CONSTANT AS THE STARS ABOVE

or how the stars are in fact not constant at all

Understanding and mitigating stellar variability for the detection of exoplanets

Federica Rescigno

Submitted by Federica Rescigno to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Physics, March 2024.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Signed: ....

Federica Rescigno

Date: 26.05.2024

Abstract

1st Supervisor: Dr. Raphaëlle D. Haywood   2nd Supervisor: Prof. Tim Naylor

In the last 20 years, the radial velocity (RV) method has successfully detected and characterised hundreds of exoplanets, from blazing hot giants to small super-Earths. With the recent sub-m s^{-1} precision reached by ultra-stable spectrographs, the signals of rocky extra-solar planets in long-orbits have finally become detectable. The greatest challenge for the characterisation of exoplanets is now stellar variability. The effects of activity on the surface of stars often strongly dominate the RV budget, and can easily obscure or mimic Keplerian signals. Gaussian Processes (GPs) have been proven to be a very successful technique for the mitigation of stellar effects, as they are able to model the variability without making any assumption about its functional form. In this work, I introduce MAGPy-RV. MAGPy-RV is a Gaussian process regression pipeline with Markov Chain Monte Carlo parameter space searching algorithm I developed in the context of exoplanet detection and characterisation. It allows to simultaneously model stellar activity, described by a GP with the chosen covariance function, and Keplerian signals in the RVs as well as transits in photometric data.

I then use this pipeline for the analysis of two planetary systems: TOI-2134 and HD 48948. The moderately active, bright K5V star TOI-2134 is orbited by an inner mini-Neptune in a 9.2292005±0.0000063 day orbit and an outer mono-transiting sub-Saturn planet in a $95.50^{+0.36}_{-0.25}$ day orbit. Based on the analysis of TESS data, I determine the radii of TOI-2134b and c to be 2.69±0.16 $R_{\oplus}$ for the inner planet, and 7.27±0.42 $R_{\oplus}$ for the outer one. The masses of both planets are derived based on HARPS-N and SOPHIE RVs via Gaussian process regression to be $9.13^{+0.78}_{-0.76}$ $M_{\oplus}$ for TOI-2134b and $41.89^{+7.69}_{-7.83}$ $M_{\oplus}$ for TOI-2134c. The outer planet is computed to have a significant eccentricity of $0.67^{+0.05}_{-0.06}$ from a combination of photometry and RVs.

The HD 48948 system comprises of three super-Earth planetary candidates with orbital periods of roughly 7.3, 38, and 151 days, and minimum masses estimated to be 4.96 ± 0.42 $M_{\oplus}$,

7.45 ± 0.75 $M_\oplus$, and 10.67 ± 0.90 $M_\oplus$, respectively. The outermost planet is also found to reside within the (temperate) habitable zone, positioned at a projected distance of 0.029″ from its star.
Both these analyses highlighted the need for caution and systematic model testing when employing GPs to model stellar-induced signals.

In parallel, I also analyse solar data in order to develop a better understanding of the processes driving stellar variability. The Sun is a fairly representative star of the sample of targets that are generally selected for RV surveys, and most crucially it is the only star we can resolve. I use the SolAster pipeline to derive disc-integrated longitudinal magnetic field data in order to asses its uses in radial-velocity surveys. I show that the mean longitudinal magnetic field is an excellent rotation period detector and a useful tracer of the solar magnetic cycle. In order to put these results into context, I compare the mean longitudinal magnetic field to three common activity proxies derived from HARPS-N Sun-as-a-star data: the full-width at half-maximum, the bisector span and the S-index. I find that the mean longitudinal magnetic field cannot be used as a one-to-one proxy, but that it outperforms all other considered indicators as a solar rotational period detector, and can be used to inform our understanding of the physical processes happening on the surface of the Sun.

Contents

List of Figures

List of Tables

Declaration

This thesis includes the work of multiple published papers, as well as on-going investigations.

- Chapter 1 and Chapter 2 are inspired by work undertaken for the Terra Hunting Experiment Spring 2023 Consortium, done by the author, A. Mortier and B. Klein. The output of this effort is a table of current mitigation technique and which source of stellar variability they aim to correct.

- Chapter 3 made use of the following resources: Rasmussen and Williams (2006), Roberts et al. (2013), Hara and Ford (2023), and Wang (2023). The Chapter also introduced the pipeline MAGPY-RV I developed for GP regression. This software was recorded at the Astrophysics Source Code Library under the record of ascl:2310.006, and at the Exoplanet Modelling and Analysis Center with the record of emac:2311.006. It can be found under the title of *MAGPy-RV: a Gaussian Process regression pipeline with MCMC parameter searching algorithm* by F. Rescigno, B. Dixon, and R. D. Haywood published in 2023.

- Chapter 4 includes the work of two published papers.
  Section 4.1 is directly taken from *A hot mini-Neptune and a temperate, highly eccentric sub-Saturn around the bright K-dwarf TOI-2134* published in 2024 under the Monthly Notices of the Royal Astronomical Society, V.527, I.3, pp 5384-6407 (doi:10.1093/mnras/stad3255) by F. Rescigno, G. Hébrard., A. Vanderburg, A. W. Mann, A. Mortier, S. Morrell, L. A. Buchhave, K. A. Collins, C. R. Mann, C. Hellier, R. D. Haywood, R. West, M. Stalport, N. Heidari, D. Anderson, C. X. Huang, M.

López-Morales, P. Cortés-Zuleta, H.M. Lewis, X. Dumusque, I. Boisse, P. Rowden, A. Collier Cameron, M. Deleuil, M. Vezie, F. A. Pepe, X. Delfosse, D. Charbonneau, K. Rice, O. Demangeon, S. N. Quinn, S. Udry, T. Forveille, J. N. Winn, A. Sozzetti, S. Hoyer, S. Seager, T. G. Wilson, S. Dalal, E. Martioli, S. Striegel, W. Boschin, D. Dragomir, A. F. Martínez Fiorenzano, R. Cosentino, A. Ghedina, L. Malavolta, L. Affer, B. S. Lakeland, B. A. Nicholson, S. Foschino, A. Wünsche, K. Barkaoui, G. Srdoc, J. Randolph, B. Guillet, D. M. Conti, M. Ghachoui, M. Gillon, Z. Benkhaldoun, F. J. Pozuelos, M. Timmermans, E. Girardin, S. Matutano, P. Bosch-Cabot, J. A. Muñoz, and R. Forés-Toribio. For this work I directed the collaboration, and wrote the paper. I undertook all the analysis personally with the following exceptions: in Section 4.1.2.1 the second processing of the TESS data was done by A. Vanderburg, the stellar characterisation in Section 4.1.3 was done by A. Mann, A. Mortier, L. Buchhave and S. Morrell, the first transit analysis in Section 4.1.5 with Fig. 4.8 was done by A. Vanderburg, and the chaos studies included in Section 4.1.8.1 and Fig. 4.13 were completed by M. Stalport.

Section 4.2 is extracted form the work *Trio of super-Earth candidates orbiting K-dwarf HD 48948: A new habitable zone candidate* currently submitted to the Monthly Notices of the Royal Astronomical Society with the following author list: S. Dalal, F. Rescigno, M. Cretignier, A. A. John, F. Z. Majidi, L. Malavolta, A. Mortier, M. Pinamonti, L. A. Buchhave, R. D. Haywood, A. Sozzetti, F. Lienhard, K. Rice, A. Vanderburg, B. Lakeland, L. Affer, A. S. Bonomo, W. Boschin, A. Collier Cameron, B. Cooke, R. Cosentino, M. Damasso, X. Dumusque, L. Di Fabrizio, D. W. Latham, M. López-Morales, A. F. Martínez Fiorenzano, M. Mayor, A. Harutyunyan, C. Lovis, M. Stalport, S. Udry, T. G. Wilson and C. A. Watson. I was heavily involved with this work from its inception. I directly produced and wrote the following sections for the paper: the activity indicator Gaussian process regression analysis in Section 4.2.6, and the entirety of Sections 4.2.7.1 and 4.2.7.2. For this thesis, I have personally summarised the other analysis and completely reworked Section 4.2.8. This analysis was lead by S. Dalal, work in Section 4.2.3 was undertaken by A. Mortier, L. Buchhave and L. Malavolta, the data extraction in Section 4.2.4 was done by M. Cretignier and A. A. John, the analyses in Sections 4.2.7.3 and 4.2.7.4 were completed by L. Malavolta and F. Z.

Majidi, and A. A. John respectively.

- Chapter 5 includes work published in two papers. The pipeline described in Section 5.1 has been published as *Leveraging space-based data from the nearest Solar-type star to better understand stellar activity signatures in radial velocity data* in The Astronomical Journal, Volume 163, Issue 6, id.272, 16 pp (doi:10.3847/ 1538-3881/ac67e6) by T. Ervin, S. Halverson, A. Burrows, N. Murphy, A. Roy, R. D. Haywood, F. Rescigno, C. F. Bender, A. S. J. Lin, J. Burt, Jennifer and S. Mahadevan. SolAster is registered at the Astrophysics Source Code Library with the record ascl:2207.009. I have been involved in the publication of this software. For this work I focus on explaining the characteristics of the pipeline within the context of my work and analyses.

  Section 5.2 is directly lifted from my paper *The Mean Longitudinal Magnetic Field and its Uses in Radial-Velocity Surveys* currently submitted to the Monthly Notices of the Royal Astronomical Society with the co-author list: F. Rescigno, A. Mortier, X. Dumusque, B. S. Lakeland, R. Haywood, N. Piskunov, M. López-Morales, S. Dalal, M. Cretignier, B. Klein, A. Collier Cameron, A. Ghedina, M. Gonzales and B. Nicholson. I led this project and wrote the paper. The only analysis I did not personally complete, Section 5.2.3.2 and Fig. 5.10, was done by B. Lakeland.

Acknowledgements

This thesis is the culmination of three and a half years of work, and it couldn't have been possible without the help, the guidance and the support of the amazing people around me.

First and foremost, I want to thank my family, who always encouraged my dream of becoming a scientist since childhood.
Thank you papá, for your help defeating trigonometry when I got my first bad grade in maths, for proof-reading all my reports in high-school and university, and for learning a completely new coding language when I was first struggling with Python. Can you believe I am publishing actual professional code now?
Thank you mamma, for always encouraging me when I was feeling at the lowest, for the trips and visits you always made sure to schedule for us whenever you could when I was living far away, and for the hours-long Facetime calls when I was living on my own. I cannot wait to show the life I made for myself when you come up next month.
Thank you Chiara, for feeding me throughout the entirety of the pandemic, and for being the greatest supporter of my love life.
Thank you to the rest of my family, who always celebrate my achievements, even when they don't really make sense to them.
Thank you Simon, for coming into my life, for being my greatest cheerleader and believer, and for keeping me alive by feeding me when my mind is 100 light years away. You have supported me throughout this process and I would not have been able to do this without you. Soon I will be a Doctor too, so get all the "because the Doctor said so" jokes out of your system before it's too late!

I also want to acknowledge the enormous help I received from my supervisor, Dr. Raphaëlle Haywood. Thank you for believing in me every time I came to you with a wacky idea, for welcoming me with open arms, and for introducing me to so many other amazing people. I am incredibly grateful for your guidance and support throughout my PhD. I learnt a lot from you, and I look back fondly to the effort you made to see me in person when I had just started during the pandemic, those picnic tables at the park had never seen so much science!
Thank you to the small Exeter EPRV group, who listened to my science problems all the time. Thank you for all the conference trips, and the journal clubs. Thank you for dealing with me all the times I poked my head in your offices asking for help.

To Prof. Nathan Mayne, you have been an aspiration to me. Thank you for all the serious future and career talks, for your mentoring, for being someone I could always turn to no matter the problem, and for always finding time for me in between your crazy busy days. I don't think you are quite aware of the impact you've had on my journey, but it cannot be understated.
To Dr. Annelies Mortier, thank you for being so incredibly friendly and welcoming. Thank you for believing in me, and I am so incredibly excited to get the chance to work with you in the next years.

Thank you to the friends around me, for all the girl dinner dates, the coffee shop chats, and the Star Wars game nights. Thank you Francesco, for the D&D Wednesdays, for always being willing to listen, and for being such an amazing friends even 1,500 km away.

To everyone whose journey intersected with mine, thank you. I will bring you all with me as I step forward in the next chapter of my life.

Federica Rescigno
Exeter, United Kingdom
31st March 2024

"Constant as the stars above

Always know that you are loved

And my love shining in you

Will help you make your dreams come true."

— Constant as the stars above,

Barbie as Rapunzel

Chapter 1

Stellar Variability in the Extreme Precision Radial-Velocity Regime

"Exploration is wired into our brains. If we can see the horizon, we want to know what's beyond."

— Buzz Aldrin, Magnificent Desolation: The Long Journey Home from the Moon

1.1 Introduction

Humans have been longing for the stars since the beginning of time. Questions such as "are we alone in the universe?" are a driving force in both science and general culture. Historians report that the first person to put to text the hypothesis of other worlds in the cosmos was Anaximander. The pre-Socratic Greek philosopher born in 610 BC in modern Turkey was one of the first to attempt to explain the cosmos with non-mythological methods. He stated that our planet is a free-floating object in space, and that Earth simply could not be the only one of such bodies in the universe. This thought was followed by his contemporary Anaximenes of Miletus, who was the first to differentiate between the burning Sun and the extinguished Earth in the categories we now refer to as stars and planets. In the midst of the Copernican revolution, Giordano Bruno, an Italian philosopher born in 1548 near Neaples, proposed the revolutionary concept that far-away stars are Suns of their own systems, and that just like Earth was orbiting our Sun there would be planets orbiting those seemingly fixed stars. In his third dialogue "De l'infinito universo et mondi" (of the infinite universe and worlds) he wrote:

> *"Io credo et intendo che oltre et oltre quella margine imaginata del cielo, sempre sia eterea regione, e corpi mondani, astri, terre, soli: e tutti sensibili absolutamente, secondo sé et a quelli che vi sono dentro o da presso: benché non sieno sensibili a noi per la lor lontanza e distanza."*
> I believe and I understand that beyond this imagined edge of the sky, there will always be an ethereal region, and worlds, stars, earths, suns: and all inherently perceptible one to another, to those which are within or near; though they may not be perceptible to us due to the distance.

He continues:

> *"Di maniera che non é un sol mondo, una sola terra, un solo sole: ma tanti son mondi, quante veggiamo circa di noi lampade luminose"*
> In such way that there is not a single world, a single earth, a single sun: but there are many worlds, as many shining lamps we see around us.

Other philosophers and later scientists supported the idea that, just like the Sun, other stars would host their own planetary systems. This idea was further reinforced as we started to understand more about how the Solar System formed and we began to see the signs of opportunity in the universe. Although claims of exoplanets detected with astrometry were made before then (e.g., Jacob 1855; See 1896), Struve (1952) was the first to propose a specific plan for the detection of new worlds. He asserted that planets around other stars would orbit at similar distances as solar planets do. They therefore should be detectable by observing the motion of the star around the common centre of mass via Doppler spectroscopy, or by monitoring the stellar luminosity in search of periodic dimmings due to planetary obstruction. After 40 years of waiting, the predictions of Otto Struve came true. The first hesitant discovery of a huge planet orbiting a star was published by Campbell et al. (1988), using radial-velocity measurements. These results were controversial and, although they were confirmed in 2003 with further observations, at the time of publication the community instead believed that the detected body was a brown

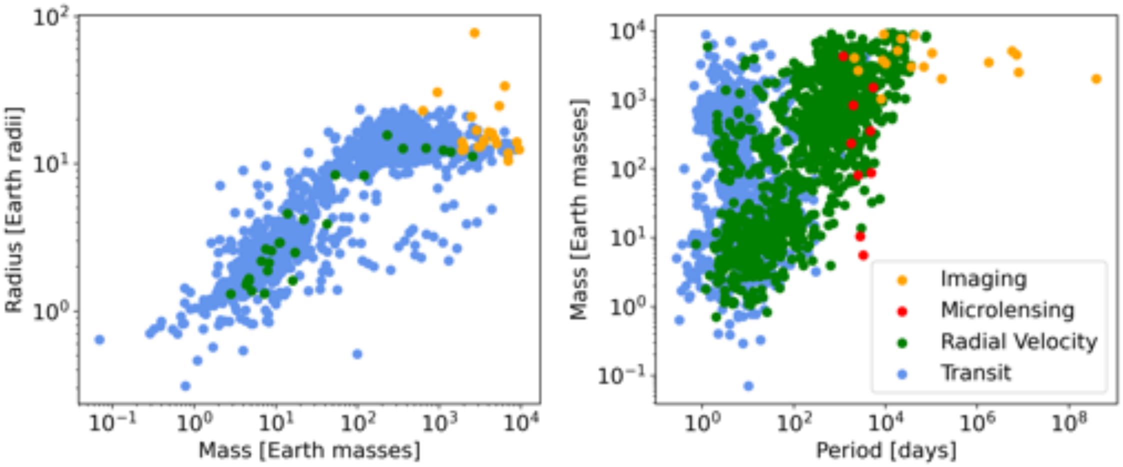

Figure 1.1: *Left panel:* Mass-Radius diagram of confirmed exoplanets with detected mass in Earth masses and radius in Earth radii, colour-coded by the technique of first detection. *Right panel:* Period-Mass diagram of confirmed exoplanets with detected mass in Earth masses and orbital period in days, colour-coded by the technique of first detection. Data acquired from the Exoplanet Archive at exoplanetarchive.ipac.caltech.edu on 2023-03-14.

dwarf rather than an exoplanet. The first accepted detection of an extrasolar planet was published by Wolszczan and Frail (1992). They detected a super-Earth orbiting the pulsar PSR 1257+12 by measuring changes in the pulsar period. This discovery fuelled groups of scientists around the word and soon after with new radial-velocity data, Mayor and Queloz (1995) presented the detection and characterisation of the first exoplanet orbiting a main-sequence star: 51 Pegasi b.

1.1.1 How to Detect an Exoplanet

As of March 2024, the community has characterised more than 5,599 exoplanets in 4,200 planetary systems, with 10,000 more candidates in the process of being confirmed. The first panel of Fig. 1.1 shows all the confirmed exoplanets with computed masses and radii, colour-coded based on the technique employed for first detection, while the second panel shows the Period-Mass diagram of all detected planets with known mass and period. Of those, the great majority were detected using the same techniques proposed in the ’50s by Otto Struve: radial velocity and transit photometry.

1.1.1.1 Transit Photometry

If the system is favourably aligned to our line-of-sight, a planet orbiting a star will transit in front of its host’s bright disc. This partial obscuring of a section of the star will cause a

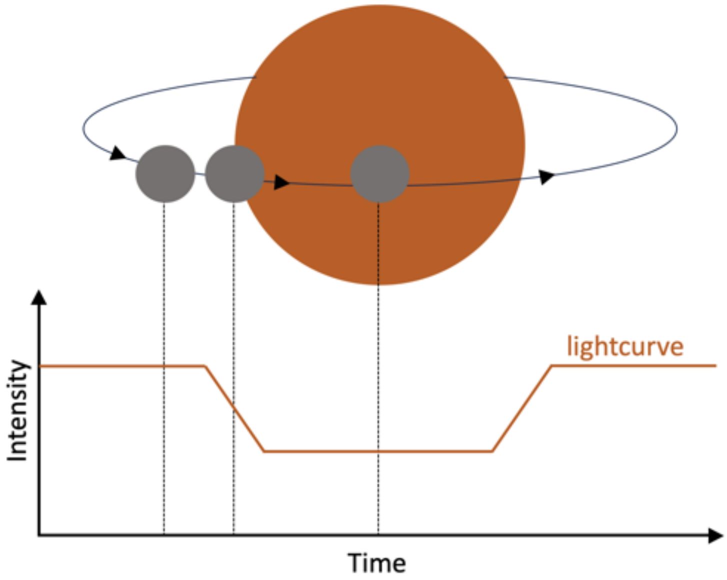

Figure 1.2: Pictorial representation of the transit photometry method, including the characteristic transit lightcurve.

momentary dimming of the total observed flux. The fraction of the light "lost" during the transit is dependent on the relative apparent sizes of the two objects. Such dimming has a characteristic shape in the stellar brightness time series shown in Fig. 1.2. Physically-motivated models can then be used to fit the observational data in order to derive the radius of the exoplanet and some of its orbital parameters. The first planet for which a transit was observed and studied was HD 209456 b (Henry et al. 2000), closely followed by the first ever planet detection with transit photometry of OGLE-TR-56 b (Udalski et al. 2002). The method has since risen to prominence as the most effective at radius detections, with the launches of the Convection, Rotation and planetary Transits (CoRoT: Baglin et al. 2006) space telescope, of the Kepler/K2 mission (Borucki et al. 2010; Howell et al. 2014), and of the Transiting Exoplanet Survey Satellite (TESS: Ricker et al. 2015) in 2006, 2009 and 2018 respectively. Transit photometry has incredible advantages that make this technique the most prolific at finding planets. It allows for the observation of multiple targets at the same time, and enables the study of stars located several thousand light years away. It is the most direct method to estimate the radius of a planet. Via secondary eclipse it also allows for the calculation of the planetary radiation, and therefore the temperature of the planet. Nevertheless, this technique has several drawbacks: first and foremost, assuming a random distribution of planetary system orientations, only a

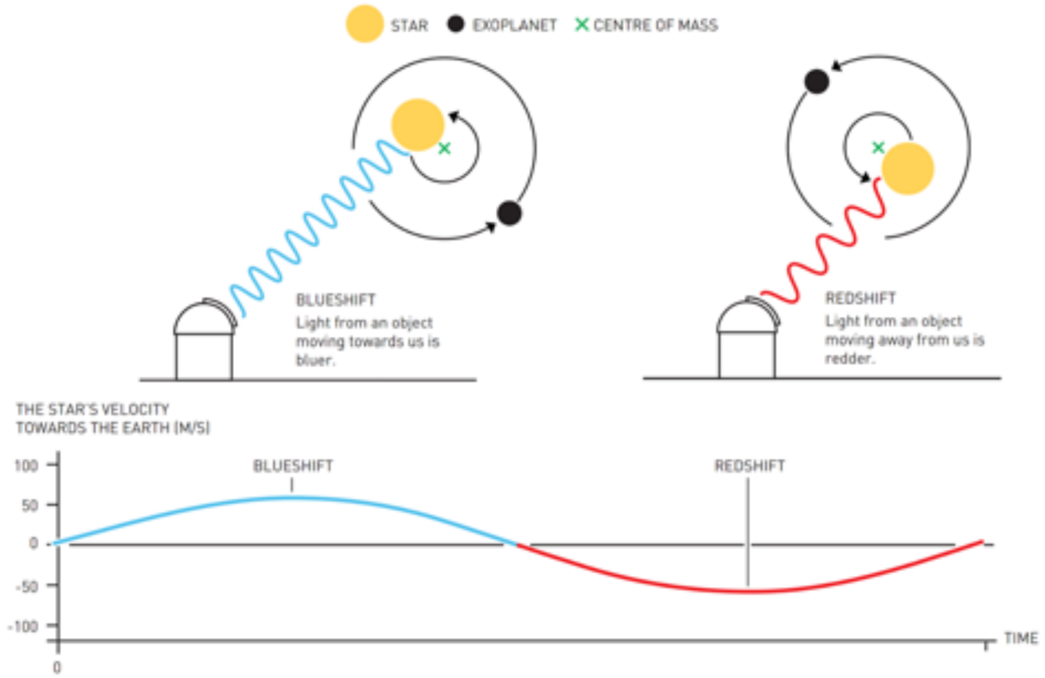

Figure 1.3: Pictorial representation of the radial-velocity method. Figure adapted from an image produced by Johan Jarnestad/The Royal Swedish Academy of Sciences.

small minority can actually be expected to transit the star on our line-of-sight. The probability of observing a transit of a planet with orbital period under 10 days is between 2 and 10%, and this percentage gets down to 0.46% for a planet in an Earth-like orbit (Collier Cameron 2016). Transit photometry is also the method that yields the most false positives, with over 35% of initial candidates detected by the Kepler mission being disproven (Santerne et al. 2012). Moreover, this technique is heavily biased towards exoplanets orbiting close to their star: they are more likely to transit, their apparent size relative to their star is larger, and they require shorter observing baselines to detect multiple transit events. Even with these downsides, transit photometry is still a powerful tool for the detection of exoplanets, especially if paired with the radial-velocity method for mass characterisation.

