C3

——. 2023, A&A, 674, A1, 2208.00211

Gray, D. F. 1992, Science, 257, 1978

——. 2005, The Observation and Analysis of Stellar Photospheres (Cambridge University Press)

——. 2009, ApJ, 697, 1032

Gregory, P. C. 2016, MNRAS, 458, 2604, 1601.08105

Guerrero, N. M. et al. 2021, ApJS, 254, 39, 2103.12538

Haas, M. R. et al. 2014, in American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts #223, 228.01

Hale, G. E. 1908, ApJ, 28, 315

Halverson, S. et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 99086P, 1607.05634

Hara, N. C., Boué, G., Laskar, J., & Correia, A. C. M. 2017, MNRAS, 464, 1220, 1609.01519

Hara, N. C., Delisle, J.-B., Unger, N., & Dumusque, X. 2022a, A&A, 658, A177, 2106.01365

Hara, N. C., Unger, N., Delisle, J.-B., Díaz, R. F., & Ségransan, D. 2022b, A&A, 663, A14, 2105.06995

Hathaway, D. H., Teil, T., Norton, A. A., & Kitiashvili, I. 2015, ApJ, 811, 105, 1508.03022

Hatzes, A. P. 2013, Astronomische Nachrichten, 334, 616, 1307.1246

——. 2014, A&A, 568, A84, 1407.0853

Hatzes, A. P. et al. 2010, A&A, 520, A93, 1006.5476

——. 2011, ApJ, 743, 75, 1105.3372

Haywood, R. D. 2016, Radial-velocity Searches for Planets Around Active Stars (Springer)

Haywood, R. D. et al. 2014, MNRAS, 443, 2517, 1407.1044

Howard, A. W. et al. 2011, ApJ, 726, 73, 1003.3444

Isaacson, H., & Fischer, D. 2010, ApJ, 725, 875, 1009.2301

Isaak, K. G., & Benz, W. 2019, Nature Astronomy, 3, 873

Jeffreys, H. 1961, The Theory of Probability, 3rd edn. (Oxford University Press)

Jenkins, J. M. et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9913, Software and Cyberinfrastructure for Astronomy IV, ed. G. Chiozzi & J. C. Guzman, 99133E

Jinglin Zhao., Ford, E. B., & Tinney, C. G. 2022, ApJ, 935, 75, 2201.03780

Johnson, D. R. H., & Soderblom, D. R. 1987, AJ, 93, 864

Kipping, D. M. 2013, MNRAS, 434, L51, 1306.4982

Klein, B. et al. 2022, MNRAS, 512, 5067, 2203.08190

Koch, D. G. et al. 2010, ApJ, 713, L79, 1001.0268

Kopp, G., & Lean, J. L. 2011, Geophys. Res. Lett., 38, L01706

Kron, G. E. 1947, PASP, 59, 261

Kumaraswamy, P. 1980, Journal of Hydrology, 46, 79

Lacedelli, G. et al. 2021, MNRAS, 501, 4148, 2009.02332

Lanza, A. F. et al. 2010, A&A, 520, A53, 1005.3602

Leamon, R., McIntosh, S., & Title, A. 2020, arXiv e-prints, arXiv:2012.15186, 2012.15186

Léger, A. et al. 2009, A&A, 506, 287, 0908.0241

Leighton, R. B. 1959, ApJ, 130, 366

Lovis, C., & Fischer, D. 2010, in Exoplanets, ed. S. Seager (University of Arizona Press), 27–53

Luhn, J. K., Wright, J. T., Henry, G. W., Saar, S. H., & Baum, A. C. 2022, ApJ, 936, L23, 2207.00612

Malavolta, L., Lovis, C., Pepe, F., Sneden, C., & Udry, S. 2017, MNRAS, 469, 3965, 1705.02343

Mantovan, G. et al. 2023, arXiv e-prints, arXiv:2310.16888, 2310.16888

——. 2022, MNRAS, 516, 4432, 2208.12276

Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D., & Doyon, R. 2008, Science, 322, 1348, 0811.2606

Mayor, M. et al. 2003, The Messenger, 114, 20

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24, 1402.5694

Meunier, N., Desort, M., & Lagrange, A. M. 2010, A&A, 512, A39, 1001.1638

Meunier, N., & Lagrange, A. M. 2019, A&A, 625, L6, 1904.09089

Meunier, N., Tkaczuk, R., & Roudier, T. 2007, A&A, 463, 745

Mishra, L., Alibert, Y., Udry, S., & Mordasini, C. 2023, A&A, 670, A69, 2301.02373

Mortier, A., & Collier Cameron, A. 2017, A&A, 601, A110, 1702.03885

Mortier, A., Santos, N. C., Sousa, S. G., Fernandes, J. M., Adibekyan, V. Z., Delgado Mena, E., Montalto, M., & Israelian, G. 2013, A&A, 558, A106, 1309.1998

Mortier, A., Sousa, S. G., Adibekyan, V. Z., Brandão, I. M., & Santos, N. C. 2014, A&A, 572, A95, 1410.1310

Mortier, A. et al. 2020, MNRAS, 499, 5004, 2010.01993

Morton, T. D. 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library, record ascl:1503.010, 1503.010

Motalebi, F. et al. 2015, A&A, 584, A72, 1507.08532

Nordlund, Å., Stein, R. F., & Asplund, M. 2009, Living Reviews in Solar Physics, 6, 2