1.1.1.2 Radial Velocity

Light emitted by astronomical objects in movement is subject to the Doppler effect. The measured frequency of the emitted photons changes based on the relative motion between the source and the observer, decreasing (redshifting) for objects moving away from the observer and increasing (blueshifting) for objects moving towards the observer. By analysing the Doppler shift in the spectral lines of celestial objects, one can compute the body's radial velocity (RV), defined as the component of the object's velocity pointing in the direction of the observer. The presence of a planet applies a gravitational pull to the star. As the smaller body travels in its orbit, it will also cause its star to "wobble" around the common centre of mass, as shown in Fig. 1.3. The more massive and the closer this second object is, the more significant this effect will be. By monitoring the Doppler shift in the stellar spectra, we can trace the star's motion and analytically compute the mass and orbit of the body causing the wobble: the exoplanet. As Otto Struve predicted, the first exoplanet orbiting a main-sequence star, 51 Pegasi b (Mayor and Queloz 1995), was in fact detected using radial-velocity observations with the ELODIE spectrograph, installed at the Haute-Provence Observatory in France (Baranne et al. 1996). As for 2023, roughly 20% of all confirmed planets were first detected using this technique. The radial-velocity method is independent of distance, but still requires good signal-to-noise ratio (SNR) for precise detections. Differently from transit photometry, we can only observe one star at a time. Given that the amplitude of the signal is directly proportional to the mass of the planet and inversely proportional to its distance from the star, the radial-velocity method has been very successful in the detection of "Hot Jupiters", giant gaseous planets with orbital periods smaller than 10 days, but it can struggle to confirm smaller Earth-like planets, especially those with wider orbits. In the first 20 years since 51 Peg b, the precision of the instruments was the greatest limitation of spectroscopy as a planet-detection method. Taking for example the planets in the Solar System, Jupiter imprints a Doppler shift on the absorption lines in the spectra of the Sun of amplitude 12.5 $\mathrm{m\,s^{-1}}$, while the Earth's signal is only 8 $\mathrm{cm\,s^{-1}}$, almost 16 times smaller. This need for higher instrumental precision and long-term stability has motivated the commissioning of a series of new high-precision, ultra-stable spectrographs such as EXPRES and ESPRESSO, which have proven to reach precision of 30 $\mathrm{cm\,s^{-1}}$ (Brewer et al. 2020) and 60 $\mathrm{cm\,s^{-1}}$ (Suárez Mascareño et al. 2020) respectively. Tests on the soon-to-be on-sky HARPS3 (Thompson et al. 2016) show projected precision of 10 $\mathrm{cm\,s^{-1}}$. New-generation spectrographs are therefore coming within reach of the precision needed to detect Earth-analogues. However, stars are not static. They are complex, active bodies. They exhibit many time-dependent radial-velocity perturbations that can impact exoplanet detection by simulating the presence of a non-existent planet or by obscuring their signal. Stellar-induced variability, in fact, often dominates the RV budget. Thus with the new precision, the greatest challenge for the detection and characterisation of exoplanets in the Extreme Precision Radial-Velocity (EPRV) regime is now stellar activity (Saar and Donahue 1997; Lindegren and Dravins 2003; Meunier et al. 2010; Dumusque et al. 2011; Fischer et al. 2016; Crass et al. 2021; Meunier 2021).

1.1.2 The Objectives of this Thesis

The community has therefore now reached a stage at which effective and efficient mitigation and modelling of stellar-induced RV signals is required for the advancement of the field. In this work, I focus on describing methods of understanding and addressing the various sources of stellar activity affecting the radial-velocity domain for the detection of exoplanets. The remaining Sections of Chapter 1 summarise the processes that give rise to the most significant stellar signals, ordered by the amplitude of the imprinted RV variation. Chapter 2 provides a detailed overview of the state of the field, and the techniques currently most used to mitigate stellar variability. Chapter 3 focuses on the definition of Gaussian Processes (GP), and the description of the Python pipeline MAGPy-RV, a tool that employs GP regression coupled with Markov Chain Monte Carlo parameter space searching algorithms to model stellar variability. In Chapter 4 the described software is applied to the detection and characterisation of two planetary systems. The first system orbits the star TOI-2134 and is comprised of an inner mini-Neptune and an outer sub-Saturn in a highly eccentric orbit. I also detect three super-Earths orbiting the bright K-dwarf HD 48948. This Chapter highlights the fact that caution and systematic model testing should always be employed when mitigating stellar signals. Chapter 5 focuses on the analysis of solar RV observations with the aim of building a better understanding of stellar activity. It introduces the pipeline SolAster for the extraction of Sun-as-a-star data from resolved solar observations. These data are then used to investigate a new stellar rotation tracer to inform planet detection analyses: the mean longitudinal magnetic field.

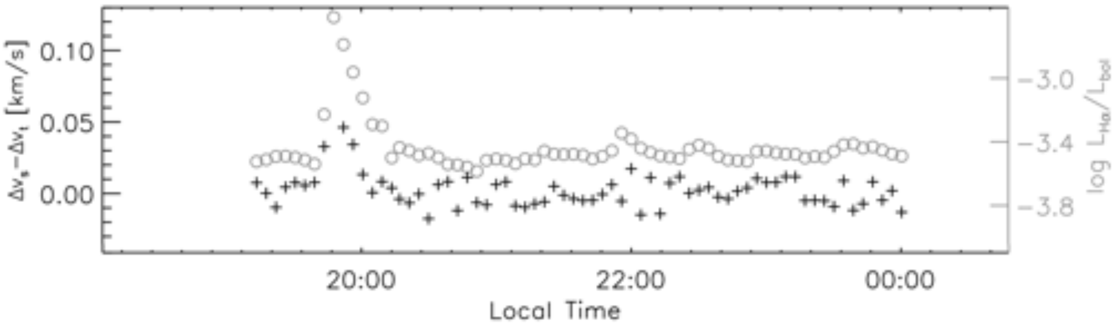

Figure 1.4: Radial-velocity scatter (black symbols) and normalised Hα-index (gray circles) of the mid-M flare star CN Leo. A flare event is recorded as the spike in both time series at roughly 20:00 local time. Figure adapted from Fig. 8 of Reiners (2009).

1.2 Sources of Stellar Variability

Stars are incredibly complex astronomical objects. They have been the focus of study since the conception of astronomy in one way or another. Stars are also major players in the hunt for exoplanets. Stellar age, composition and rotation axis can be vital information to understand the characteristics of a planet and its formation history. Most crucially stars are active, meaning that their surface and atmosphere are in continuous motion. In the context of exoplanet detection with the radial-velocity method, this wealth of processes shaping the photo- and chromosphere of the host star can generate signals that mimic or obscure the presence of a planet. This Section introduces the major sources of stellar variability in RV data. The list is organised based on the amplitude of the injected signal.

1.2.1 Flares and Mass Ejections

Stars have magnetic fields. They are created by the motion of charged plasma in the convective zone of the star. These magnetic fields generate electric currents which in turn, due to a combination of induction, differential rotation and Coriolis force, sustain the magnetic field. Stellar dynamos can be extremely complex, and they evolve in time. The stellar magnetic fields interact with the convective plasma in a variety of ways. One of the most extreme of these processes are flares. Coronal flares are intense, localised emissions of electromagnetic radiation. They have been extensively observed in the Sun and have been detected in stars (e.g., Fletcher et al. 2011; Davenport et al. 2019; Katsova et al. 2022). They are caused by the interaction between charged particles and the plasma medium. Studies have associated flare events to magnetic reconnection in solar arcades (Zhu et al. 2016).

Figure 1.5: Section of a continuum intensity observation captured by NASA's Solar Dynamics Observatory of two active regions on the Sun rotating out of sight. Multiple sunspots are visible, with the temperature gradient between umbra and penumbra clearly observed. Faculae surrounding the spots are only visible closer to the limb, given their significantly smaller temperature difference with the quiet surface and their tubular geometric configuration.

Flares are often accompanied by ejections of matter trapped in the unconnected magnetic fields that violently expand outwards in the heliosphere. These occurrences are called coronal mass ejections, or CMEs (Howard et al. 2023). These large scale events can impact the results of RV investigations. However, flares and CMEs are dramatic and can be easily identified as anomalies, since they generate spikes with respect to the average RV of several tens of $\mathrm{m\,s^{-1}}$ (Reiners 2009). They have lifetimes of minutes to hours. Their presence in the dataset is usually confirmed by jumps in Hα and S-index values, as shown in Fig. 1.4, and their effects are mitigated by discarding the data. Overall they do not represent a significant obstacle in typical exoplanet searches.

1.2.2 Active Regions

When the stellar magnetic field coagulates in small areas on the stellar surface, active regions arise. In the photosphere, strong local fields manifest themselves as regions with different luminosity than the surroundings. Magnetic fields can locally halt convective motions, thus decreasing the temperature of the upper layer of plasma. This cooling is expressed by a reduced brightness of the affected region as shown in Fig. 1.5, commonly referred to as a star spot. Spots usually appear in pairs of opposite polarities (Solanki 2003), and their temperatures range from 500 to 2000 K less than the surrounding regions (Schrijver 2002). Their lifetime depends on their size: smaller spots can dissipate in days, while larger ones can last months on the stellar surface (Giles et al. 2017). In the Sun, spots arise with varying sizes, with diameters between 1,500 and 20,000 km. Starspots are often surrounded by bright faculae. They are regions in which narrow magnetic flux tubes are embedded in the quiet stellar atmosphere (Walton 1987). They are usually grouped into networks (Berger and Title 1996) and are most visible near the stellar limb, given their tubular shape. Faculae are only 10% brighter than the surrounding quiet star as shown in Fig. 1.5, and a few hundred Kelvin hotter (Thomas and Weiss 2008). When isolated, they have short lifetimes, of a couple of hours, but they can persist up to years when grouped (Foukal 1998). Faculae always anticipate, accompany and outlive starspots. In the Sun faculae tend to dominate over spots, meaning that their filling factor (the percentage of area they cover on the visible stellar surface) is consistently larger than the one of sunspots (Haywood et al. 2022). In the chromosphere, strong magnetic fields can also cause areas of increased brightness, referred to as plages. Their position often closely maps the faculae in the photosphere below (Hall 2008). They are made of small bright points called floccule surrounded by thin, dark up-moving jets known as spicules (Zirin and Howard 1966). The number and size of active regions on the stellar surface changes with time and is dependent on the star's magnetic cycle. They generally arise at higher latitudes, migrating towards the equator as the cycle progresses to maximum (Strassmeier 2009). While active regions are relatively spatially stable during their lifetimes, their signal is significantly modulated by the stellar rotation, as they contribute to the stellar activity RV signal only when they are in view of the observer. For a full review on the birth and evolution of active regions and their fundamental characteristics, the reader is directed to Van Driel-Gesztelyi and Green (2015), Thomas and Weiss (2008) and similar reviews. In this work I primarily focus on how the presence of active regions affects radial-velocity measurements.

1.2.2.1 Suppression of Convective Blueshift

As is fully addressed in Section 1.2.5, plasma motions cause an overall blueshift of the stellar spectra. The strong magnetic fields that generate active regions inhibit or completely halt convection on the stellar surface, and therefore reduce the total blueshift observed. This effect is commonly referred to as the suppression of convective blueshift. The amplitude of the "lost signal" is dominated by the amount of surface covered by active regions. In Sun-like stars, faculae filling factors tend to be larger than those of spots. Faculae are therefore the driving force of this suppression of blueshift in Sun-twins (Meunier et al. 2010; Haywood et al. 2016; Milbourne et al. 2019). Younger, fast rotating stars are more likely to present much stronger stellar activity and to be instead dominated by starspots, with filling factors up to 22% of their surface (Strassmeier 2009). The suppression of convective blueshift due to the presence of active regions can be tricky to model. In most cases this signal is modulated by the stellar rotation period, as active regions rotate in and out of view of the observer. Active regions however have limited lifetimes, and will dissipate in the surrounding quiet star, as new ones emerge in different locations. The generated signal has therefore a quasi-periodic behaviour. In the Sun, RV variations induced by the suppression of convective blueshift have amplitudes ranging between 0.5 and 10 $\mathrm{m\,s^{-1}}$. In stars, the strength of this signal is dependent on the overall activity level of the star, but it can reach amplitudes of $\mathrm{km\,s^{-1}}$ in young stars. In most cases, the effects of the suppression of convective blueshift are mitigated with the analysis of activity proxies (as addressed in Section 2.2), or with machine learning modelling techniques such as Gaussian process regression (further covered in Chapter 2 Section 2.3.4, and in Chapter 3).

1.2.2.2 Photometric Inhomogeneities

The brightness of active regions is different from the quiet-star average. Their increased or significantly decreased luminosity therefore impacts the rotational profile of the star. In a rotating object with rotational axis perpendicular to the observer's line-of-sight, half of the disc is moving towards the observer, and the other half is moving away. The flux emitted by the first section is blueshifted, while the light emitted by the second is redshifted. In a perfectly homogeneous case, these two sides will perfectly balance, leaving no net shift in the RVs. However, the presence of bright faculae or a dark spot breaks this symmetry by respectively enhancing or diminishing the total signal of the disc half they are on. For example, in the case of a dark spot emerging in the field of view, the summed brightness of the blue limb is smaller than the one of the red limb, as shown in Fig.1.6, and an overall

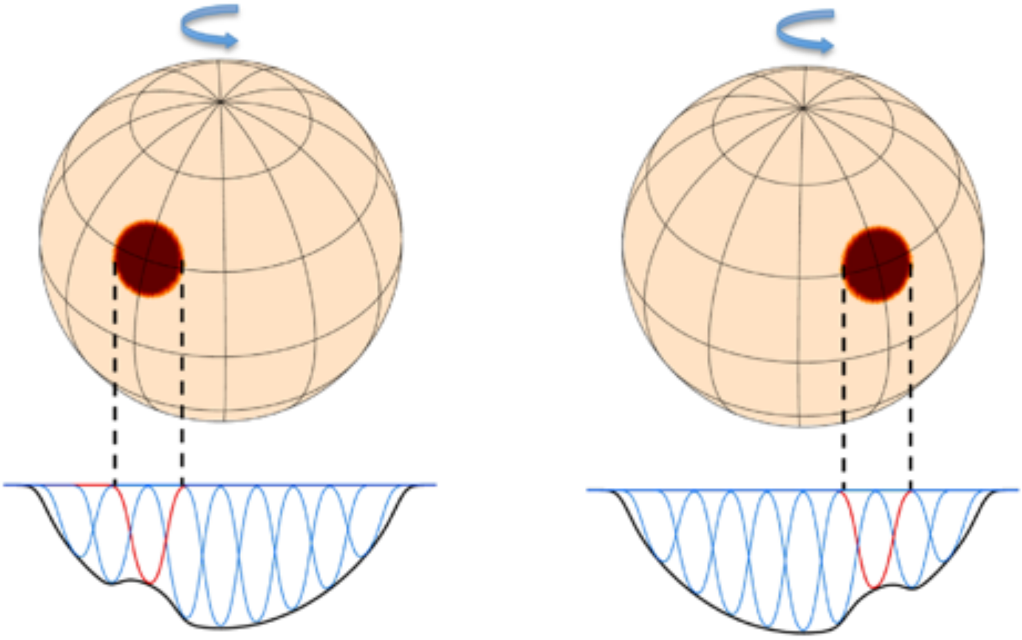

Figure 1.6: Schematic illustrating the distortion caused by the presence of dark spots in the line-profile of stars. The introduced asymmetries lead to a shift of the centre of the best-fit Gaussian function to the profile and therefore to variations in the extracted radial velocities (more information on this last step in Section 2.1). As the spot rotates on the visible disc, the velocity experiencing flux deficiency also travels across the profile. Figure adapted from Kochukhov (2016).

redshift will be observed. The opposite happens as the spot rotates out of view. The RV signal generated by these photometric inhomogeneities is mostly driven by starspots, even in Sun-like stars. While faculae may dominate the active region filling factors, the difference in luminosity with the surroundings is much less significant than the one between starspots and the quiet star. In the Sun, the amplitude of these variations is of the order of tens of $\mathrm{cm\,s^{-1}}$ at most. Similarly to the signal due to suppression of convective blueshift, photometric inhomogeneities are modulated by the rotation period of the star and by the lifetime of the active regions, thus exhibiting quasi-periodic behaviours. Their effects in the RVs are most often modelled with simultaneous photometric techniques (see Section 2.3.4.2), and with Gaussian process regression.

1.2.3 Magnetic Cycles

The Sun has been observed to experience a nearly-periodic variation in its magnetic activity expressed in the form of active regions and explosive events. Over roughly 11 years (Hathaway 2010), solar activity increases to a maximum, then slowly decreases to a minimum, normally defined by the maximum and minimum values of sunspot count. The solar brightness also varies over this period, and the magnetic field flips polarities (Charbonneau 2020). The study of the solar magnetic cycle and its prediction are some of the most active fields of solar science (e.g., Shepherd et al. 2014; Ng 2016). Stars with internal compositions similar to the Sun, and therefore with comparable stellar dynamo processes, are also expected to experience magnetic cycles (Lanza 2010; Jeffers et al. 2023). These cycles however have been consistently hard to measure due to the long observational baselines needed. Starting in 1966, various programs monitored a variety of stars, and in 1978 the first evidence of stellar magnetic cycles was published (Wilson 1978). Since then there has been a continuous effort towards building a connection between solar observations and stellar models. In the context of RV surveys, magnetic cycles affect the detectability of exoplanets in a combination of ways. Most relevantly, they produce significant long-term radial-velocity variations, of the order of tens of $\mathrm{m\,s^{-1}}$ (Lovis et al. 2011), with periodicities that may mimic the presence of outer gas giants. As an example, Jupiter's 12-year orbit imprints roughly a 12 $\mathrm{m\,s^{-1}}$ RV signal on the Sun, a variation comparable to the 7 $\mathrm{m\,s^{-1}}$ 11-year trend induced by the solar magnetic cycle. It is therefore important to distinguish between the two. One of the ways to isolate the long-term effects of magnetic cycles is to turn to activity proxies to isolate common trends. More complex modelling methods are currently being developed, such as employing neural networks or non-stationary kernels in Gaussian processes. However, the lack of training data represent a significant limitation.

1.2.4 Oscillations

Stars are not static objects. They oscillate in a variety of ways. These oscillations manifest themselves as motions in the stellar photosphere as well as density or temperature changes in the full stellar body (García and Ballot 2019). Stars can be excited to oscillation by different effects, but most main-sequence stars, sub- and red-giants undergo the same process: turbulent motions in the external convective layers generate repeated stochastic excitation and damping. These oscillations are known as "solar-like". Solar-like oscillations are resonances occurring at specific frequencies dependent on the characteristics of their considered star. They therefore can be considered resonance modes. By studying the

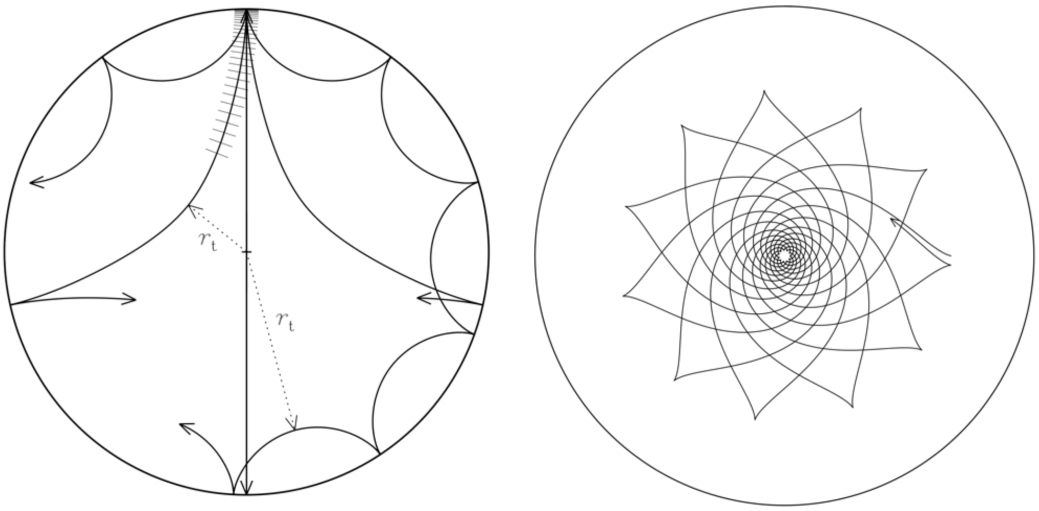

Figure 1.7: Schematics illustrating the propagation of oscillation modes in stars. *Left panel:* pressure modes propagating through the stellar interior. *Right panel:* gravity modes propagating in the non-convective stellar areas. Both diagrams are adapted from Di Mauro (2016).

Sun, an array of different modes caused by different effects has been identified. While they all impact stellar observations in the extreme precision radial-velocity regime, most produce signals too small to be resolved with current spectrograph precision. In this work, I only address three modes. For a more detailed description of asteroseismology as a whole, see the following reviews: Chaplin and Miglio (2013), Hekker and Christensen-Dalsgaard (2017), and García and Ballot (2019).

1.2.4.1 Pressure Modes

The most significant mode of oscillation is the pressure mode, or p-mode. They are acoustic standing waves, and arise from the pressure gradient of the star. Their resonant frequencies are mainly dependent on the stellar outer layer, the section between the top of the photosphere and a characteristic internal turning point. The left panel of Fig. 1.7 shows a diagram of their typical behaviour. They are one of the best ways to study the interior of stars. P-modes result in RV variations of the order of a few $\mathrm{m\,s^{-1}}$ (Schrijver and Zwaan 2000) with characteristic timescales between 5 and 15 minutes. Given their stochastic nature, the best way to mitigate the contribution of pressure modes to the radial velocities relies on simple "averaging-out", by taking observations with at least 10-minute long exposures (Dumusque et al. 2011; Chaplin et al. 2019).

1.2.4.2 Gravity Modes

Gravity modes, or g-modes, are driven by buoyancy forces due to gravity and density fluctuations. In Sun-like stars, they are confined to the radiative interiors or the atmosphere, as shown in the right panel of Fig. 1.7, as they can only propagate in non-convective areas. Due to this characteristic, they are difficult to detect and isolate. Their injected RV amplitude is very small, and the variations are expected to repeat with periods comparable to hours. In current RV surveys, g-modes are not a focus of mitigation techniques, given their minimal impact.

1.2.4.3 Quasi-toroidal Modes

Quasi-toroidal modes, or r-modes (given their connection to Rossby waves), are generated by the rotation of the star (Papaloizou and Pringle 1978). They are driven by the Coriolis force. The amplitude of the signal that they imprint on RV observations is directly proportional to the stellar mass and inversely proportional to the square of the stellar radius and rotation frequency (Provost et al. 1981). Variability generated by r-modes has regular but complex behaviours, with periodicity dependent on the mode considered. In the Sun, r-modes periods ranging from 20 to 38 days have been detected (Wolff and Blizard 1986). Lanza et al. (2018) found that the periods of RV oscillations due to r-modes are related to the stellar rotation rate as $3/4$ and $2/5$ P_{rot}. These variations are of the order of tens of $\mathrm{cm\,s^{-1}}$, and are therefore now of interest to more precise spectrographs such as HARPS3. No mitigation technique has been developed yet to account for these signals.

1.2.5 Magnetoconvection

Matter in the outer layers of stars is in continuous motion. While radiation is the most efficient mode of energy transport in the inner parts of the star, convection dominates closer to the surface due to the increased opacity of plasma. In Sun-like stars, the unstable hydrogen ionisation zone under the visible photosphere is in constant cyclical motion. Plasma at the bottom of this area is heated up by radiative energy, and tends to expand in order to maintain pressure-temperature equilibrium. This decrease in density forces the hot plasma to rise towards the surface of the convective zone, driven by buoyant forces. As it reaches the upper layers, it cools down and thus sinks back to the bottom, re-starting

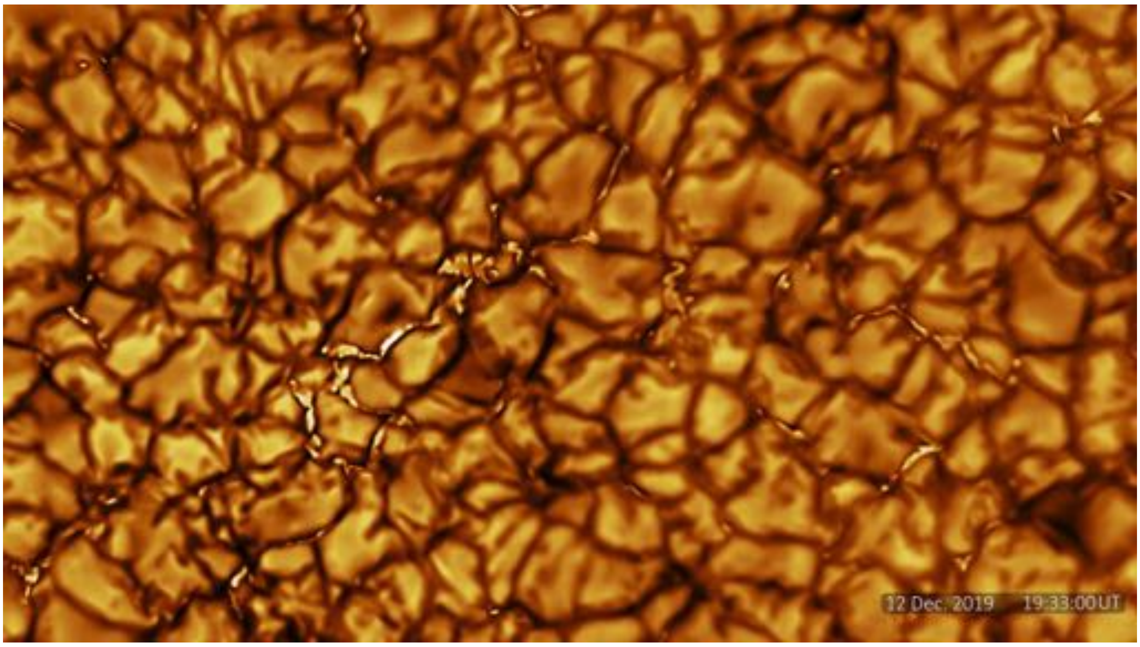

Figure 1.8: Image of the solar surface observed with the Daniel K. Inouye Solar Telescope taken at a wavelength of 705 nm. The granulation structure is clearly visible. Image credit: NSO/NSF/AURA.

the cycle. Convective motions happen at multiple spatial and temporal scales. For a more in-depth analysis of magnetoconvection in the Sun see Stein (2012) and Proctor and Weiss (1982).

1.2.5.1 Granulation

The most obvious expression of magnetoconvection is granulation. The continuous rising and sinking of plasma generates the typical "patchy" look clearly visible on the surface of the Sun, as shown in Fig. 1.8. Bright areas with up-welling flow of hot material are referred to as granules, while the dark filament-like regions surrounding them are called intergranular lanes and are the zones in which the cooled plasma sinks back down. The size and lifetime of granules change based on the magnetic field that creates them, ranging from kilometres to megametres in diameter (Del Moro 2004), with lifetimes of minutes to days, respectively (Hall 2008). Granules are distinctly quasi-polygonal in shape (Leighton 1963). Since the intergranular lanes are dimmer and occupy a smaller fraction of the surface, the Doppler shifts generated by the upward and downward motions of plasma do not cancel out. In a radial-velocity observation, this effect leads to an overall blueshift (motion towards the observer) due to each convective cell. This blueshift measures roughly 3.5 km s^{-1} in the Sun (Lindegren and Dravins 2003; Dravins 1999). In the absorption lines

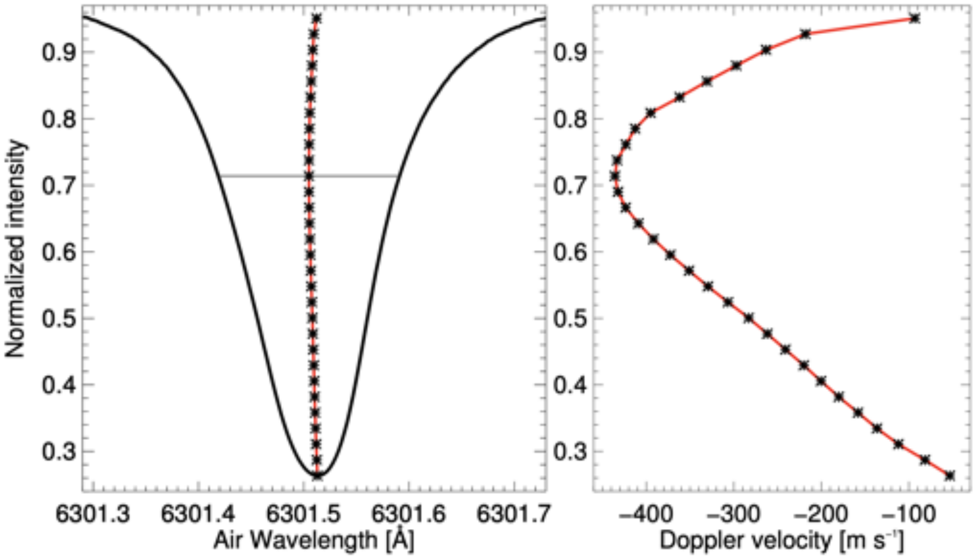

Figure 1.9: Spectral line profile and bisector of the Fe I 6301.5 Å line observed in the magnetically quiet Sun. *Left panel:* the intensity of the spectral line normalised to continuum is plotted as a solid black line, the computed bisector points are shown as black asterisks and their interpolation as a solid red line. *Right panel:* the same bisector line is here shown with a different x-axis (in Doppler velocity) to highlight the typical C-shape caused by granulation. Figure taken from Fig. 13 of Löhner-Böttcher et al. (2017).

and in the cross-correlation function (CCF) of the spectrum, this general blueshift is visible as a distortion of the Gaussian shape (e.g., Gray 2005), highlighted by the tilt towards positive velocities of the bisector. The typical C-shape of the bisector of the CCF is shown in Fig. 1.9 (Löhner-Böttcher et al. 2017). However, the granulation pattern does not remain constant in time. Granules have lifetimes correlated to their average size. The surface they each cover also changes with time, tending to decrease for granules smaller than 0.76 Mm^2 and increase for those above that typical size (Meunier et al. 2015). Changes in the total number of granules and in the bright-dark balance of granules and intergranular lanes result in RV variations of the order of $\mathrm{m\,s^{-1}}$. These velocity fluctuations obey Poisson statistics, and their amplitude can be approximated to the square root of the total number of granules. Simulations of the Sun show that granulation leads to a root mean squared (RMS) scatter in the RVs between 0.3 and 0.8 $\mathrm{m\,s^{-1}}$ (Meunier et al. 2015; Collier Cameron et al. 2019; Al Moulla et al. 2023; Dalal et al. 2023). Few RV surveys are directly assessing the effects of granulation. Those that do aim to average out the signal by taking multiple exposures in the same night. In the past years, more sophisticated techniques have been proposed, mostly focused on the disentangling of the granulation contribution directly in the stellar absorption lines (Cegla et al. 2019) or in the shape of the cross-correlation function (Palumbo et al. 2022).

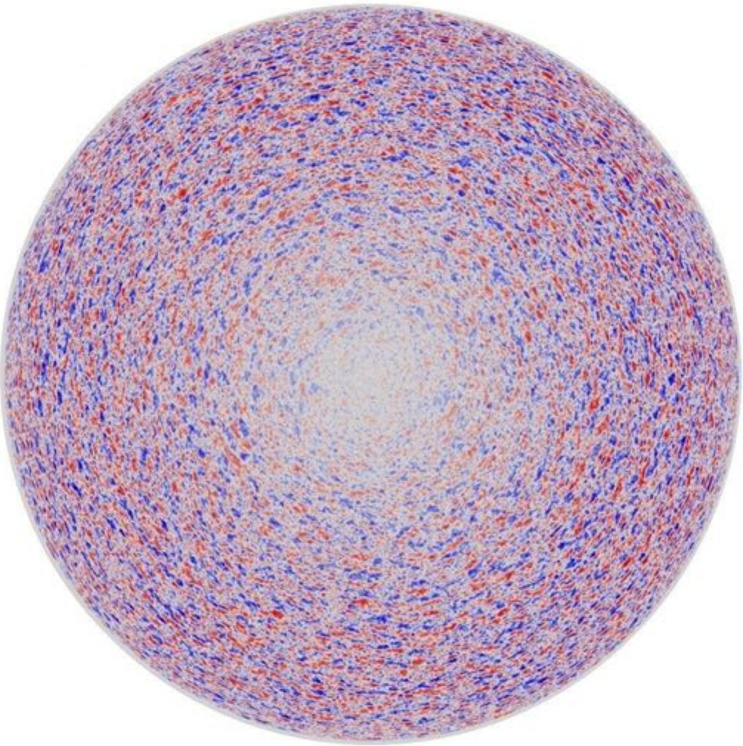

Figure 1.10: SOHO Dopplergram observation of the solar chromosphere. The figure is colour-coded based on the line-of-sight velocity observed. The supergranular structure is clearly evident. The mostly-horizontal nature of the motion of plasma in supergranules is here highlighted by the radially-symmetric visibility gradient from the centre of the disc towards the limb. Image credit: NASA/MSFC Hathaway.

1.2.5.2 Supergranulation

A second pattern of convection cells also introduces significant radial-velocity variations: the supergranulation. It was first proposed by Hart (1954). They detected a fluctuating velocity field with different irregularities and larger scales than what would have been expected from "traditional" convective instabilities. These results were later confirmed by Hart (1956). Since then, many separate measurements of the size and effects of this process have been published (e.g., Leighton et al. 1962; Simon and Leighton 1964) and supergranulation has been accepted as an expression of convection at larger spatial scales, as shown in Fig. 1.10. Although the most likely origin of this convective motion is considered to be thermal convection, building a true understanding of the causes and driving forces of supergranulation has been extremely challenging (Rincon and Rieutord 2018). Solar supergranules have been measured to have diameters of roughly 35 Mm (Cegla 2019) and lifetimes of 1.5 to 2 days. The majority of the plasma motion is horizontal, flowing radially out from the middle of the supergranule with velocities of the order of a few hundred m s^{-1}. Supergranulation has been modelled to contribute to up to 0.8 m s^{-1} to the stellar activity RMS budget in RV observations (Meunier et al. 2019), and these values have been reproduced by indirect observations (Lakeland et al. 2024). Given their long lifetimes, granulation-like averaging-out mitigation techniques are generally unfeasible. No other modelling method has yet been developed to directly address the effects of supergranulation.

1.2.6 Meridional Circulation

Matter on the solar surface is in continuous motion from the equator to the poles. In order to preserve equilibrium, a second return flow is also present underneath the surface in order to substitute the displaced matter. These flows arise from the redistribution of angular momentum due to turbulence (Meunier and Lagrange 2020). In the Sun, poleward flows have a maximum velocity of 10-20 m s^{-1} that change with varying solar cycle stage (Meunier 2021). The integrated RV contribution of solar meridional flows has been measured to be between 0.5 and 1.7 m s^{-1}, depending on activity level and on the orientation of the solar rotation axis (Meunier and Lagrange 2020; Markov et al. 2010). These flows were found to be maximally significant on the descending phase of the solar cycle, and when viewing the Sun pole-on. In other stars, the amplitude of this effect ranges between 0.1 m s^{-1} for quiet stars to 4 m s^{-1} for more active ones. Similarly to the Sun, meridional circulation in stars has also been observed to change in time (Chou and Dai 2001; Komm et al. 2015; Hathaway and Rightmire 2010), growing weaker with increasing sunspot number (Choudhuri 2021). Its imprinted signal on the RVs therefore also changes with time, introducing long-term variations with periods of the order of the stellar magnetic cycle. Meridional flows also suppress convective blueshift, and weaken the connection between activity proxies and active regions filling factor (more information on this correlation in Chapter 2, Section 2.2). Currently this effect is not taken into consideration in most RV planet searches, and no mitigation technique has been widely accepted.

1.2.7 Gravitational Redshift

As photons escape the gravitational well of their star, they lose energy to the stellar gravitational potential. This causes an overall redshift of the absorption lines observed in the spectrum. This redshift has been measured to be of the order of 600 m s^{-1} in the Sun (Lindegren and Dravins 2003). The amplitude of this signal is dependent on the size of the emitting object. Over timescales of days and longer, stars expand and contract, varying their radius. In the Sun, a change in radius of 0.01% introduces a change in velocity of ~6 cm s^{-1} (Cegla et al. 2012). For low mass stars, Cegla et al. (2012) found that effects due to gravitational redshift become significant over timescales of 10 days or longer. Thus, while radius fluctuations due to pressure modes are not a significant concern, any size deformation due to magnetoconvection or Wilson depression of starspots (Suzuki 1967) can successfully mimic or obscure the signal of an Earth-analogue. Variations in the overall gravitational redshift are currently not addressed or mitigated in any RV survey.

1.2.8 Other Effects

In this Chapter, I have covered the major sources of stellar variability, but the list is not complete. There are a wealth of other processes that directly or indirectly affect radial-velocity data. In most cases, these effects have amplitudes under the precision levels of most spectrographs and are therefore not accounted for in RV correction, extraction or modelling. This is the case for small scale behaviours in and around spots, such as Evershed flows (Evershed 1909; Rempel et al. 2009) - in which radial motion outwards from the centre of the starspot can reach velocities of several km s^{-1}- and differences in brightness between umbra and penumbra that affect the way our models of photometric inhomogeneities impact the RV data. Other processes, while they can produce significant signals, evolve over timescales of the order of decades. For example, solar Gleissberg cycles (e.g., Hathaway 2010) modulate the amplitude of the solar magnetic cycle over a period of 70-100 years. These changes are therefore too slow to be resolved in the great majority of RV surveys, which have significantly shorter baselines. For one reason or another, while interesting these effects are not of primary focus.

Chapter 2

Mitigating and Modelling Stellar Variability

> *"Not just beautiful, though — the stars are like the trees in the forest, alive and breathing. And they're watching me. What I've done up till now, what I'm going to do — they know it all. Nothing gets past their watchful eyes. As I sit there under the shining night sky, again a violent fear takes hold of me. My heart's pounding a mile a minute, and I can barely breathe. All these millions of stars looking down on me, and I've never given them more than a passing thought before. Not just the stars — how many other things haven't I noticed in the world, things I know nothing about?"*
>
> *— Haruki Murakami, Kafka on the Shore*

In this Chapter, I list and describe the relevant mitigation and modelling techniques that have been developed and are currently being employed to address the many sources of non-planetary signals in RV data. In this work, I focus on highlighting techniques aimed at the mitigation of stellar variability, instead of methods developed for the correction of instrumental systematics. The Chapter is organised as follows: Section 2.1 covers the various techniques for the extraction of radial velocities. It is subdivided in three parts. I fist describe the basic methodologies to derive RVs from spectral observations. I then introduce techniques that include stellar activity mitigation before or simultaneously to the RV extraction using information from either the spectra or the cross-correlation function. Section 2.2 describes the common activity proxies and how they can be used to better understand stellar variability in the radial velocities. Finally, Section 2.3 covers a subsection of the most employed methods of modelling the RV time series. The Chapter concludes with a brief summary of other independent observational data that can be studied for a better understanding of stellar activity or used for simultaneous modelling with the radial velocities.

2.1 Radial Velocity Extraction

The radial velocity of an object can be measured as the Doppler shift of the centre of the absorption lines of its spectrum. With infinitely high precision, the motion in time of each spectral line can be measured independently, each yielding the same value. However, spectrographs are in fact not infinitely precise, and the stellar lines are affected by stellar activity in a variety of different ways such that the derivation of RVs is not straightforward. In this Section, I describe common methods to combine information over the spectral range and compute radial velocities.

2.1.1 Basic Methods for RV Extraction

While there exist a large variety of techniques to extract the radial velocities starting from spectra, only the most used ones are included in this work. All of the described extraction methods rely on the fact that the stellar motion caused by the presence of a planet shifts all the absorption lines in the same way and by the same amount without affecting their shape. In this Section, I focus on methods that perform little to no modelling or correction for stellar variability, and instead only aim to minimise its significance and to reduce the uncertainties on the RV measurements. Generally radial-velocity extraction is performed with three main numerical techniques: the cross-correlation function (Section 2.1.1.1), least-squared deconvolution (Section 2.1.1.2), or least-squares template matching (Section 2.1.1.4).

2.1.1.1 Cross-Correlation Function

The most commonly accepted and widely used method to extract the RV variations from the stellar spectrum is the cross-correlation function, or CCF, method (Baranne et al. 1996;

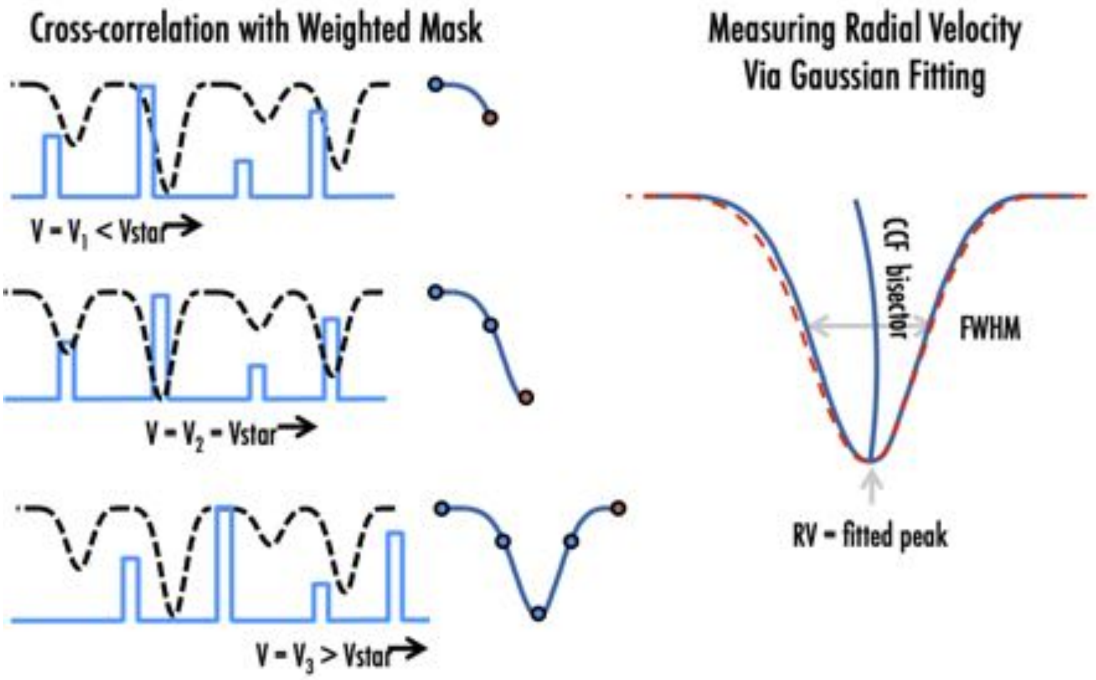

Figure 2.1: *Left panel:* Pictorial representation of the computation of the cross-correlation function (CCF). A weighted line mask, here in blue, is shifted over the spectral observations, shown as black dashed lines, and their correlation is calculated for each velocity, as depicted for each step on the right. *Right panel:* example diagram of a CCF in blue with its main characteristics labelled, the full-width at half-maximum (FWHM) and the bisector (for an extended explanation, see Section 2.2.2). The radial velocity of the observation can be computed by fitting a Gaussian function, shown as a dashed red line, to the profile and finding its peak. Figure adapted from Roy et al. (2016) and Arpita Roy's 2018 Sagan Summer Workshop Talk.