Pagano, I. 2013, in Planets, Stars and Stellar Systems. Volume 4: Stellar Structure and Evolution, ed. T. D. Oswalt & M. A. Barstow, Vol. 4 (Springer), 485

Palethorpe, L. et al. 2024, arXiv e-prints, arXiv:2403.04464, 2403.04464

Pepe, F. et al. 2021, A&A, 645, A96, 2010.00316

Perryman, M. 2018, Transits, 2nd edn. (Cambridge University Press), 153–328

Phillips, D. F. et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9912, Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II, ed. R. Navarro & J. H. Burge, 99126Z

Pollacco, D. L. et al. 2006, PASP, 118, 1407, astro-ph/0608454

Pont, F., Aigrain, S., & Zucker, S. 2011, MNRAS, 411, 1953, 1008.3859

Priest, E. R. 1982, Solar magneto-hydrodynamics., Vol. 21 (Cambridge University Press)

Quanz, S., Carrion-Gonzalez, O., Noack, L., Lichtenberg, T., & LIFE Initiative. 2023, in AAS/Division for Planetary Sciences Meeting Abstracts, Vol. 55, AAS/Division for Planetary Sciences Meeting Abstracts, 311.12

Queloz, D. et al. 2009, A&A, 506, 303

——. 2001, A&A, 379, 279, astro-ph/0109491

Quirrenbach, A. et al. 2013, in American Astronomical Society Meeting Abstracts, Vol. 221, American Astronomical Society Meeting Abstracts #221, 149.05

Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269, 1506.07304

Reddy, B. E., Lambert, D. L., & Allende Prieto, C. 2006, MNRAS, 367, 1329, astro-ph/0512505

Rescigno, F. et al. 2023, arXiv e-prints, arXiv:2310.13623, 2310.13623

Ricker, G. R. et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Sairam, L., & Triaud, A. H. M. J. 2022, MNRAS, 514, 2259, 2205.11187

Salabert, D. et al. 2016, A&A, 596, A31, 1608.01489

Santerne, A. et al. 2012, A&A, 545, A76, 1206.0601

Santos, N. C., Gomes da Silva, J., Lovis, C., & Melo, C. 2010, A&A, 511, A54, 0912.2901

Santos, N. C. et al. 2014, A&A, 566, A35, 1404.6135

Schrijver, C. J., & Zwaan, C. 2000, Cambridge Astrophysics Series, 34

Schwabe, H. 1844, Astronomische Nachrichten, 21, 233

Shah, S. P., Wright, J. T., Isaacson, H., Howard, A. W., & Curtis, J. L. 2018, ApJ, 863, L26, 1801.09650

Silva, A. M. et al. 2022, Astronomy & Astrophysics, 663, A143

Skilling, J. 2004, AIP Conference Proceedings, 735, 395

——. 2006, Bayesian Analysis, 1, 833

Soderblom, D. R. 1985, AJ, 90, 2103

Sousa, S. G. 2014, ARES + MOOG: A Practical Overview of an Equivalent Width (EW) Method to Derive Stellar Parameters (Springer International Publishing), 297–310

Sousa, S. G., Santos, N. C., Israelian, G., Lovis, C., Mayor, M., Silva, P. B., & Udry, S. 2011, A&A, 526, A99, 1012.1528

Spina, L. et al. 2020, ApJ, 895, 52, 2004.09771

Stalport, M. et al. 2023, arXiv e-prints, arXiv:2308.05669, 2308.05669

Standing, M. R. et al. 2023, Nature Astronomy, 7, 702, 2301.10794

——. 2022, MNRAS, 511, 3571, 2112.05652

Struve, O. 1952, The Observatory, 72, 199

Trotta, R. 2008, Contemporary Physics, 49, 71, 0803.4089

Tuomi, M., Anglada-Escude, G., Jenkins, J. S., & Jones, H. R. A. 2014, arXiv e-prints, arXiv:1405.2016, 1405.2016

Ulrich, R. K. 1970, ApJ, 162, 993

Vaughan, A. H. 1980, PASP, 92, 392

Vaughan, A. H., Baliunas, S. L., Middelkoop, F., Hartmann, L. W., Mihalas, D., Noyes, R. W., & Preston, G. W. 1981, ApJ, 250, 276

Vaughan, A. H., Preston, G. W., & Wilson, O. C. 1978, PASP, 90, 267

Wenger, M. et al. 2000, A&AS, 143, 9, astro-ph/0002110

Wilson, O. C. 1978, ApJ, 226, 379

Wright, J. T. 2004, AJ, 128, 1273, astro-ph/0406338

——. 2005, PASP, 117, 657, astro-ph/0505214

Yi, S., Demarque, P., Kim, Y.-C., Lee, Y.-W., Ree, C. H., Lejeune, T., & Barnes, S. 2001, ApJS, 136, 417, astro-ph/0104292

Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577, 0901.2573

Zeng, L. et al. 2021, ApJ, 923, 247, 2201.02125

Zeng, L., Sasselov, D. D., & Jacobsen, S. B. 2016, ApJ, 819, 127, 1512.08827

Zhao, J., & Tinney, C. G. 2020, MNRAS, 491, 4131

Zhao, L. L. et al. 2022, AJ, 163, 171, 2201.10639