Pepe et al. 2002). It requires the corrected order-by-order spectrum as well as a generic stellar line mask. The process is simple: an appropriate numerical stellar mask of the expected absorption lines given the stellar type is superimposed to the observed spectrum and is then shifted over all the available velocity space. At each velocity-step the cross-correlation between the mask and the observation is measured. With this method, the information of every line in the spectrum is condensed to a single line profile with high signal-to-noise ratio (SNR). The diagrams in Fig. 2.1 depict this basic process and its output. The RV of the observation can then be derived by finding the velocity at which the correlation peaks. This last step is sometimes done by fitting a Gaussian function to the profile. The CCF technique is a fast extraction method, taking minutes at most for each observation. It generates quick and reliable RVs and a high-SNR averaged line profile that can be further studied in the search of stellar activity proxies (see Section 2.2.2). However, the best feature of the CCF method is also its main limitation, as by averaging over all the spectral lines it loses the majority of the spectral information. Moreover, it relies on generic stellar masks that are not optimised to the specific target. While new adaptations of the CCF method are being developed to better account for stellar variability (e.g., Simola et al. 2019), the Extreme Precision Radial-Velocity (EPRV) community is moving away from this technique.

2.1.1.2 Multi-Mask Least-Squares Deconvolution

The Multi-Mask Least-Squares Deconvolution technique (MM-LSD: Lienhard et al. 2022)[1] is a recent advancement on the Least-Squares Deconvolution (LSD) method widely applied to the extraction of weak magnetic signals in spectropolarimetric data (Donati et al. 1997) and later to the extraction of RVs (Barnes et al. 2012). It is based on the assumption that the stellar spectrum can be modelled via the convolution of a mask of the absorption lines (in the form of multiple delta functions centred at the middle wavelength of each line and weighted by their depth) and a common weak-line profile. The LSD technique therefore aims to find this common weak-line profile by deconvolution between the observed spectrum and a stellar line mask. The basic method is fully detailed in Donati et al. (1997). The technique relies on the assumption that the line profile generated by multiple over-

---

[1] Code available at: https://github.com/florian-lienhard/MM-LSD

lapping lines can be described by the sum of the line profile of each separate line. This assumption holds in most cases, as long as the lines are not intrinsically overlapping. The major improvement of MM-LSD over LSD focuses on the introduction of a multi-mask approach: the method described in Lienhard et al. (2022) relies on several combinations of parameters to extract the deconvolved LSD profile. The RV of the observation is extracted by, for example, fitting a Gaussian curve to the computed common line profile to find its peak. This technique has been shown to yield RVs with generally lower scatter than those produced with the CCF method, and the line mask chosen for deconvolution can be easily adapted to the stellar spectral type. However, similarly to the CCF method, the MM-LSD method also averages out all the information of hundreds of lines into a single profile.

2.1.1.3 Line-by-Line

The Line-By-Line method (LBL: Dumusque 2018) is an intermediate solution between averaging out all the spectral information into a common profile and locally measuring the shift of each absorption line. Since its inception, the method has been further tested on solar-type stars (Cretignier et al. 2020a) and has been expanded to the near-infrared wavelength range (Artigau et al. 2022)[2]. Starting from a line mask of the appropriate stellar type, this technique measures the RV shift of each spectral line separately, and then performs a weighted mean of all the extracted values to obtain the final precise RV. It requires the reduced 2D stellar spectrum pre-corrected of any known instrumental systematics. A crucial part of the LBL technique is the choice of the lines used in the computation. The ideal selection includes the maximum number of available lines while excluding those that are sensitive to magnetic activity (such as the Ca II H and K, Hα and the Mg II lines) as well as all lines affected by tellurics (absorption features created by the passage of the stellar light through Earth's atmosphere). The LBL method is particularly successful at minimising the effect of outliers that can bias the RV measurements. An effective selection of lines can also minimise the effects of stellar activity on the extracted RVs, and reach sub-m s^{-1} scatter. However, the larger the number of the rejected lines is, the more significant the photon noise becomes, so excluding all lines affected by activity is not viable, and stellar variability cannot be fully corrected for. Moreover, this technique requires the chosen

---

[2] Code available at: https://lbl.exoplanets.ca

spectral lines to have high enough resolution that a velocity shift can be measured. This is only possible for a subsection of lines and is successful systematically only for bright stars. Cretignier et al. (2020a) extended the LBL method by optimising the line selection, and by computing line centres directly from the master stellar spectrum with an extremum localisation algorithm. The LBL technique has been shown to be very efficient at recognising lines affected by stellar activity, but it is not immune to other effects such as the propagation of uncertainties in the wavelength solution.

2.1.1.4 SERVAL

In order to remove or minimise the uncertainties derived from an imperfect fit between the selected line mask and the observations, some methods choose to create the template directly from the data, rather than using a pre-determined mask generated based on stellar type. The SpEctrum Radial Velocity AnaLyser (SERVAL: Zechmeister et al. 2018)[3] is one of such methods. Instead of using standard pre-computed numeric masks, SERVAL generates its template spectrum uniquely for each star. It follows a sequential and iterative approach: it first calculates an approximate RV shift for all observed spectra. All observations are then coadded following this preliminary shift into a master spectrum. The final RVs are computed as the shift between this new template and each observation using a χ^2 minimisation approach. SERVAL is a fully data-driven approach that aims to use all the information in the spectral lines. It has been shown to perform well on M-dwarfs observed with the CARMENES spectrograph (Quirrenbach et al. 2016), and has been tested on High Accuracy Radial-velocity Planet Searcher (HARPS: Mayor et al. 2003) and NEID (Schwab et al. 2016) data (e.g., Trifonov et al. 2020; Cañas et al. 2022). Nevertheless, SERVAL remains a template-matching technique, the reliability of which is dependent on the quality of the produced template. Further work is required to systematically show how SERVAL works when applied to F-, G-, and K-type stars.

2.1.1.5 Template-free Gaussian Process Matching

Most recently a new template-free method has been proposed, which relies on the use of Gaussian Processes (GPs) to model the spectrum (for more information on Gaussian

---

[3] Code available at: www.github.com/mzechmeister/serval

processes, refer to Chapter 3). Rajpaul et al. (2020) presented a starting proof-of-concept methodology for RV extraction using this technique. They model each spectrum using a GP described by a Matérn 5/2 kernel. By then aligning the GP model of each observation to all other spectral models, they are able to infer the RV shifts between them. With a large enough number of observations, this method essentially provides the perfect template. The GP model is not limited by uncertainties such as photon noise and mismatch, and it is not tied to pre-existing line lists or masks. At the current stage, this GP matching method assumes that the only variation in the spectra of stars is a perfect Doppler shift caused by the presence of exoplanets or other orbiting bodies. As has been addressed in Section 1.2, this is not the case. Further work is therefore required to extend this formalism to include stellar variability with evolving kernel hypeparameters. This technique is also computationally limited, as modelling the entirety of the visible spectra would require prohibiting amounts of computational resources and time. Moreover, its efficacy is directly dependent on the amount of observations available, working best with large numbers of data and long baselines. The GP matching method has, nevertheless, been proven to significantly improve the RV scatter due to stellar variability even in this preliminary version.

2.1.2 RV Extraction with Simultaneous Spectral-level Stellar Activity Mitigation

In this Section, I describe the most common and promising techniques that include spectral-level activity mitigation before or simultaneously to RV extraction. These methods focus on correcting for stellar variability in the wavelength-domain.

2.1.2.1 Line-by-Line Statistics

The Line-by-Line method (Section 2.1.1.3) can either enhance or minimise the contribution of stellar variability by careful selection of lines. A basic application of the method aims to utilise all available unblended, symmetric lines to maximise the amount of information included. However, further work has focused on building systematic line selection techniques which are tailored to isolate pure Doppler shifts from effects caused by stellar activity. This push started from the incipit, at formalisation (Dumusque 2018; Cretignier et al. 2020a), but it has since gained more momentum, and a series of different line selection approaches have been investigated (e.g., Bellotti et al. 2022). Ould-Elhkim et al. (2023) introduced Wapiti, a new data-driven approach of the line-by-line method coupled with weighted Principal Component Analysis (wPCA: Delchambre 2015)[4] for the correction and extraction of RV data. PCA is a method to reduce the dimensionality of large datasets. It relies on the creation of principal components, generated as the linear combination or mixtures of the initial dataset. These combinations are created in such a way that the new uncorrelated variables (the principal components) summarise all the information from the original dataset. Essentially, PCA takes N-dimensional data and generates N principal components. In doing so, it re-organises the information budget so that the great majority of it is assigned to the first component, then the majority of the remaining information is assigned to the second component, and so forth. The late components contain very little of the initial information, and therefore can be discarded with minimum loss, going thus from N-dimensional data to generally less than 10 principal components. By assuming that the RV time series extracted with the LBL method is dominated by "spurious" stellar-induced signals, the Wapiti algorithm aims to isolate non-Keplerian signals with wPCA reconstruction. The RVs generated by the gravitational pull of the planet are then computed as the subtraction between the original LBL time series and the new wPCA-reconstructed stellar RVs. For this calculation Wapiti determines the appropriate number of principal components that best describe only the stellar signal by permutation tests and leave-one-out cross validations (Cretignier et al. 2022). The Wapiti pipeline has however only been introduced recently, and further testing on stellar and solar data is required to properly assess its efficacy.

2.1.2.2 Temperature-dependent RVs

A complementary RV extraction approach building on the idea of measuring the radial-velocity shift from individual spectral lines was developed by Al Moulla et al. (2022). They assume that spectral lines formed at different temperatures in the photosphere are affected by stellar activity differently. They therefore aim to isolate line temperatures that are least sensitive to stellar variability and use those for RV calculation. After subdividing

---

[4] Code available at: https://github.com/jakevdp/wpca

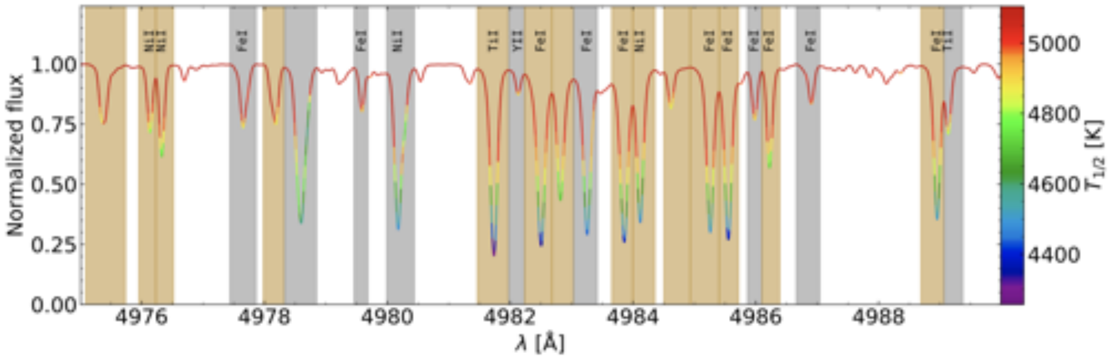

Figure 2.2: Pictorial description of the temperature-dependent RV extraction method. A 15 Å section of a corrected observed spectrum is plotted as a solid line colour-coded based on the temperature bin each line segment is assigned to. The number of bins is arbitrary. The formation temperature of each wavelength section is computed with spectral synthesis. The shaded areas indicate whether a spectral line has been selected (brown) or rejected (gray) for the computation of the RVs: only absorption lines segmented over more than three bins are chosen. Some lines are also described by a label to indicate their element and ionisation. Figure taken from Fig. 15 of Al Moulla et al. (2022).

the temperature range in N bins (in which N is arbitrarily chosen), they perform local thermodynamic equilibrium spectral synthesis in order to estimate which sections of each absorption line belong to each bin based on their average formation temperature, as shown in Fig. 2.2. Each spectral line is thus divided in up to $2N$ segments. To better understand the temperature gradient, only lines that are well-modelled by the spectral synthesis and that are divided over more than three temperature bins are selected. The mean RV of each temperature bin is obtained as the weighted average of the RVs derived from each selected line segment. This method is completely planet-insensitive, meaning it is not expected to absorb the signal of the planet for any choice of temperature bins. While this technique has been shown to be successful at isolating some effects of stellar activity such as rotation period and magnetic cycle, the ideal temperature selection to minimise the effects of stellar variability is still in discussion, as the dependence of stellar activity on line temperature formation is still not well understood. Moreover, the temperature sectioning is currently not based on physically-motivated assumptions.

2.1.2.3 WOBBLE

In the past few years, a series of data-driven algorithms have been developed to minimise the amount of assumptions and uncertainties introduced in RV extraction by imperfect templates. WOBBLE[5] (pronounced like Michael Bublé) is one of these techniques (Bedell

---

[5] Code available at: https://github.com/megbedell/wobble

et al. 2019). It is a simple linear model for the simultaneous derivation of stellar spectra, telluric spectra and RVs. It depends on information extracted from the data, rather than on pre-defined templates, or on physical knowledge of the star or of the Earth's atmosphere. Wobble models the flux of spectral observations as the sum of a stellar and a telluric contribution. The first is itself dependent on both magnetic activity and pure Doppler-shifts induced by planets. With the limiting assumption that the stellar variability and the telluric component do not change in time, Wobble fits the flux model to the observed spectra by maximising its logarithmic likelihood and allowing a varying Doppler-shift. The RVs of each epoch are then extracted from this process as the shift component that best fits each observation. This pipeline has been tested on a variety of HARPS stars, and it can produce RVs of comparable precision to the standard CCF method. Wobble has shown significant promise in the direct correction of telluric and micro-telluric lines, which makes it ideal for M-dwarf targets. Nevertheless, this technique assumes constant telluric and stellar activity spectral forms, a premise that is inherently incorrect. Wobble is currently being re-worked and extended to include different algorithms to model the spectrum. Further enhancements produced by the community are also welcome.

2.1.2.4 YARARA

Building on the success of wobble, Yarara (Cretignier et al. 2021) is a recent post-processing pipeline for the derivation of RV measurements from high-resolution spectra. The technique aims to correct the one-dimensional order-merged spectra of effects from cosmic rays, telluric lines, instrumental systematics (such as interference patterns, point spread function variations, fibre B contamination and ghosts), and stellar activity simultaneously with a data-driven approach, and to extract high-precision RVs. The workflow of yarara begins with spectrum normalisation with Rassine (Cretignier et al. 2020b). It then produces a master spectrum by stacking all individual observations, and uses it to derive residual spectra for all epochs. Yarara then cleans spectral ranges that show strong residuals with either physically-motivated models or by PCA and component fitting. As an example, stellar activity can be corrected by linearly fitting a scaled version of the S-index on the spectra time series matrix. The RVs are then extracted with the line-by-line method, using a line selection optimised for the chosen target as described in Cretignier et al. (2020a).

A second iteration of the pipeline was published in Cretignier et al. (2023). YARARA Version 2 includes a further correction to the LBL RVs using the shell methodology, introduced by Cretignier et al. (2022). The shell method performs variant template matching. Instead of measuring the RV shift, it extracts the line profile distortions orthogonal to the pure Doppler shift (and their time-domain coefficients) between each observation and a reference spectrum. The RVs generated by these distortions are then subtracted from the total radial velocities to isolate Keplerian signals. This pipeline has been extensively tested and is currently being implemented for the analysis of a series of targets (e.g., Stalport et al. 2023; Dalal et al. 2024). It can both provide the RVs corrected of stellar variability and instrumental systematics, and isolate the effects of stellar activity in a separate time series. This method assumes that a master spectrum free of all systematics and stellar variability can be built, although that is not always the case. YARARA is limited by the fact that it best performs on datasets with large numbers of observations and requires high signal-to-noise ratio in order to properly model the stellar lines.

2.1.2.5 AESTRA

Machine learning is a field of growing interest for astrophysics as a whole (e.g., Bloom et al. 2012; Domıénguez Sánchez et al. 2018). In the context of planet detection, neural network architectures have been successful at predictive modelling of large datasets for the assessment of false positives in photometric surveys (e.g., Shallue and Vanderburg 2018; Osborn et al. 2020). Deep learning is a subset of machine learning that uses multiple hidden layers in a neural network, and is generally interpreted in terms of probabilistic inference. Given the ability of neural networks at identifying repeating features in data, they have been employed to also correct stellar variability in large datasets. Liang et al. (2024), for example, presented Auto-Encoding STellar Radial-velocity and Activity (AESTRA), a deep learning method designed to distinguish between pure Doppler-shifts introduced in the spectra by orbiting objects and stellar-induced perturbations in the line profiles. In order to isolate all stellar variability patterns and derive high precision and accuracy planetary RVs, AESTRA uses a large sample of stellar observations covering a variety of activity levels and states. This pipeline combines deep learning methods for RV estimation with a spectrum auto-encoder (SPENDER: Melchior et al. 2023; Liang et al. 2023) for activity modelling. The two are jointly trained to determine the RV offsets. AESTRA does not make any assumptions about activity and is a fully flexible framework. It however has only been tested on simulated data, includes significant risks of overfitting and is computationally expensive. Moreover, like many of the previous techniques, it is not always clear what signals it is removing.

2.1.2.6 Other Methods

Section 2.1.2 only includes the most used methods for extracting radial-velocity measurements from spectral observations with simultaneous stellar activity correction, and it is not a complete list. Some honorable mentions are therefore due: Doppler-constrained principal component decomposition (Jones et al. 2017), joint modelling of telluric and stellar variability (Gilbertson and Ford 2022), and RV extraction with short-time Fourier Transform (STFT) are all also promising methods.

2.1.3 RV Extraction with Simultaneous CCF-level Stellar Activity Mitigation

Spectrum-based computations that include activity mitigation can be extremely computationally expensive, and can take long amounts of time. In some cases, simpler methods have been shown to perform just as well. In fact, CCF-based RV extraction still remains the most widely used technique. Rather than fitting a simple Gaussian curve to the line profile, stellar activity can be isolated directly in the CCF before RV computation. In this Section, I focus on RV extraction and stellar variability mitigation techniques that rely on information from lower-dimensional data such as the cross-correlation function.

2.1.3.1 FIESTA

In an attempt to develop more robust techniques to quantify the effect of stellar variability in radial-velocity measurements, interest turned to methods to separate activity and pure Doppler shifts starting from the CCF level. To do so, Zhao and Tinney (2020) introduced a new technique for the study of CCF variability in Fourier space: FourIEr phase SpecTrum Analysis (FIESTA or ϕESTA: Zhao and Tinney 2020; Zhao et al. 2022). This method relies on the translation property of Fourier transforms. In summary, the phase change of a shifted signal with the same power spectrum can be analytically derived in the Fourier domain.

FIESTA decomposes the CCF into its orthogonal Fourier basis functions and calculates the shift for each. It is therefore able to separate between pure shifts in the line profile, extracted as frequency modes all shifted by the same amount, and line distortions due to activity, in which the composing frequency modes are shifted with different phases. The original version of FIESTA identifies and separates regions of the line profile that are more and less sensitive to line deformation, and uses basic algebra to derive the stellar-signal contribution to the extracted RVs. An updated version, FIESTA II (Zhao et al. 2022) decomposes the line profile into all available Fourier modes up to the CCF sampling limit, which facilitates the separation of effects due to multiple sources of stellar variability. Via simulations, FIESTA has been shown to increase the detectability of planets, even in the case of semi-amplitudes comparable to the stellar variability, and it can be used to classify relative amplitudes of planetary and stellar signals in periodogram analysis. However, FIESTA works best with high-SNR observations and further modelling of the extracted time series with techniques such as GPs (see Section 2.3.4) is often required.

2.1.3.2 SCALPELS and TWEAKS

Following a similar idea, Collier Cameron et al. (2021) presented Self-Correlation Analysis of Line Profiles for Extracting Low-amplitude Shifts (SCALPELS), a CCF-domain method to isolate shifts generated by Doppler effect from apparent shifts due to line profile deformation. This technique relies on the translation invariant property of the autocorrelation function (ACF, for a description see Section 2.3.3) of the CCF. Collier Cameron et al. (2021) describe the ACF of the CCF as the expectation value of the vector cross-product of the CCF with itself at a sequence of lags. SCALPELS applies singular-value decomposition to both the CCFs of the data and their ACFs to define the principal modes of variability. It derives the orthonormal basis in the time-domain and the orthonormal modes of profile variability. The appropriate number and ranking of leading principal components is found using leave-one-out cross-validation. The radial-velocity variations due to shape distortions, often called the shape-RVs, can then be isolated by projecting the raw RVs on the time-domain subspace spanned by the amplitude coefficients of the derived ACF basis. The shift-RVs, containing only the signals from pure Doppler shifts, can then be computed as the subtraction between the raw RVs and the shape-RVs. The SCALPELS formalism has been successfully applied to both solar and stellar data (Wilson et al. 2022; Anna John et al. 2022), and it has since been extended to include exoplanet characterisation. Time and Wavelength-domain stEllar Activity mitigation using KIMA and SCALPELS (TWEAKS: Anna John et al. 2023) was specially designed to achieve sub-m s^{-1} detection threshold for longer-period exoplanets by combining stellar activity mitigation in both the wavelength- and the time-domain. KIMA[6] (Faria et al. 2018) is a nested-sampling (NS) package for the detection of exoplanets and the analysis of their orbital characteristics that includes the number of planets as a free parameter. It uses a diffusive NS algorithm (Brewer et al. 2009) to sample from the posterior distribution of the model parameters, and to ultimately extract the Bayesian evidence for each model in order to compare various system configurations. The TWEAKS workflow begins with the construction of the basis vectors from the ACFs using SCALPELS, which represent the shape-driven components of the CCF. These vectors are then inputted into KIMA for stellar activity decorrelation, following the method in Anna John et al. (2022). TWEAKS has been shown to successfully achieve sub-m s^{-1} precision and accurate planet characteristics (Anna John et al. 2023; Dalal et al. 2024). While fast and able to successfully preserve any planetary signal, this method, as the one presented in the previous Section, still shows significant quasi-periodic variations left in the "clean" RVs, and therefore needs to be paired with other mitigation techniques.

2.1.3.3 CCF-based Deep Learning

Neural networks can also help to differentiate between Doppler shifts and stellar activity-induced distortions at the CCF level (De Beurs et al. 2021). While the training set needed for successful mitigation of stellar variability requires large numbers of observations, it does not need to be well- or densely-sampled. De Beurs et al. (2021) used a Convolutional Neural Network (CNN) to predict the difference between a simple Gaussian fit to the CCF (as described in Section 2.1.1.1) and the true Doppler shift in the data by learning the contribution of stellar variability to the RVs. They find that a CNN architecture works better for this task than other NN algorithms, as they have only sparse interactions. This means that the neural network is forced to learn local features over the entirety of the training set at the same time, and then exploits the spatial structure. This method relies

---

[6] Code available at: https://github.com/j-faria/kima

on the fact that planets impart translational shifts to the spectrum and therefore to the CCF, while stellar variability does not. To allow the CNN to only learn distortions of the line profile, all the CCFs are normalised to a common zero-point velocity. All CCFs are essentially shifted by the radial velocity measured with a simple Gaussian fit. A ΔCCF is then extracted by subtracting a "quiet" reference frame. This master CCF profile is meant to contain no effects due to stellar activity. In the case of solar data, it was selected to be the average all observations taken on a day with no visible active region on the solar surface. In the case of stellar data, this reference CCF is usually computed as the mean of all observations. All the epoch of normalised ΔCCFs are then inputted in the neural network as a training set. De Beurs et al. (2021) trained the CNN on both simulated data and solar observations, and found that they successfully removed a significant amount of stellar scatter, reaching sub-$\mathrm{m\,s^{-1}}$ precision. This method has however yet to be extended to stellar observations, which could be done in one of two ways. On one hand, the technique could be applied on single stars with large enough datasets. This would however limit the complexity of the model, as it is unlikely that all types and levels of stellar variability would be included in the dataset, and could lead to overfitting. On the other hand, the CNN could be trained on multiple stellar datasets from the same instrument, and attempt to predict stellar variability based on the whole ensemble. While this method would account for a large variety of complexity, all planetary signals from the different systems must be carefully completely removed before training. A further downside of this technique is that it requires a "quiet" master CCF, the creation of which is complicated at best.

2.2 Stellar Activity Proxies

The analysis of stellar variability in radial-velocity data can be complemented by the use of other, ideally independent, time series that directly or indirectly probe the effects due to the various sources of activity. These measurable quantities can be used as activity indicators, or proxies, when they are not affected by planetary interaction, and all their variation can instead be attributed to stellar processes. Any signal present in both the RVs and an activity indicator can therefore be attributed to stellar activity (note that however the opposite is not necessarily true). Proxies most often correlate directly to their contemporaneous RVs. In this Section, I describe the most commonly employed activity indicators, and catalogue them as either derived from the spectral data, or from the shape of the cross-correlation function.

2.2.1 Derived from the Spectrum

Stellar activity affects the spectrum of stars by introducing distortions in the profile of the spectral lines. In some cases, strong lines can be used to compute proxies for the activity by, for example, assessing the difference between their measured shape and the shape they are expected to have in the case of no magnetic fields.

2.2.1.1 S-index

The S-index (Wilson 1968) is the most commonly used spectroscopic activity indicator, and it is most often employed for removing long-term signals in the RVs due to the magnetic cycles of stars. It is a chromospheric index, as it relies on information derived from chromospheric lines. Although the exact mechanisms by which the stellar chromosphere is heated remain an argument of debate, the existence of a link between heating and magnetism is undeniable (Brun and Browning 2017). Therefore, the presence and intensity of specific heating phenomena can be used as proxy for stellar magnetic activity (Luhn et al. 2020). Long-term monitoring of the solar surface has shown that during times of high activity the flux of various calcium absorption lines brightens significantly in narrow band filters centred on the near-UV Ca II wavelengths. These lines, arising at 3969 and 3934 Å from slightly ionised calcium, are the strongest spectral features observable with ground telescopes, and were dubbed the H and K lines by Fraunhofer in 1814. The Mount Wilson Program (1966 – 2003) established that these lines could in fact be used as proxies of the solar activity cycle. The unit-less index S_{HK}, or simply S-index, is defined as the ratio between flux in the Ca II line cores and the flux in the nearby continuum regions (Vaughan 1983). The S-index doesn't require absolute flux calibration, and it provides distance-independent activity information (Isaacson and Fischer 2010). Mathematically it can be summarised as

$$S_{\mathrm{HK}} = \alpha \frac{\Psi_{\mathrm{H}} + \Psi_{\mathrm{K}}}{\Psi_{\mathrm{V}} + \Psi_{\mathrm{R}}}, \tag{2.1}$$

in which Ψ_H and Ψ_K are the fluxes of the H and K line cores, Ψ_V and Ψ_R refer to the fluxes of the continuum bands respectively on the violet and the red sides of the calcium lines, and α is a normalisation factor.

However, the S-index contains a colour term, derived from the reference bandpasses, and is therefore dependent on spectral type of the star. It also includes a photospheric contribution, making it a difficult parameter to use when considering stellar ensembles (Hall 2008). Although Middelkoop (1982) developed a method to remove the colour dependence and recent works have proposed new extraction techniques (Cretignier et al. 2024), S_HK values are often parameterised in the form $\log(R'_\mathrm{HK})$ (Noyes et al. 1984), the logarithmic fraction of the star's bolometric luminosity radiated as chromosperic H and K emission (e.g., Thompson et al. 2017). $\log(R'_\mathrm{HK})$ values range from -4.4 for very active stars to -5.1 for inactive ones (Hall 2008). Mathematically R'_HK can be expressed as

$$R'_\mathrm{HK} = \frac{\Psi'_\mathrm{H} + \Psi'_\mathrm{K}}{\sigma T_\mathrm{eff}^4}, \tag{2.2}$$

where Ψ'_H and Ψ'_K are the fluxes of the H and K line cores measured at the stellar surface (rather than by the observer as for the S-index) with chromospheric contribution of the reference star pre-subtracted, σ is the Stefan-Boltzmann constant and T_eff is the effective temperature of the star (Martıénez-Arnáiz et al. 2010). Nevertheless, the $\log(R'_\mathrm{HK})$ still shows dependency on stellar metallicity and luminosity. Both $\log(R'_\mathrm{HK})$ and S_HK are not expected to fully behave as the radial velocities, as they are affected by foreshortening and limb darkening in different ways to the observed RVs: Ca II emissions remain bright even near the stellar limb. They are also not expected to be as effective over all levels of activity, and they may not be sensitive enough to mitigate stellar variability in the low-activity regime (Desort et al. 2007). Moreover, they both require high SNR for accurate computations, as do all activity proxies derived from the spectrum.

2.2.1.2 Hα-index

The Balmer Hα line at 6562.808 Å is also often used as an activity indicator (Zarro 1983; Herbst and Miller 1989). The flux of Hα emission lines is, like the S-index, able to trace the long-term trends due to stellar magnetic cycles. In particular, it is expected to be especially sensitive to chromospheric plages, as the brightening of Hα cores due to magnetic activity is well-established (Athay 1976). Like in the case for the S-index, the depth of Hα lines shows dependencies on chromospheric heating, and thus on the presence of magnetic fields. Reiners et al. (2012) estimated the continuum around Hα by taking the median of two regions on either sides of the line and used this indicator to inform their analysis of stellar activity. The Hα index is most often defined as

$$I_{\mathrm{H}\alpha} = \frac{\Psi_{\mathrm{H}\alpha}}{\Psi_{\mathrm{R}} + \Psi_{\mathrm{B}}}, \tag{2.3}$$

in which $\Psi_{\mathrm{H}\alpha}$ is the flux computed at the centre of the Hα line, and Ψ_{R} and Ψ_{B} are the continuum fluxes at respectively the red and the blue sides of the line (Boisse et al. 2009). The flux of the Hα line is now one of the staple stellar chromospheric proxies (e.g., Robertson et al. 2014). Gomes da Silva et al. (2011) found a correlation between the Ca II lines indicator and the flux of Hα for high activity, but in the low-activity regime this correlation disappears. In some cases, Hα fluxes can be preferred to S_{HK} values, as they are less affected by changes in the flux of the ThAr calibration lamp in the telescope. The Hα-index still has its limitations: as a tracer of plages it is only truly effective when considering plage-dominated stars (e.g., Medina et al. 2022), and due to the photoionisation process on which its signal depends it is more useful in M-dwarfs rather than F-, G- or K-type stars.

A few other spectral activity indicators are worth mentioning. The flux and equivalent width of Nα, or Na I D lines (Andretta et al. 1997), the Balmer Hϵ line and the He I D line (Maldonado et al. 2019) have all been studied. For a comprehensive list, Wise et al. (2018) presented a pipeline to extract 40 different activity indices from observational spectra. In all cases, spectral activity proxies require high signal-to-noise ratios in order to resolve individual line changes. Depending on the chosen line and computation technique, they are also often prone to issues due to blending effects.

2.2.2 Derived from the CCF

Alternatively, the shape distortions caused by stellar activity can be investigated starting from the cross-correlation function. The CCF can be considered the average profile of

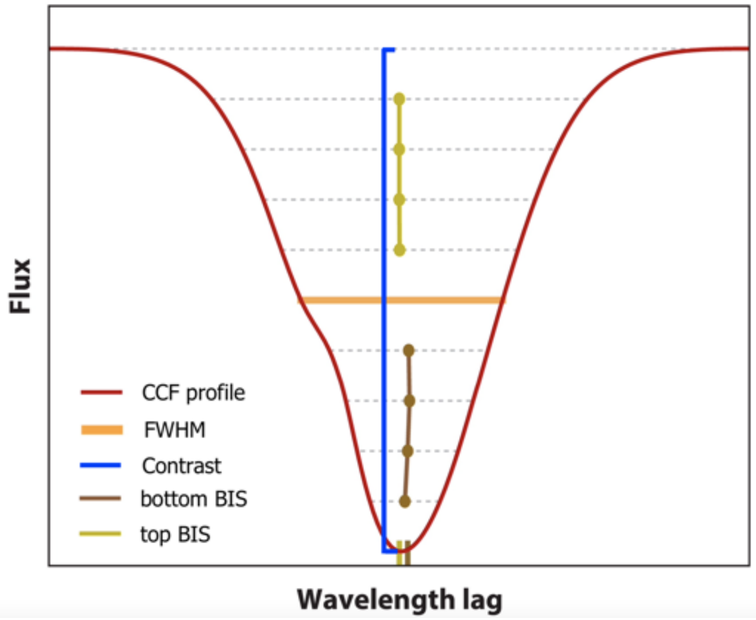

Figure 2.3: Schematic of the cross-correlation function (CCF) and its common activity indicators. The CCF is plotted as a red solid line, and a distortion caused by stellar activity can be seen in its left limb (for the sake of visibility the stellar activity inducing this change in line shape is exaggerated as a spot covering 20% of the stellar surface). The full-width at half-maximum (Section 2.2.2.1) of the CCF is indicated by the orange band. The contrast is shown as a blue bar (Section 2.2.2.2). The two velocities (here in the form of wavelengths) computed from the top and the bottom sections of the bisector (Section 2.2.2.3) are highlighted in light and dark brown respectively. Figure adapted from Fig. 4 of Hara and Ford (2023).

all the spectral lines. Common deformations in different lines will therefore contribute to distortions in the CCF. CCF-derived activity proxies are particularly useful in cases where the signal-to-noise ratio is too low for the computation of the previously addressed spectral indicators.

2.2.2.1 Full-Width at Half-Maximum

The most conventional activity proxy is the Full-Width at Half-Maximum of the CCF (FWHM: e.g., Queloz et al. 2009; Hatzes et al. 2010). The FWHM is computed as the difference between the two velocities (or wavelengths) on either side of the Gaussian fit of the line profile at which the flux of the CCF is equal to half its peak value, as shown in Fig. 2.3. The FWHM is dependent on the projected stellar rotation rate, $v \sin(i)$, and on the intrinsic line width due to thermal and turbulent photospheric motion (Desort et al. 2007). Most importantly, the FWHM changes as dark or bright regions traverse the visible disc of the star. Specifically, it increases in the presence of spots: as the missing flux region crosses the profile, it reduces the depth of the normalised line while simultaneously mostly preserving its area. The width of the CCF will thus widen to compensate (Collier

Cameron et al. 2019). The opposite is true for bright regions, although the sensitivity of this indicator to faculae and plages is significantly reduced due to the less drastic temperature contrast. Because of this difference, the FWHM is generally considered a proxy for spot activity, rather than for RV variability induced by all active regions. Late G- and K-type stars present positive correlation between the FWHM of the CCF and the $\log(R'_{\mathrm{HK}})$, indicating that their rotationally modulated activity is most likely driven by dark spots (Costes et al. 2021). For mid-G-type stars this behaviour transitions to a weak correlation. Costes et al. (2021) speculated that this change in behaviour can be explained by the shallower convective envelopes in early G- and F-type stars, resulting in weaker magnetic fields which are less likely to form spots. They therefore confirmed that the FWHM is not an effective activity proxy for faculae-dominated stars. The FWHM also presents other limitations: it is instrument-dependent (being sensitive to e.g. focus drifts), it is time-shifted with respect to the RVs, and it aims to trace distortions averaged over hundreds of lines with different physical characteristics and different responses to stellar variability.

2.2.2.2 Contrast

The contrast of the CCF is defined as the depth of the CCF at its central wavelength (e.g., Lanza et al. 2019), as depicted in blue in Fig. 2.3. In the Sun, it is anticorrelated with the stellar magnetic cycle, increasing with decreasing activity (Collier Cameron et al. 2019). It is thought to be similarly affected by bright and dark regions. In particular, the long-term variability of the contrast indicates that is it also sensitive to the facular network (Costes et al. 2021). The contrast of the CCF is generally anticorrelated to the FWHM, showing sharp dips at the passage of dark spots on the solar disc. By extension of this behaviour, the area of the CCF does not seem to be particularly affected by rotationally modulated effects, as changes in the width and depth of the line profile compensate each other. The contrast shows strong anticorrelations with the $\log(R'_{\mathrm{HK}})$ for more active stars, and weak to non-existent correlations in the case of quieter stars. The observed anticorrelation between the contrast and the $\log(R'_{\mathrm{HK}})$ is expected: heightened stellar activity results in an increase of bright region filling factor boosting the continuum (Cegla et al. 2013) and in fictitious "emission" bumps in the line profile (Thompson et al. 2017), both of which decrease the profile contrast and enhance the flux in the Ca II lines (Costes et al. 2021). Nevertheless, the contrast shares the same limitations as the FWHM, and its computed value is also dependent on the choice of CCF line mask.

2.2.2.3 Bisector

Another way to measure the deformation of the CCF due to stellar variability is to assess the changes in the bisector of the line profile. Generally CCF bisectors display the classic "C" shape generated by convective blueshift due to granulation, as shown in Fig. 1.9 in Chapter 1 (Gray 1989). The presence of active regions inhibits the convective flow and therefore reduces the total amount of convective blueshift. Deviations from this characteristic shape can therefore be used as a proxy of active-region variability in the radial velocities. Recent works have also studied the "dance" of the bisector caused by the evolution of the granulation pattern on high-SNR observations of the Sun, and have modelled the effects of granulation in the line profile (Palumbo et al. 2022). There are a variety of methods to assess bisector distortions. The earliest formalism was developed by Queloz et al. (2001). They introduced the bisector span (BIS) as a measurement of the asymmetry of the CCF. They select two regions at the top and bottom of the line profile in order to measure changes in the orientation of the bisector. The difference between the average velocity of the top of the bisector, V_{t}, and the one of the bottom section, V_{b}, is a measure of the inverse mean slope of the bisector, as shown in Fig. 2.3. The choice of sections of the CCF used for this computation changes between works, but generally the top is defined to include from 10 to 40% of the CCF based on line depth, while the bottom segment covers the 60 to 90% segment. The interpretation of this proxy is simple: any variability in the RVs with the same periodic evolution as changes in the BIS is most likely generated by shape distortions rather than pure shifts, and thus can be attributed to stellar activity. In particular, the BIS is generally anticorrelated with stellar-induced signals in the RVs. However, the computation of the BIS is particularly unreliable at low signal-to-noise ratio, and this anticorrelation is not straightforward. In fact, the relationship between the BIS and the RVs has been proven to not be monotonic (Boisse et al. 2011; Figueira et al. 2013). It instead follows a tilted "8" shape, with the angle of the tilt and the width of the shape dependent on the configuration of activity regions on the stellar surface. Desort et al. (2007) also showed that in some cases the bisector span is actually less sensitive to the presence of deformations in the CCF than the RVs. Boisse et al. (2011) further highlighted that any correction applied to the RVs using the BIS does not remove the periodicities due to stellar activity, but only reduces their amplitudes. They instead proposed another approach for the analysis of bisector asymmetry: the V_{span}. They computed two radial velocities by iteratively fitting Gaussian functions to the top and the bottom parts of the CCF, so that each RV is only sensitive to distortions in its respective segment of CCF. They then defined V_{span} to be the subtraction of the bottom RV from the top one. This method relies on the fact that the bottom part of the CCF is most sensitive to stellar activity. As for the BIS, the selected ranges can change. The V_{span} is more robust than the BIS at low signal-to-noise ratio. Similarly to its predecessor, the V_{span} is generally anticorrelated to the RV variations due to stellar activity, but their relationship still maps a tilted "8", with similar properties as the BIS-RV correlation. The calculated values of BIS and V_{span} are very similar in most cases, with variations proportional to the spot filling factor and to the projected rotational velocity of the star. A further evolution of the formalism was then introduced by Figueira et al. (2013): the ΔV and the V_{asy}. The ΔV is calculated as the subtraction between the central RV computed by fitting a bi-Gaussian curve to the CCF and the central RV computed by fitting a traditional Gaussian. The bi-Gaussian function allows for the two wings of the distribution to be characterised by different half-widths at half-maxima (Nardetto et al. 2006), and is therefore more sensitive to line deformations. ΔV is thus a measure of the shift induced to the RVs by asymmetries in the CCF, and it shows a positive correlation to activity-induced variations in the radial velocities. V_{asy} describes the difference of information content between the blue and the red wings of the CCF. It is spiritually closer to an RV calculation rather than a direct line profile indicator, and it is expected to be positively correlated to the RVs. In fact, the correlation between V_{asy} and the radial velocities is always stronger than the one between the BIS and the RVs (Figueira et al. 2013). Although the relationship with the RVs is still not wholly monotonic, it is significantly more linear than for all other bisector indicators. Further parametrisations have recently been proposed such as the skew normal density (Simola et al. 2019), but they are still in their infancy.

2.3 Modelling the RV Time Series

As covered in Section 1.2, stellar variability affects the radial velocities in complex ways, over multiple timescales and with different intensities. Even the most successful techniques are not able to fully account for all sources of stellar variability. Further modelling of the extracted RVs in the time-domain is therefore often recommended to at least understand the structure of the remaining stellar signal. Moreover, even in cases with little to no stellar variability present in the RVs, preliminary analysis of the time series allows for a better understanding of the planetary signals. This Section describes the main techniques to model and analyse radial-velocity time series.

2.3.1 Periodograms

One of the most useful and straightforward ways of investigating a variable time series is to compute its periodogram. The aim of a periodogram analysis is to identify the leading frequencies dominating the oscillations of the data. A periodogram is a tool that allows for the analysis of a time series in the frequency-domain by fitting a function to the observational data. Periodograms rely on the assumption that any time series can be expressed as a sum of cosine and sine waves of different amplitudes and periods. Scargle (1982) introduced the periodogram as “the estimation of the spectral density of a signal”. The algorithm for the now staple Lomb-Scargle periodogram was designed specifically for the detection and characterisation of periodic signals, such as planet-induced radial-velocity oscillations, in unevenly sampled time series. On a very basic level, the Lomb-Scargle periodogram is equivalent to iteratively fitting sine waves in the form $y = A \cos \omega t + B \sin \omega t$ to the time series, in which A and B are amplitude coefficients and ω is every frequency in the chosen range. The algorithm then measures how well the fitted model matches the inputted data. A peak in the periodogram means that the observational data can be well-modelled by a sine curve with the specified frequency. Since their inception, periodogram algorithms have become increasingly complex and tailored to their specific function. In this Section, I cover the most commonly employed formalisms of periodograms in the context of exoplanet hunting and mitigation of stellar variability.

2.3.1.1 Generalised Lomb-Scargle Formalism

Zechmeister and Kürster (2009) introduced the Generalised Lomb-Scargle (GLS) periodogram[7]. This updated formalism of the Lomb-Scargle algorithm aims to correct its two main drawbacks: not taking into account the uncertainty values of the measurements, and the assumption that the mean of the data and the mean of the fitted sinusoidal are the same. The first shortcoming was solved by introducing weighted sums to the definition (Gilliland and Baliunas 1987; Irwin et al. 1989). The second was overcome with the inclusion of an offset c, generalising the periodogram fitting function to a weighted full sine wave described as $y = A\cos\omega t + B\sin\omega t + c$ (Cumming et al. 1999). The GLS is a hybrid algorithm that employs in part an analytical solution for linearised parameters and in part steps through non-linear parameters. An example of a periodigram computed with the GLS formalism is shown in the top panel of Fig. 2.4, for more examples see Chapter 4. However, the GLS still expresses its results in an arbitrary power, which does not allow for easy comparison between models. In order to circumvent this drawback, a False Alarm Probability (FAP) is estimated. FAPs are a way of expressing uncertainties in the periodogram results and to differentiate significant peaks from the spurious background noise. They measure the probability that a peak at the same frequency could be produced by coincidental random scatter in data (VanderPlas 2018). Periodogram powers that exceed sufficiently small FAP levels can therefore be considered statistically significant (for further examples, see Chapter 5). There are a variety of methods to compute false alarm probabilities, but all of them are dependent on the normalisation applied to the periodogram. Alternatively, a False Inclusion Probability (FIP: Hara et al. 2022) has been proposed to substitute the FAP and to better compare the probability of different signals, in particular in the case of aliasing.

2.3.1.1.1 Bayesian Generalised Lomb-Scargle Formalism A further approach to allow for direct comparison between peaks is to extend the definition of the GLS to include Bayesian probability theory. Mortier et al. (2015) further generalised the GLS to the Bayesian Generalised Lomb-Scargle periodogram (BGLS)[8]. The BGLS computes the prob-

---

[7] Code included in the astrophysical Python package ASTROPY and available at: https://github.com/mzechmeister/GLS

[8] Code available at: https://anneliesmortier.wordpress.com/sbgls/

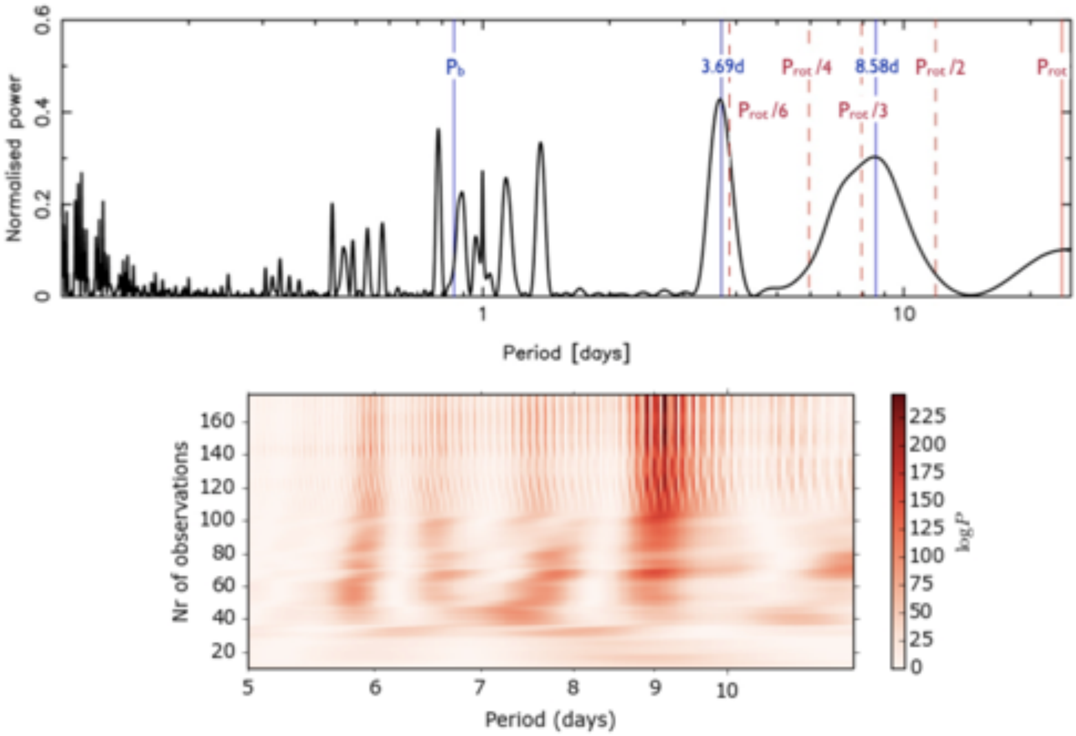

Figure 2.4: Example periodograms of radial-velocity data of CoRoT-7. *Top panel:* Generalised Lomb-Scargle periodogram of the 2012 CoRoT-7 RV dataset. The stellar rotation period and its harmonics are identified by red vertical lines. The orbital period of CoRoT-7b computed via transit photometry (Barros et al. 2014) is shown in blue, alongside the periodicities highlighted by Haywood et al. (2014). The Figure is taken from Fig. 3 of Haywood et al. (2014). *Bottom panel:* Stacked Bayesian Generalised Lomb-Scargle periodogram of the same data. The logarithmic probability of each period is depicted as a colour intensity. The ~8.5 days signal found in the top GLS can here be clearly seen to be of incoherent nature, making its origin more likely to be stellar rather than a planet. The Figure is taken from Fig. 3 of Mortier and Collier Cameron (2017).

ability that a signal of a specific frequency is present in the data sample, therefore allowing for direct assessment of the relative probability between multiple periods. Following the maths in the paper, the probability power computed in a GLS is part of the exponent that expresses the normalised probability of the BGLS. Overall, the BGLS performs comparably to the GLS at recovering the correct frequencies, and it allows for a clearer comparison between peaks.

2.3.1.1.2 Stacked Bayesian Generalised Lomb-Scargle Formalism Mortier and Collier Cameron (2017) further extended the BGLS to a Stacked Bayesian Generalised Lomb-Scargle (SBGLS) formalism. This technique aims to distinguish between periodicities generated by the presence of planets, and those due to the effects of stellar activity. It relies on the fact that pure Doppler shifts are coherent signals, meaning that with increasing number of observations their probability power will only increase. On the other hand, most stellar variability is inherently incoherent, with varying amplitude and phase. The probability of an incoherent signal changes with different observations and different phase coverage. The SBGLS computes the power associated with fitting the sine wave to each frequency for n observation epochs. Then, another random datapoint is added and the process is repeated. This routine continues until all data have been included. Plotting the probability power as an intensity against period and number of data points, as shown in the bottom panel of Fig. 2.4, highlights how the significance of a signal changes with number of observations. This "stacking" of periodograms allows to assess the coherency of the considered best-fit signal. A peak with a probability that oscillates with number of epochs can therefore be attributed to stellar variability. However, caution is recommended and not all seemingly coherent frequency peaks can automatically be assumed to represent the period of Keplerian oscillations. Mortier and Collier Cameron (2017) also showed that aliases of the same periodicity behave similarly to each other and to their true signal, allowing for easier connection between them. However, it is also of note that all the described periodogram formalisms share a vital flaw: they assume that only one frequency is able to fit all the variability in the data. In other words, they do not account for the presence of multiple sources of signal. They fit one frequency at a time, and can therefore fail to properly distinguish the correct model.

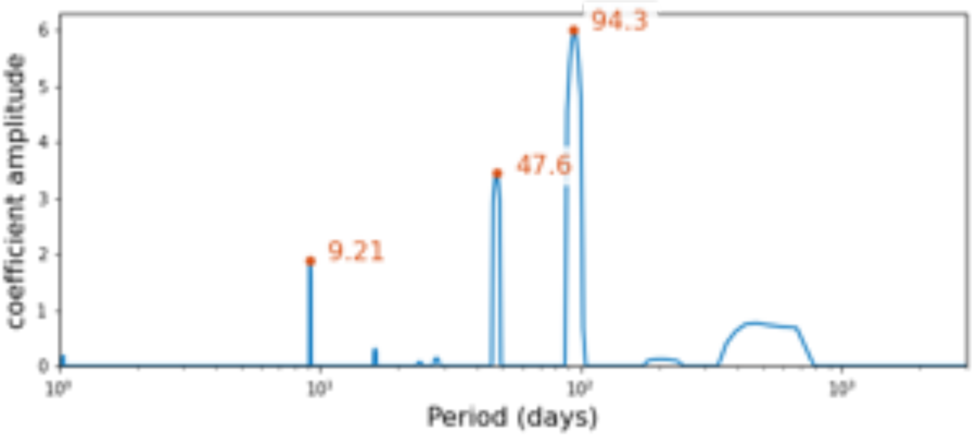

Figure 2.5: Example of $\ell 1$ peridogram of the HARPS-N radial-velocity data published in Rescigno et al. (2023b) for the characterisation of the TOI-2134 system (see Chapter 4, Section 4.1). The most probable periods are highlighted in red.

2.3.1.2 $\ell 1$ Formalism

To overcome this limitation, Hara et al. (2017) introduced the $\ell 1$ periodogram (Hara and Mari 2021)[9]. This formulation aims to avoid the residual distortions that are generated by iterative fitting. To do so, they use the theory of compressed sensing, adapted for handling correlated noise. Differently from a GLS, the $\ell 1$ periodogram searches for a representation of the input signal as a sum of sinusoidal waves, the frequencies of which are taken from the considered frequency range. In basic terms, it fits all frequencies at the same time. By assuming that the number of frequencies that describe the inputted signal is much smaller than the number of observations, they rely on "sparse recovery" tools for fast computations, more precisely on the Basis Pursuit minimisation problem (Chen et al. 1998) and the group-LASSO (Yuan and Lin 2005). The $\ell 1$ periodogram uses the assumption that the noise in the data is Gaussian and that it is correlated through a weighting matrix. On top of the usual frequency grid, this formulation thus requires the definition of a covariance model to describe the noise, and of a list of unpenalised model vectors (the assumed properties of the data, such as an offset). The resulting periodogram usually presents much fewer peaks than a traditional Lomb-Scargle periodogram, and instead it hones-in on the main periodicities, as shown in Fig. 2.5. Generally, $\ell 1$ periodograms also tend to be more robust against aliasing effects. Hara et al. (2017) recommended the use of the $\ell 1$ periodogram as a complement to classical GLS algorithms. Most interestingly, a failure of

---

[9] Code available at: https://github.com/nathanchara/l1periodogram

this formalism (meaning the presence of forests of peaks, rather than clean single-peaked signals) can indicate that the noise in the data is either higher than expected or that it is non-Gaussian.

2.3.1.3 Welch Formalism

A different approach to frequency searches from the traditional Lomb-Scargle algorithm was developed by Welch (1967) building on the work of Bartlett (1948). It does spectral density estimation based on magnitude-squared coherence. This bivariate statistic aims to isolate oscillations that manifest in more than one observable. In the case of planet searches, the considered time series are the RVs and the stellar activity proxies. It can be interpreted as a frequency-dependent correlation coefficient. The Welch formalism yields cleaner spectral windows and lower variance than the GLS, and is primarily useful for the identification of stellar signals (Dodson-Robinson et al. 2022). For unevenly sampled data, no direct cross-correlation can be computed. Welch's method overcomes this problem by utilising non-uniform fast Fourier transforms (similarly to Lomb-Scargle periodograms). The uniqueness of this formalism lies in the preparation of the data. Given that it aims for a variance decrease with increasing number of observations (the opposite to traditional periodograms), the Welch formalism requires a segmentation of the dataset. While many non-overlapping segments are ideal, this method introduces significant bias for too-small numbers of observations in each data section. There is therefore a trade-off between reducing the variance (by increasing the number of splits) and avoiding consistent bias (by keeping the number of observations in each segment as large as possible). Welch (1967) therefore proposed to use overlapping bins with tapered ends. A 50% overlap between splits can significantly increase the number of datapoints, but the computed spectral estimates are no longer independent. Tapers are thus added to each segment in order to minimise spectral leakage and most vitally to increase the independence between estimates of overlapping sections. Tapers are also particularly useful for enhancing the detectability of weak signals in the neighbourhood of much stronger ones. The formalism pioneered in Welch (1967) is also fully described in Dodson-Robinson et al. (2022), and has since been packaged as NWelch (Dodson-Robinson 2022)[10]. While Welch periodograms are

---

[10] Code available at: https://github.com/sdrastro/NWelch

computationally cheap and can be easier to interpret than GLS periodograns, they also present a new set of limitations. First and foremost, they generally require larger datasets to successfully work. A split in two or more segments is recommended, with each section including at least 100 datapoints. This formulation therefore requires times series with a minimum of 150 epochs, assuming a 50% overlap.

2.3.2 Structure Functions

Structure functions (SF) are a technique for the analysis of time series data that complements traditional periodograms. They provide a method of variability quantification that is not sensitive to common problems such as windowing and aliasing (Hughes et al. 1992). They are effective even with extremely unevenly-sampled data, assuming enough variety in the epoch separations, and they are most useful to characterise aperiodic signals, which Lomb-Scargle periodograms fail to detect. Structure functions as a tool have been extensively used in extra-galactic work (e.g., Simonetti et al. 1985), and have recently been introduced to the analysis of radial velocity data (Lakeland et al. 2024). On a basic level, structure functions consider all pairs of observations $f(t)$ separated by the same time τ, and for each τ bin they assess the scatter within the data. A structure function is computed as

$$\mathrm{SF}(\tau) = \left\langle (f(t) - f(t-\tau))^2 \right\rangle . \tag{2.4}$$

To allow for easier interpretation, the value of the structure function is usually related to the RMS of the time series as $\mathrm{SF}(\tau) = 2{\cdot}\mathrm{RMS}^2$ for a Gaussian noise signal. A small value of SF means that two observations separated by the considered time lag τ show little variation, and one can be used to more precisely predict the other. Conversely, large SF values indicate significant variance between observations spaced by the considered τ. Typically, for most types of variability, structure functions increase with increasing time separation between the considered epochs, as sources of scatter from different processes at different timescales are cumulative. Periodic signals are the most notable exception to this rule: the SF of a perfectly periodic signal decreases to zero at multiples of the period. This behaviour is straightforward to interpret, as observations separated by a full period are equivalent in value. In most astrophysical cases, where a periodic signal is

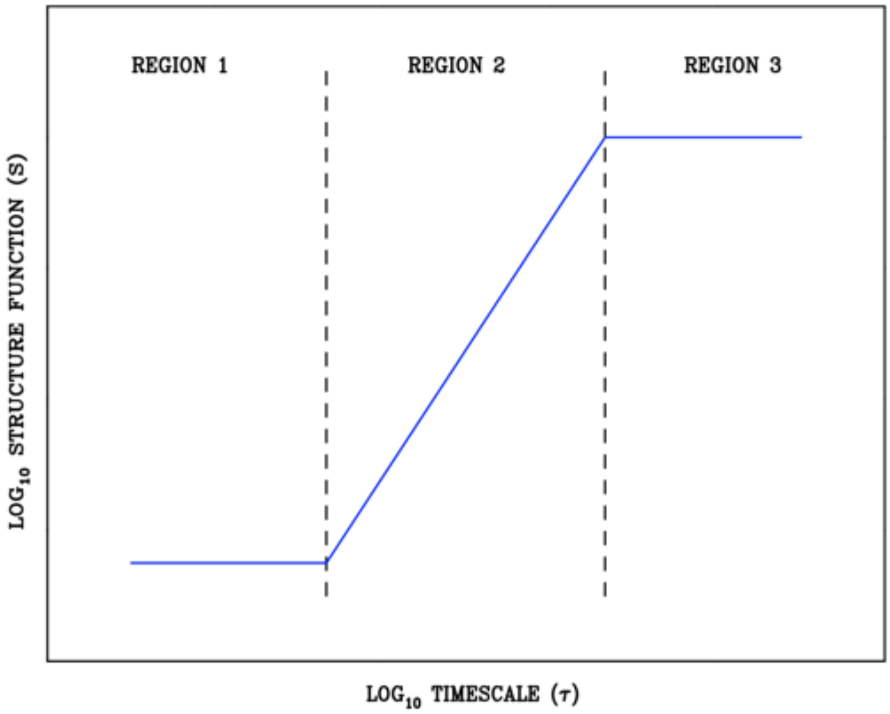

Figure 2.6: Diagram of the typical behaviour of structure functions, and the separation in three main regions, described in the text. Figure adapted from Fig. 7 of Sergison et al. (2020).

either partially disrupted by measurement uncertainties or is subject to changes in time, the SF would still decrease but not reach zero, as observations with time difference close to the period are still more similar than those with separation of any other phase. As Sergison et al. (2020) highlight, a typical structure function can be separated in three main regimes, shown in Fig. 2.6. In the first region, the SF is dominated by the uncertainties of the data and other uncorrelated jitter. This variability level can be interpreted as the noise floor of the observations. In Region 2, the SF starts increasing, generally following a power-law behaviour. In this regime, the structure function is probing the variability of the correlated noise, and its gradient is dependent on the frequency spectrum of the sources of scatter. Region 3 is characterised by a plateau. For timescales longer than the starting τ of this regime, no additional intrinsic variability is present. The transition point between Regions 2 and 3 is called the knee of the structure function and represents the characteristic timescale of the probed source of scatter. In practical terms, to effectively sample a specific source of variability, observations should be taken with time lags within the second region of the SF. In cases of complex signals, it is possible for the structure functions to present multiple increasing regions followed by plateaus, each probing dif-

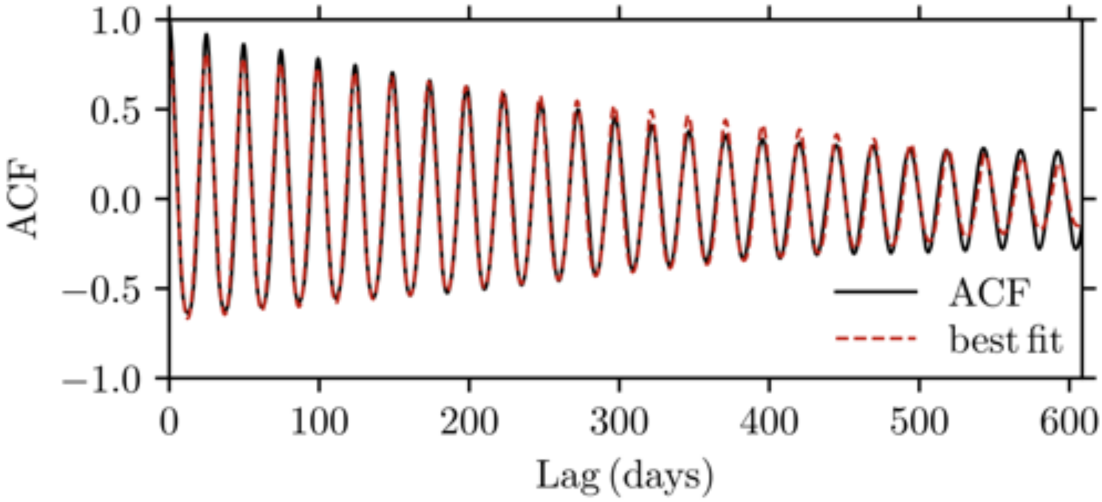

Figure 2.7: Example autocorrelation function of a simulated Kepler-like lightcurve plotted as a black solid line. The best-fit model defined as the sum of a cosine and sine multiplied by a linear decay term is shown as a red dashed line. Figure adapted from Fig. 2 of Santos et al. (2021).

ferent sources of variability. Different signals with different characteristic timescales will contribute to different knees in the structure function. Overall, structure functions are a tool to assess the frequency spectrum of a variable signal. They are best used in combination to Fourier techniques, but have their advantages. In particular SFs are significantly less dependent on sampling, as they are computed in the time-domain. Moreover, as long as enough epoch time-differences are sampled, they are not impacted by large gaps in the data. However, structure functions do not take into consideration the uncertainties on the measurements, and they require a large enough number of observations so that all timescale bins are properly populated in order to avoid small number statistics.

2.3.3 Auto-Correlation Function and Lag Analysis

The Auto-Correlation Function (ACF: McQuillan et al. 2013; Giles et al. 2017; Collier Cameron et al. 2019) is another method to identify variability timescales in time series data. The ACF measures the degree of self-similarity of a dataset over a range of time lags. Mathematically it is described as:

$$r_\tau = \frac{\sum_{t=t_0}^{t=t_{\rm tot}-\tau}(y(t)-\overline{y})(y(t+\tau)-\overline{y})}{\sum_{t=t_0}^{t=t_{\rm tot}}(y(t)-\overline{y})^2}, \tag{2.5}$$

in which r_τ is the auto-correlation coefficient of the time series y with epochs $t = [t_0, ..., t_{\text{tot}}]$ at time lag τ. In practice the ACF is a measure of how well the original dataset correlates with itself when shifted in time of τ. The auto-correlation function generally requires evenly-sampled data, but clever binning techniques have been developed to allow the computation of ACFs for unevenly-sampled datasets (Edelson and Krolik 1988; Robertson et al. 2015). If the considered time series includes a periodic signal, the ACF will present repeated peaks in the correlation value at time lags regularly spaced by integer multiples of the period of the variability, as observations separated by a period are more similar and therefore more correlated, as shown in Fig. 2.7. For a systematic method of period extraction, see McQuillan et al. (2013). In the case of complex signals, such as the effects of multiple active regions rotating in and out of the visible stellar disc, the ACF can show peaks at time lags equal to harmonics of the period. The decrease in amplitude of these side-lobes is driven by the evolution of the signal, such as the decay of the active regions. We can therefore relate the decay rate of the secondary peaks to the evolution timescale of the probed variability. In an auto-correlation analysis, a varying amplitude or phase of the periodic signal yields a steady decrease of the height of the main peaks, creating a sort of envelope. The slope of this envelope is dependent on the magnitude of the change. Given these behaviours, the ACF is often analytically modelled as an underdamped simple harmonic oscillator with an added inter-pulse term (Giles et al. 2017) or as the sum of a cosine and sine multiplied by a linear decay term (Santos et al. 2021) (plotted as a red dashed line in Fig. 2.7). When the analysed time series is characterised by long-term trends or includes "jumps" in the median value, a significant signal for large time lags can be introduced in the ACF, changing the shape of the damping envelope. These signals will however only change the amplitude of the repeated peak, not its τ. They therefore do not impact period extraction. Overall, the auto-correlation function can be a powerful tool to identify the periodicity of a dataset, with generally clear results even with evolving signal structure, as for RV signals dominated by the effects of active regions. ACF are particularly good at producing robust results in cases of long instrumental systematics and other uncorrected correlated noise, as they only affect the height of the maxima, not their time lag.

Lag analysis follows the same formalism as ACF, but instead of computing the correlation of a dataset with itself, it calculates how a time series correlates with a second one shifted in time. In this case, a peak in the correlation coefficient represents a likely time lag between the two considered time series. This method has been used for example to identify the phase difference between radial velocities and stellar activity indicators (e.g., Collier Cameron et al. 2019; Costes et al. 2021). Sections 5.2.3.6 and 5.2.3.7 in Chapter 5 use the ACF and lag techniques to analyse the longitudinal magnetic field of the Sun.

2.3.4 Gaussian Process Regression

Gaussian processes have been been extensively employed in astrophysical literature to model stellar activity-induced variations and instrumental noise in both radial-velocity and photometric surveys (e.g., Haywood et al. 2014; Faria et al. 2016; Serrano et al. 2018; Barros et al. 2020; Barragán et al. 2022; Rescigno et al. 2023b). GPs allow the user to model the variability of a signal without needing to assume its functional form. They only require pre-existing knowledge of the correlation matrix of the dataset. An in-depth explanation and analysis of Gaussian processes and GP regression for the detection and characterisation of exoplanets is covered in Chapter 3. In summary, GPs represent a data-driven approach for activity mitigation. They output a variability model built solely on a covariance function, and learnt from the data offered as a training set. Gaussian process regression allows to model the stellar variability simultaneously to the planetary signals. Within the context of mitigation methods, GPs are used in combination with parameter space searching algorithms such as Markov Chain Monte Carlo (e.g., Foreman-Mackey et al. 2013) or χ^2 optimisation in the quest to find the hyperparameters that best fit the stellar variability. The quality that makes Gaussian processes the ideal tool for stellar activity modelling is their ability to recognise patterns. The covariance hyperparameter optimisation step is what is referred to as the "regression". GP regression is a very active field of study, with multiple new developments being implemented every year (e.g., Rajpaul et al. 2015). In this Section, I briefly cover the three main formulations to apply Gaussian processes to the analysis of radial-velocity data.

2.3.4.1 One-dimensional GPs

A "one-dimensional" analysis is the most straightforward application of GPs. In this formalism, the GP is trained only on the considered dataset, most often the RVs. I implement this method in Chapter 4 (Sections 4.1.6 and 4.2.7.2) and Chapter 5 (Section 5.2.4.3). This analysis is most efficient when some characteristics of the stellar signals, such as the period, are known or can be approximated. In some cases, stellar variability and planetary signals are modelled simultaneously with transits in the photometric data (as in Chapter 4, Section 4.1.7). While one-dimensional GPs have been successful at reducing the scatter due to stellar activity in RV datasets, they should be used with caution. GPs have a tendency towards overfitting, especially in cases in which the signals due to activity are not well-sampled. Overfitting is in this case defined as the incorrect fitting of random scatter present in the data as part of the correlated noise described by the covariance model. In order to minimise this behaviour, in some cases a white noise jitter term is added to the diagonal of the covariance matrix (Rescigno et al. 2023b). There are multiple publicly available code packages that can perform one-dimensional GP regression, each with their specific focus and computational algorithms. I here include a brief list of the most commonly used: GEORGE[11] (Ambikasaran et al. 2015) focuses on efficient evaluation of the likelihood of the GP model for large datasets, SPLEAF[12] (Delisle et al. 2022) extends the CELERITE (Foreman-Mackey et al. 2017) formalism and includes a more general class of noise models, TINYGP[13] is a small library built in jax for GP regression (Foreman-Mackey et al. 2024), and finally MAGPY-RV is the pipeline I developed and fully address in Chapter 3 (Rescigno et al. 2023a).

2.3.4.2 Multi-dimensional GPs

The flexibility that GPs offer often remains their major drawback. New techniques have therefore been developed to introduce further constraints to the fitting process. In this context, Rajpaul et al. (2015) pioneered the multi-dimensional Gaussian process regression. This technique builds on the algorithms of the FF′ method (Aigrain et al. 2012), which aims to model RV variation due to the presence of active regions on the stellar disc start-

---

[11] Code available at: https://github.com/dfm/george

[12] Code available at: https://gitlab.unige.ch/delisle/spleaf

[13] Code available at: https://github.com/dfm/tinygp

ing from information in the photometric data. The FF′ method relies on the assumption that, in the simplest case, a single spot on the equator imprints a signal in the photometry which oscillates as

$$\mathrm{F}(t) = \cos\left(2\pi t / P_{\mathrm{rot}}\right), \tag{2.6}$$

in which $\mathrm{F}(t)$ represents the projected area of the active region and P_{rot} is the rotational period of the star. In the RVs, the contribution of the same active region is two-fold (as addressed in Section 1.2.2): the suppression of convective blueshift can be approximated to vary as a function of $\cos 2\pi t / P_{\mathrm{rot}}$ (i.e. F), while the photometric inhomogeneities vary as a function of $\sin 2\pi t / P_{\mathrm{rot}}$, which is proportional to the first derivative of F. The variability due to active regions can therefore be modelled as the sum of the two effects multiplied by the projected spot area as

$$\Delta RV = A \cdot \mathrm{F}(t)\mathrm{F}'(t) + B \cdot \mathrm{F}^2(t), \tag{2.7}$$

in which A and B are constants that can be related to physical quantities. Aigrain et al. (2012) showed how this simple approximation can successfully mitigate and isolate stellar activity in a variety of cases.

Multi-dimensional GPs follow a similar formalism, but instead exploit correlations between spectroscopic activity indicators and RVs to constrain the stellar activity component of the radial velocities. They assume that the stochastic process giving rise to the variability is shared across multiple time series and that it can be described by a GP with the appropriate covariance function. The radial velocities are in this context modelled by latent GPs, rather than photometric variations. The second term of Eq. 2.7 can then be expressed in terms of a GP model as $\mathrm{G}(t) = \mathrm{F}^2(t)$, and by extension its first derivative can be written as $\mathrm{G}'(t) \propto \mathrm{F}(t)\mathrm{F}'$. Following this reparametrisation, the variability in the RVs can then be expressed as the sum of a suppression of convective blueshift term $\mathrm{G}(t)$ and a photometric term $\mathrm{G}'(t)$ in the form

$$\Delta \mathrm{RV} = A \cdot \mathrm{G}(t) + B \cdot \mathrm{G}'(t). \tag{2.8}$$

Activity indicators can also be modelled with the same Gaussian process function, depending on which physical processes the proxy is sensitive to. For example, variability in the S-index is dominated by the suppression of blueshift, thus in this formalism it is fully described by only the $\mathrm{G}(t)$ term multiplied by a different constant C as

$$\text{S-index} = C \cdot \mathrm{G}(t). \tag{2.9}$$

In cases in which the variability of the activity indicator is dependent on both the photometric inhomogeneities generated by the broken red- and blueshift symmetry, and the suppression of convective blueshift, as for the BIS, their model includes both terms as

$$\mathrm{BIS} = L \cdot \mathrm{G}(t) + M \cdot \mathrm{G}'(t). \tag{2.10}$$

This formulation can be extended to any time series whose signal is dominated by the same stellar activity processes that drive the scatter in the RVs, as long as a mathematical relation can be found. By modelling multiple time series with the same Gaussian process, the optimisation algorithm that finds the best covariance model parameters is less likely to overfit the data and attempt to include white jitter in the red noise model. This method has been shown to successfully disentangle stellar-induced variability and planetary signals (e.g., Barragán et al. 2019; Mayo et al. 2019). Various pipelines with adaptations of this technique are publicly available, most notably both PYANETI[14] (Barragán et al. 2022) and PYORBIT[15] (Malavolta et al. 2016; Malavolta et al. 2018) are able to model the stellar variability from multiple activity indicators simultaneously to planetary signals in the RVs, and GPLinearODEMaker[16] (GLOM: Gilbertson and Ford 2022) is a Julia package aimed at likelihood computation for multivariate GPs which also includes terms proportional to the second derivative of the latent GP. In all its forms, the multi-dimensional GP framework models all inputted time series of stellar indicators simultaneously to the RVs. It "reigns in" the flexibility of GPs by using the proxies as constraints for the stellar activity model. It is more sensitive to shared signals and periodicities than a simple correlation analysis and detrending. The ideal number of latent Gaussian processes needed to model RV data has also been investigated, and using two latent GPs (the first and second derivatives of $\mathrm{G}(t)$)

---

[14] Code available at: https://github.com/oscaribv/pyaneti

[15] Code available at: https://github.com/LucaMalavolta/PyORBIT

[16] Code available at: https://github.com/christiangil/GPLinearODEMaker.jl

was found to be most successful at reducing scatter (Tran et al. 2023). The main limitation of this formalism is that it only allows to describe activity-sensitive auxiliary time series as linear combinations of the GP and its derivatives, barring more complex relationships and therefore limiting its adaptability to other proxies. Moreover, in all cases current formulations of GPs all share a similar drawback. The ability of the GP to model the stellar signal in the data is dependent on the quality of the covariance function chosen to describe it. Current covariance functions such as the Quasi-Periodic kernel (QP, for a description see Chapter 3 Section 3.1.3) have been designed to model the variability in radial-velocity data produced by the presence of active regions. They therefore are not able to mitigate the effects of other sources of scatter such as supergranulation.

2.3.4.3 GP Regression Network

Another of the most egregious limitations of traditional GP analysis is that it assumes that the covariance form of the variability stays constant in time. Stationary covariance functions are applied even to data spanning multiple years and different phases of stellar magnetic cycles. However, recent works (e.g., Demin et al. 2018) have shown that Sun-as-a-star RV data is not well-modelled using a single stationary kernel, and that instead non-stationarity is required when analysing time series with longer baselines. In practice, over the course of the magnetic cycle the variability generated by activity on the Sun changes best-fitting hyperparameters. Therefore, for more precise prediction one should model separate chunks of the data with different GPs. However, this approach undermines the value of long baselines and has severe impacts on the detectability of small long-period planets, which require consistent observations over long timescales. In this context, Gaussian Process Regression Networks (GPRN: Wilson et al. 2012) represent the only tested technique to overcome this limitation. GPRNs combine the structural properties of neural networks with the non-parametric flexibility of GPs. They are able to create non-stationary linear combinations of independent stationary Gaussian processes. The covariance form can be shared between all latent GPs, or can be assigned individually. In the simplest case, when only two latent variables are considered, the resulting posterior is defined by the product of two independent Gaussian distributions. Camacho et al. (2023) further extended this formulation to include the possibility to use stellar activity indicators to constrain the GP model, following the formalism presented in the previous Section. They show that GPRNs can model highly non-linear correlations and are significantly more resistant to overfitting. This technique is however still in its infancy. Camacho et al. (2023) tested a GPRN one-dimensional analysis on solar data and found that it does yield sensible results, but when activity indicators are included in the fit the model is significantly worse at predicting the variability in the RVs. Moreover, it is computationally very expensive.

Overall, Gaussian processes in all their forms are a promising tool to account for stellar activity in radial-velocity data, and are quickly evolving to address the increased precision of new spectrographs and the renewed interest of the community in the elusive small signals of long-period planets.

2.4 Other Sources of Information

Often information in the obtained spectra is not enough to successfully constrain the effects of stellar variability. In these cases, we can turn to other sources of data. Several mitigation techniques in fact rely on independent observations. First of this list is the FF′ method (Aigrain et al. 2012), described in Section 2.3.4.2, which aims to constrain the RV variations due to active regions by modelling the photometric noise due to bright faculae and dark spots. However, in most cases, photometry is not as effective as radial velocities to characterise the star. Moreover, in order to model activity in the RVs the photometry needs to be near-simultaneous. The community is therefore looking to other independent sources of information. Spectropolarimetry is a promising avenue of study, as is addressed Section 5.2. Finally, a good understanding of not only stars but also the instrument is required, especially to model small long-term variations that could prevent the detection of Earth-analogues. Solar observations can then be invaluable datasets to study, both for their long baselines and their cadence, as is highlighted in Chapter 5, and as testing ground for instrumental systematics.

Chapter 3

An Analysis Toolkit: Gaussian Processes and MAGPy-RV

"Any sufficiently advanced technology is indistinguishable from magic."

— Arthur C. Clarke

As briefly covered in Chapter 2, Gaussian Processes (GPs) are one of the most used tools for the mitigation and modelling of stellar variability in astrophysical literature (e.g., Haywood et al. 2014; Faria et al. 2016; Serrano et al. 2018; Barros et al. 2020; Rescigno et al. 2023b). In this Chapter, I summarise the mathematical definition of GPs and introduce the analysis toolkit MAGPy-RV, a pipeline for GP regression with Markov Chain Monte Carlo parameter space searching algorithms for the detection of exoplanets and the modelling of stellar activity signals.

3.1 Mathematical Definition of Gaussian Processes

In statistical terms, a Gaussian process can be defined as an N-dimensional collection of random variables, such that the joint probability distribution drawn from it is a Gaussian distribution in N dimensions. In other words, a Gaussian process is a non-parametric stochastic model based on the Gaussian distribution (Rasmussen and Williams 2006; Pavliotis 2015). GPs are a generalisation of multivariate normal distribution theory, which is itself an extension to multiple dimensions of basic Gaussian distributions.

3.1.1 From Gaussian Distributions to Multivariate Normal Distribution Theory

A random variable is Gaussian (or normally) distributed if its probability density function takes the form of

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \tag{3.1}$$

in which σ^2 represents the variance of the distribution, and μ its mean. In statistical terms, it is usually presented as

$$p(x) \sim \mathcal{N}(\mu, \sigma^2), \tag{3.2}$$

in which the $\sim$ symbol means "is distributed as". Multiple independent normal distributions of correlated variables can be expressed as joint Gaussian distributions. A system defined by multiple "characteristics" $\vec{x} = [x_1, x_2, ..., x_n]$ can thus be summarised as a multivariate normal distribution fully described by its mean vector $\vec{\mu}$ and its covariance matrix **K** in the form

$$p(\vec{x}|\vec{\mu}, \mathbf{K}) \sim \mathcal{N}(\vec{\mu}, \mathbf{K}), \tag{3.3}$$

or more explicitly

$$p(\vec{x}|\vec{\mu}, \mathbf{K}) = \frac{1}{(2\pi)^{n/2} \cdot |\mathbf{K}|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \mathbf{K}^{-1} (\vec{x}-\vec{\mu})\right). \tag{3.4}$$

Multivariate normal distributions enjoy a variety of properties. In particular, four are of interest to this work.

- Normalisation property: the density function normalises to unity as

$$\int p(x|\vec{\mu}, \mathbf{K}) dx = 1. \tag{3.5}$$

- Marginalisation property: a Gaussian distribution always marginalises to a Gaussian distribution. Therefore, given two variables x_A and x_B, expressed by a multivariate normal distribution such that

$$\vec{\mu} = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix} \qquad \text{and} \qquad \mathbf{K} = \begin{bmatrix} \mathbf{K}_{AA} & \mathbf{K}_{AB} \\ \mathbf{K}_{BA} & \mathbf{K}_{BB} \end{bmatrix}, \tag{3.6}$$

then their marginals can be written in the form

$$p(x_A|\vec{\mu}, \mathbf{K}) = \int_{x_B} p(x_A, x_B|\vec{\mu}, \mathbf{K}) dx_B \sim \mathcal{N}(x_A|\mu_A, \mathbf{K}_{AA}) \tag{3.7}$$

and

$$p(x_B|\vec{\mu}, \mathbf{K}) = \int_{x_A} p(x_B, x_A|\vec{\mu}, \mathbf{K}) dx_A = \mathcal{N}(x_B|\mu_B, \mathbf{K}_{BB}). \tag{3.8}$$

- Conditioning property: the conditional densities of multivariate normal distributions are also Gaussian, expressed as

$$p(x_A|x_B) = \frac{p(x_A, x_B|\vec{\mu}, \mathbf{K})}{\int_{x_B} p(x_A, x_B|\vec{\mu}, \mathbf{K}) dx_A} \sim \mathcal{N}(\mu_A + \mathbf{K}_{AB}\mathbf{K}_{BB}^{-1}(x_B - \mu_B), \mathbf{K}_{AA} - \mathbf{K}_{BA}\mathbf{K}_{BB}^{-1}\mathbf{K}_{AB}). \tag{3.9}$$

- Summation property: given two independent random variables z and z' with the same dimensionality and both described by normal distributions, when summed their resulting distribution is also Gaussian, as

$$\begin{gathered} \text{for} \qquad z \sim \mathcal{N}(\vec{\mu}, \mathbf{K}) \qquad \text{and} \qquad z' \sim \mathcal{N}(\vec{\mu}', \mathbf{K}') \\ z + z' \sim \mathcal{N}(\vec{\mu} + \vec{\mu}', \mathbf{K} + \mathbf{K}'). \end{gathered} \tag{3.10}$$

3.1.2 Gaussian Processes

As mentioned, Gaussian processes are a generalisation of multivariate Gaussian distributions to infinite dimensions. A multivariate normal distribution is dependent on its finite NxN covariance matrix and mean vector. The number N of rows and columns of the matrix is defined by the number of variables that the joint distribution includes. A GP extends this dimensionality to infinity and is thus instead fully described by covariance and mean functions, as

$$f \sim GP(\mu(x), k(x_i, x_j)), \tag{3.11}$$

where $\mu(x)$ is the mean function of the GP model (often taken to be equal to zero) and $k(x_i, x_j)$ is its covariance function. The latter is often chosen to have parametric form and is written as a kernel function in the form $k(x_i, x_j, \theta)$, in which θ is the list of the hyperparameters required to define the covariance. We define the two functions as

$$\begin{aligned} \mu(x) &= \mathbb{E}[f(x)] \\ k(x_i, x_j) &= \mathbb{E}[(f(x_i) - \mu(x_i))(f(x_j) - \mu(x_j))], \end{aligned} \tag{3.12}$$

so that the mean model is only dependent on the set of variables x, while the covariance depends on both the mean function and the relationship between each pair of variables x_i and x_j. Here $\mathbb{E}$ can be read "as expected value of".

The probability density $p(f|x)$, in the case of parametric kernels also written as $p(f|x, \theta)$, can be extrapolated from the same definition as before, in the multivariate case, to be in the form of

$$p(f|x,\theta) \sim \mathcal{N}(f|\mu, k) = (2\pi)^{-n/2}|\mathbf{K}_{xx}|^{-1/2} \exp\left(-\frac{1}{2}(\vec{f} - \vec{\mu})^T \mathbf{K}_{xx}^{-1}(\vec{f} - \vec{\mu})\right), \tag{3.13}$$

in which $\mathbf{K}_{xx}$ is the covariance matrix built by evaluating the kernel function on a finite set of inputs $\vec{x}$, and $\vec{f}$ and $\vec{\mu}$ are the GP function and the mean model respectively evaluated at those same x values. $f(x)$ represents the GP function that fits the behaviour of the data. Therefore in most cases, $\vec{f}$ is actually a vector of the training data values themselves. Equation 3.13 is often called the likelihood, $\mathcal{L}$, of the derived Gaussian process model to fit the data. The first term of the equation is a normalisation factor, dependent on the number of known variables n, the size of the vector $\vec{x}$. The second term, the inverse of the square root of the determinant of the covariance matrix evaluated on the training set $\vec{x}$, is often considered an Occam Razor term, which penalises complex covariance models against simpler ones. Most often this formula is presented in its logarithmic form

$$\ln(\mathcal{L}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{K}_{xx}|) - \frac{1}{2}(\vec{f} - \vec{\mu})^T \mathbf{K}_{xx}^{-1}(\vec{f} - \vec{\mu}). \tag{3.14}$$

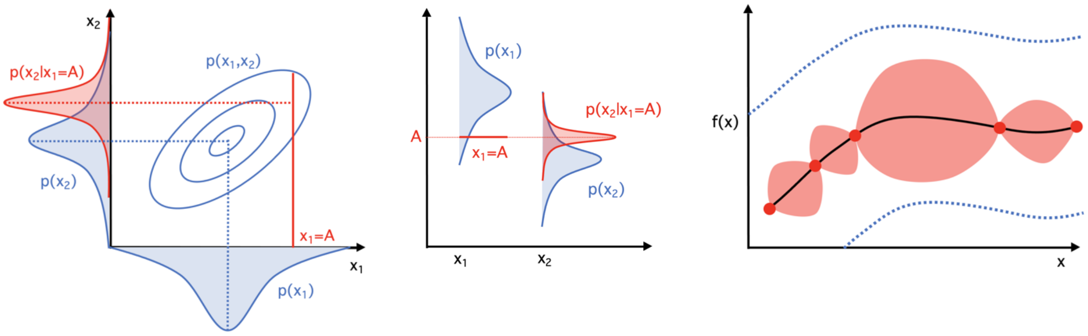

Figure 3.1: Diagram visually describing the conceptual basis of Gaussian processes.
Left panel: representation of a two-dimensional multivariate Gaussian distribution. The joint distribution of the two variables x_1 and x_2 is depicted as elliptic contours (covariance ellipse) in blue, and the marginal distributions $p(x_1)$ and $p(x_2)$ that form it are also included in blue on their respective axes. Observing x_1 at the value A indicated by the red vertical line allows for a better determination of the possible x_2 values based on the pre-existing knowledge of their covariance. The conditional distribution $p(x_2|x_1 = A)$ can therefore be extracted and is here plotted in red.
Middle panel: alternative representation of the previous panel, presented here in a "time series" form. The variables x_1 and x_2 are plotted on the same axis x, and their possible values are indicated on the vertical axis. The marginal distributions $p(x_1)$ and $p(x_2)$ are still plotted in blue, and the conditional distribution $p(x_2|x_1 = A)$ is also included in red.
Right panel: the diagram of the middle panel is then extended for an arbitrarily dense array of x values, giving rise to a continuous marginal distribution here symbolised by the space between the two dashed blue lines. By measuring $f(x)$ at multiple x values highlighted as red dots, a continuous posterior distribution can be derived, plotted in red, similarly to the previous panels. A Gaussian process computes the posterior conditional distribution given the observed data x and extracts all the possible functions that can fit it. The most probable function $f(x)$ is plotted as a black line.

3.1.2.1 Intuitive Explanation

In order to exhaustively explain Gaussian processes, often a more intuitive route is preferred to the strictly mathematical one. In this Section, I follow the example of Roberts et al. (2013) and describe GPs in more practical terms. We begin by considering a two-dimensional Gaussian distribution of two random variables x_1 and x_2. The multivariate normal distribution $p(x_1, x_2)$ of these two variables is described as in the previous Sections by a mean vector, which for simplicity will be zero for both variables, and a 2x2 covariance matrix. We can visualise this problem in the leftmost panel of Fig. 3.1 as blue elliptic contours for a 2D distribution, with the corresponding marginal one-dimensional distributions, $p(x_1)$ and $p(x_2)$, also plotted on the respective axes in blue.

Let's now consider the effect of observing variable x_1, and assigning it the value $x_1 = A$. By drawing the red vertical line at value A in the left panel of Fig. 3.1, we can refine our prediction of the most probable value of x_2 given the pre-existing information we have about their covariance relationship (the shape and orientation of the contours of the joint distribution). We can therefore extract the resulting conditional distribution

$p(x_2|x_1 = A)$ in red. This conditional distribution is different from its original marginal probability distribution in blue, and has shrunken uncertainties. The middle panel of Fig. 3.1 is an alternative way of visualising the same problem using the horizontal axis to list the variables x and the vertical axis to represent their possible values. The conditional distribution for x_2 from the knowledge of x_1 and their relationship is still plotted in red. Finally, we can extend this same concept to as many variables as we desire, to the limit in which x_i is infinitely dense, as shown in the last panel of Fig. 3.1. Starting from a marginal distribution here symbolised by the space between the blue dashed lines, we can extract the conditional distributions for all x values between each of the measured data represented by the red dots. The posterior of the GP is here depicted in red, and the most probable function $f(x)$ is plotted as a black line. The joint distribution over all these infinite variables becomes itself of infinite-dimensions and therefore can be described by a function space, a Gaussian process. Gaussian processes are thus a distribution over functions.

3.1.3 Covariance Functions

A Gaussian process requires only two parameters, the mean and the kernel functions, although the covariance is in most cases further described by hyperparameters. The role of inference in GP models is to refine the wide and uninformed distribution of possible functions into a subset of curves that best fit the available training data. The form of those curves is uniquely described by the covariance function. Covariance functions map pairs of variables into real values. They are positive, semi-definite functions over all the space defined by each pair of variables. They have to satisfy three main criteria:

- They must be symmetric

$$k(x_i, x_j) = k(x_j, x_i). \tag{3.15}$$

- They must be bilinear

$$k(ax_i, bx_j) = ab \cdot k(x_i, x_j). \tag{3.16}$$

- They must be positive for all values of x larger than zero, and be equal to zero for

$x = 0$, mathematically written as

$$\begin{aligned} k(x, x) &\geq 0 \qquad \forall \qquad x \\ k(0, 0) &= 0. \end{aligned} \tag{3.17}$$

In practice, covariance functions are evaluated on a finite set of datapoints over which the GP model is then interpolated. These kernel matrices all share the same property: all their eigenvalues $\vec{v}$ are positive, expressed mathematically as

$$\vec{v}^T \mathbf{K} \vec{v} \geq 0 \qquad \forall \qquad \vec{v}. \tag{3.18}$$

The covariance function of the Gaussian process can be interpreted as the measure of similarity or correlation between each point in time (Foreman-Mackey et al. 2017). However, in reality with precision limits, we cannot perfectly describe the covariance between observations with a function, we can only approach the best description. White noise must therefore be considered in all computations. Most often the derived kernel matrix is therefore not only expressed by the covariance function, but also depends on a further parameter, here β, which describes the uncertainty of the observations. White noise can then be mapped by the diagonal elements of the matrix, while red, correlated noise is represented by all non-diagonal elements in the form

$$\mathbf{K}_{x_i x_j} = \beta^2 \delta_{x_i, x_j} + k_\theta(x_i, x_j). \tag{3.19}$$

There are many covariance functions currently employed in GP analyses. In this work, I only list those that have shown to be useful in the context of stellar activity modelling.

- The Squared Exponential Kernel is one of the standard covariance functions. It is also known as the Radial Basis Function Kernel or the Gaussian Kernel. It has the form

$$k(x_i, x_j) = \sigma^2 \exp\left(-\frac{|x_i - x_j|^2}{2l^2}\right), \tag{3.20}$$

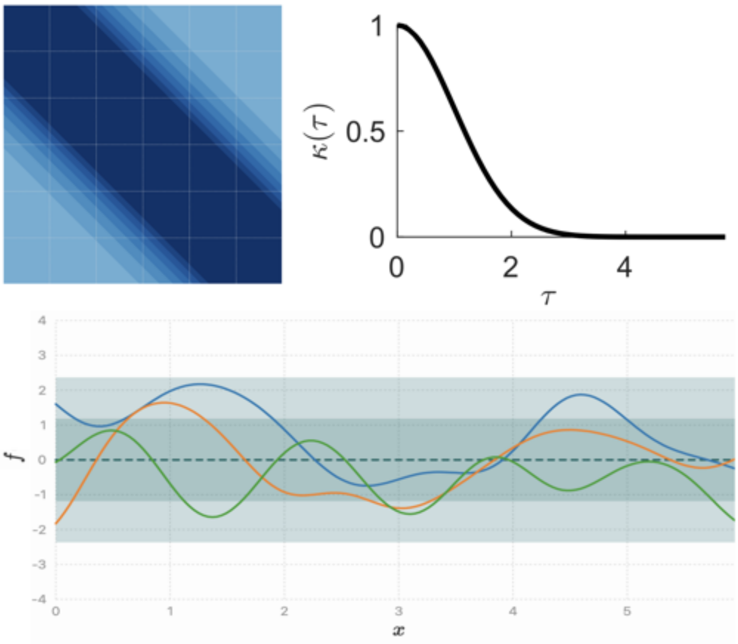

Figure 3.2: Collection of example visualisations of the squared exponential covariance function. In order, the covariance matrix, functional form with respect to $\tau = |x_i - x_j|$, and three random samples of GP functions described by the kernel. Figure adapted from Nayek et al. (2019), with supplementary material derived from the interactive GP visualisation found at: github.com/st--/interactive-gp-visualization/.

in which σ^2 is the variance of the GP function, and l is its lengthscale. An example of the GP covariance matrix, its functional form with respect to $\tau = |x_i - x_j|$, and a selection of three random GP function samples are shown in Fig. 3.2.

- The Exponential Sine Squared Kernel is the most common periodic covariance function, and is often referred to simply as the Periodic Kernel. It is defined as

$$k(x_i, x_j) = \sigma^2 \exp\left[-\frac{2}{l^2}\sin^2\left(\frac{\pi|x_i - x_j|}{P}\right)\right], \tag{3.21}$$

  in which σ^2 is the variance, l is the timescale of the periodicity and P is the period of the oscillation. Fig. 3.3 shows example visualisations for this covariance function.

- The Matérn kernel is a family of stationary and isotropic covariance functions. They are a generalisation of the squared exponential formulation. They include an extra ν term and are usually written in the form

$$k(x_i, x_j) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|x_i - x_j|}{l}\right)^{\nu} \mathcal{K}_\nu\left(\frac{\sqrt{2\nu}|x_i - x_j|}{l}\right) \tag{3.22}$$

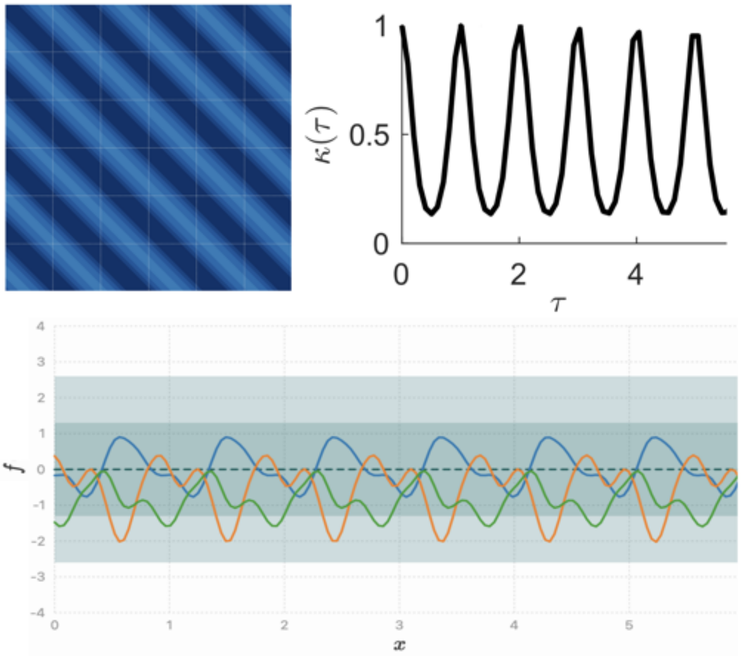

Figure 3.3: Collection of example visualisations of the exponential sine squared covariance function. In order, the covariance matrix, functional form with respect to $\tau = |x_i - x_j|$, and three random samples of GP functions described by the kernel. Figure adapted from Nayek et al. (2019), with supplementary material derived from the interactive GP visualisation found at: github.com/st--/interactive-gp-visualization/.

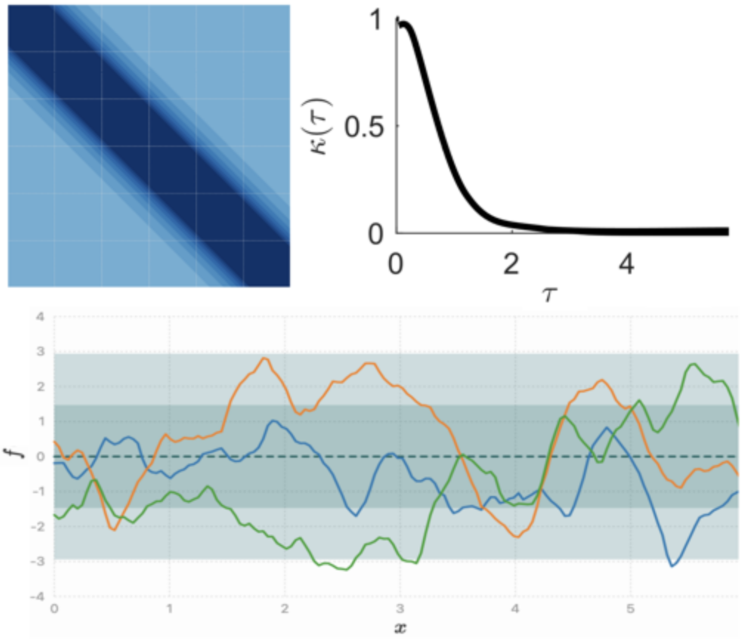

Figure 3.4: Collection of example visualisations of the Matérn 3/2 covariance function. In order, the covariance matrix, functional form with respect to $\tau = |x_i - x_j|$, and three random samples of GP functions described by the kernel. Figure adapted from Nayek et al. (2019), with supplementary material derived from the interactive GP visualisation found at: github.com/st--/interactive-gp-visualization/.

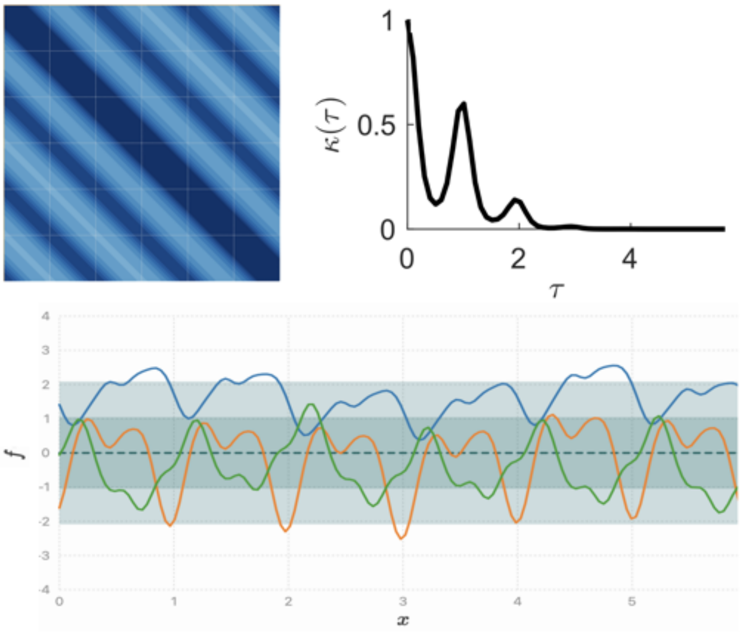

Figure 3.5: Collection of example visualisations of the quasi-periodic covariance function. In order, the covariance matrix, functional form with respect to $\tau = |x_i - x_j|$, and three random samples of GP functions described by the kernel. Figure adapted from Nayek et al. (2019), with supplementary material derived from the interactive GP visualisation found at: github.com/st--/interactive-gp-visualization/.

in which σ^2 is the variance, $\Gamma(\nu)$ is the Gamma function, l is the lengthscale, and $\mathcal{K}_\nu$ is a modified Bessel function. ν can take any positive value. For $\nu = 0$ the kernel is reduced to the Ornstein-Uhlenbeck kernel, and for $\nu \longrightarrow \infty$ the covariance function recovers the squared exponential kernel. The ν parameter controls the smoothness of the function: smaller ν values yield more complex functions, while larger ones produce smoother curves. Matérn kernels with half-integers ν values ($\nu = i + 1/2$ for any $i \in \mathbb{N}_+$) have simpler forms and can be expressed as the product of an exponential and a polynomial function of order i. In most cases ν is taken to be equal to $1/2$, $3/2$ or $5/2$. As an example, the Matérn 3/2 covariance function is defined as

$$k(x_i, x_j) = \left(1 + \frac{\sqrt{3}|x_i - x_j|}{l}\right) \exp\left(-\frac{\sqrt{3}|x_i - x_j|}{l}\right). \tag{3.23}$$

Fig. 3.4 shows an example covariance matrix and random samples of functions for this last kernel formulation.

- The Quasi-Periodic (QP) Kernel combines a square exponential term with periodic oscillations as

$$k(x_i, x_j) = \sigma^2 \exp\left[-\frac{|x_i - x_j|^2}{l^2} - \frac{\sin^2\left(\frac{\pi|x_i - x_j|}{P}\right)}{h^2}\right], \tag{3.24}$$

in which as before σ^2 is the variance, l is the timescale of the periodicity, and P is the period of the oscillation. The fourth hyperparameter h is the harmonic complexity of the fit, and determines the "inner-period smoothness" of the output function, its amount of high-frequency structure. Fig. 3.5 shows example visualisations for this covariance function. Given its periodic damped behaviour, the QP kernel is the covariance function most used to model stellar variability. Its periodic modulation mimics the periodic signal due to active regions rotating in and out of view on the stellar disc, and the exponential damping term accounts for the change in the signal as the active regions dissipate into the quiet star. All hyperparameters can therefore be directly connected to physical properties of the star: P represents the stellar rotation period, l the average lifetime of spots and faculae and h can be connected to the distribution of active regions on the stellar disc (Nicholson and Aigrain 2022).

Building on the properties of covariance functions, new and more complex kernels can always be developed by simple addition or multiplication of standard kernel functions. This can be especially useful when one is aiming to simultaneously model signals generated by multiple processes. As an example grounded in the field of stellar activity mitigation, the combined variability of active regions and of the effects of the stellar magnetic cycle can be modelled by summing a QP kernel to describe the first, and a long-period periodic kernel for the second.

3.1.4 Predictive Distribution

With an assigned mean function and a fully descriptive covariance matrix (with all hyperparameter values assigned), the Gaussian process is fully defined, meaning we can derive the best-fit function. In practice this is done by evaluating the GP posterior distribution $p(\vec{f}_\star)$ at a series of test datapoints, $x_\star$, dense enough for interpolation. We start by considering

$$p\left(\begin{bmatrix}\vec{f}\\ \vec{f}_\star\end{bmatrix}\right) \sim \mathcal{N}\left(\begin{bmatrix}\vec{\mu}(x)\\ \vec{\mu}_{x_\star}\end{bmatrix}\right), \left(\begin{bmatrix}\mathbf{K}_{xx} & \mathbf{K}_{xx_\star}\\ \mathbf{K}_{x_\star x} & \mathbf{K}_{x_\star x_\star}\end{bmatrix}\right). \tag{3.25}$$

With some manipulation to obtain the mean $\vec{m}_\star$ and variance $\mathbf{V}_\star$ of the set of datapoints values $\vec{f}_\star$, we can define the GP posterior distribution as

$$p(\vec{f}_\star) \sim \mathcal{N}(\vec{m}_\star, \mathbf{V}_\star), \tag{3.26}$$

in which the mean vector is

$$\vec{m}_{x_\star} = \vec{\mu}_{x_\star} + \mathbf{K}_{x_\star x}\mathbf{K}_{xx}^{-1}(\vec{f}(x) - \vec{\mu}(x)), \tag{3.27}$$

and

$$\mathbf{V}_\star = \mathbf{K}_{x_\star x_\star} - \mathbf{K}_{x_\star x}\mathbf{K}_{xx}^{-1}\mathbf{K}_{x_\star x}^T \tag{3.28}$$

represents the matrix of the covariance. The mean values $\vec{m}_{x_\star}$ of the predicted posterior distribution are often referred to as the predictive data and used via interpolation to find the most probable GP function to the training data x. The uncertainties on each data prediction $\vec{m}_{x_\star}$, and by extrapolation of the predicted function, can then be computed as the square root of the diagonal of the covariance matrix $\mathbf{V}_\star$.

Finally, let us then summarise all the previous Sections. GPs are a tool for isolating a probability distribution of functions that describe the available observations. They require the definition of mean and covariance functions. By assigning values to the kernel hyperparameters, an NxN covariance matrix can be computed, with N equal to the number of training data fed to the GP. Equations 3.27 and 3.28 can then be used to derive the most probable function and its uncertainty by defining a subset of epochs $\vec{x}_\star$ over which the GP can be interpolated. Finally, by applying Equation 3.13, the likelihood of the derived function can be calculated in order to assess how well it fits the observations. But how do we find the hyperparameter values to plug in the covariance function so that we extract a GP model that maximises the likelihood? We perform a Gaussian process regression.

3.1.5 Gaussian Process Regression

In machine learning a regression aims to learn a function based on information derived from some number of evaluations of said function. In the context of probabilistic studies, a regression has the task of learning a probability distribution over all possible functions. It is an efficient way of quantifying the uncertainty of our knowledge. In the case of Gaussian processes, we define a GP regression as the search for the kernel hyperparameter values which yield the GP function that best describes the available observations.

3.1.5.1 Priors

In order to perform inference we need to marginalise over hyperparamters. This can be done by assigning a prior $p(\theta)$ to each hyperparameter θ, in order to inform the covariance domain. In Bayesian theory, a prior is the probability distribution that best expresses the previous knowledge about the considered quantity. In the absence of pre-existing knowledge, these priors can be non-descriptive. From a prior, the probability of θ being equal to a value A can be computed as $p(\theta = A)$. This calculated probability can then be multiplied by the likelihood $\mathcal{L}$ of the GP in order to more precisely assess the quality of the model fit based on prior knowledge of the hyperparameter in question. In other words, we can marginalise over the known parameter θ such that

$$p(f, x) = \int p(f|x, \theta)p(\theta)d\theta. \tag{3.29}$$

There are a variety of priors, each described by their characteristic probability distribution and the related parameters. In this work I only list the most common priors, useful in the context of stellar variability mitigation.

- A uniform prior is the simplest prior formulation. Its probability distribution $p(\theta)$ looks like a rectangular top hat, in which all points in a finite interval $[a, b]$ are equally likely, in the form

$$p(\theta) = \begin{cases} \frac{1}{b-a} & \forall \qquad \theta \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

For all values outside of the considered interval, the logarithmic probability approaches negative infinity, therefore applying an insurmountable penalty term to the model.

- The Gaussian prior, also referred to as the normal prior or bell prior, can be written in the form

  $$p(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta - \mu}{\sigma}\right)^2\right], \tag{3.30}$$

  in which μ is the mean of the Gaussian distribution, and σ is its standard deviation. A Gaussian prior is usually applied to cases in which some approximate knowledge is gained of the hyperparameter θ. μ can then be set as the hyperparameter expected value, and σ its uncertainty level.

- The Jeffreys prior is a non-informative prior distribution. Its probability function is proportional to the square root of the determinant of the Fisher information matrix. Similarly to the uniform prior, its probability density is bound between minimum and maximum values a and b. It is usually expressed in the form

  $$p(\theta) = \frac{1}{\ln\left(\frac{b}{a}\right)} \cdot \frac{1}{\theta}. \tag{3.31}$$

  The Jeffreys prior is an improper prior, as it cannot always be normalised. It is useful when we are ignorant about the scale of the value of θ, and thus require a prior which has the same form no matter the order of magnitude.

- The modified Jeffreys prior is a modified version of the previous prior. It extends the Jeffreys formulation for the limit case of the minimum allowed value $a \longrightarrow 0$, for which the probability function in a Jeffreys prior goes to 1. Beside a minimum a and a maximum b it is also characterised by a knee value $c \in [a, b]$, as

  $$p(\theta) = \frac{1}{\ln\left(\frac{b-c}{a-c}\right)} \cdot \frac{1}{\theta - c}. \tag{3.32}$$

  The modified Jeffreys prior essentially behaves like a uniform prior for $a < \theta < c$, and like a traditional Jeffreys prior for $c < \theta < b$.

3.1.5.2 Markov Chain Monte Carlo

We have now defined all elements required for the computation of the GP posterior distribution $p(f, x)$, but its derivation is not straightforward. We now need a method that can approximate the integration over posterior distributions. Markov Chain Monte Carlo algorithms are one group of such methods. There are many techniques to perform regression and find the hyperparameters that yield the function that best describes the data. While direct optimisation methods may be applied, probabilistic inference methods are generally preferred. Markov Chain Monte Carlo algorithms are a combination of two techniques: Monte Carlo sampling, and Markov Chain processes. In general, Monte Carlo sampling is the typical solution for probabilistic inference. Monte Carlo methods are a class of computational algorithms. In basic terms, they consist of drawing independent samples from the probability distribution, and repeating this process many times to derive the best-fit value, using randomness to approximate solving deterministic principles which cannot be solved analytically. Monte Carlo methods are widely employed in all areas of science to simulate complex systems (Gregory 2010). They are named after the Monte Carlo Casino in Monaco. Basic Monte Carlo sampling, however, presents some drawbacks. Firstly, Monte Carlo methods assume that each random sample is fully independent from the others, and therefore can be independently drawn (Bishop and Nasrabadi 2007), but this is rarely the case. Secondly, they do not perform well in high-dimensions, since the volume of the sample space to explore increases exponentially with the number of parameters.

In order to address these limitations, Markov Chain Monte Carlo (MCMC) techniques were developed to perform well in high-dimensional cases (Murphy 2013; Gregory 2010). They were first created around the same time as computers and were originally employed in particle physics calculations required for the development of the atomic bomb. A Markov Chain is a type of stochastic process that aims to characterise a series of values (Robert and Casella 2005). It is a systematic method for generating a sequence of random sample variables, in which each new parameter value is dependent on the previous one. Combining the two methods, MCMC algorithms allow for random sampling of high-dimensional probability distributions while taking into consideration the dependence be-

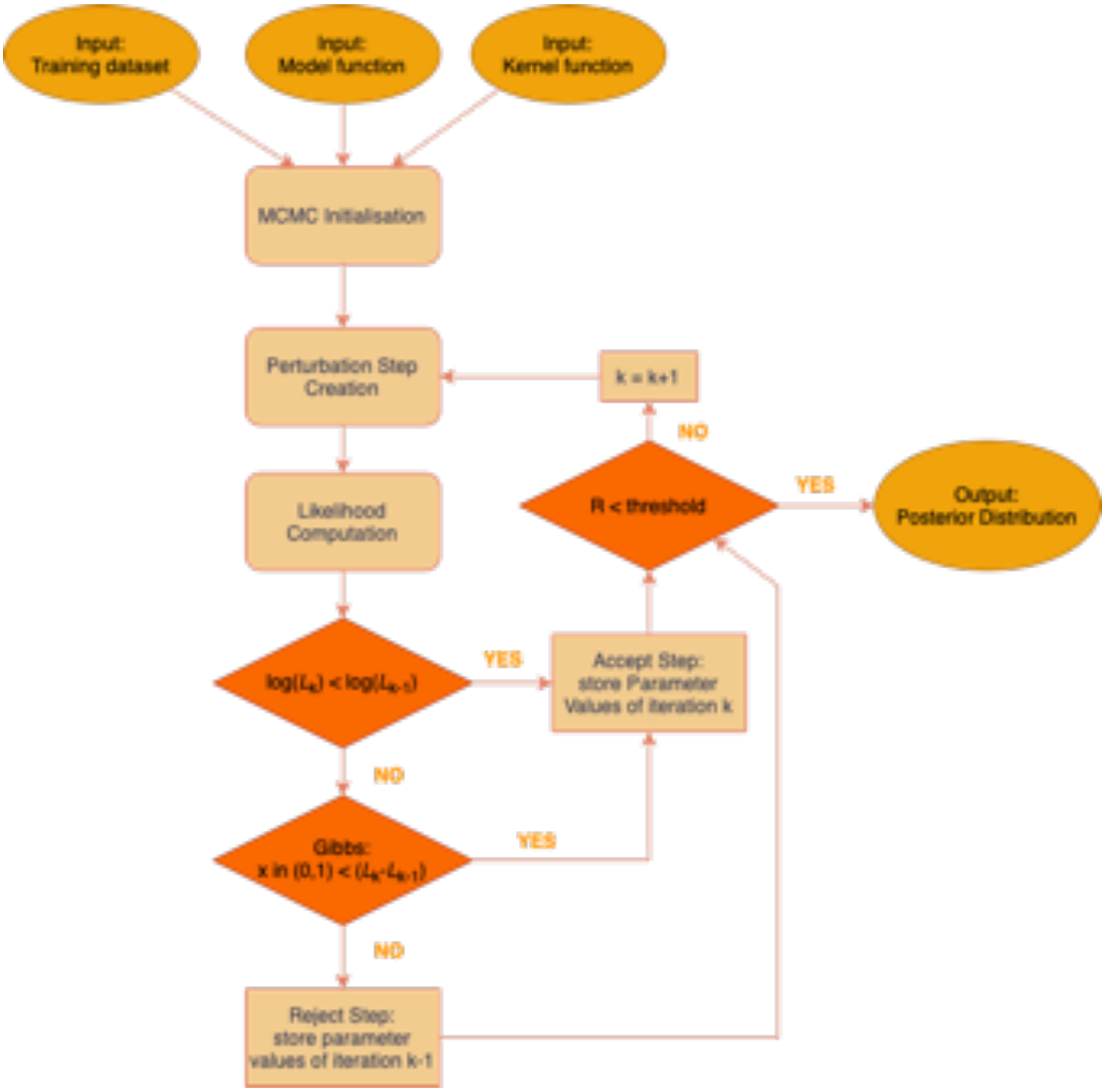

Figure 3.6: Flowchart representing the basic workflow of MCMC techniques. Inputs and outputs are represented by ovals. After MCMC initialisation, the algorithm goes through 5 basic steps which are iterated until convergence is reached: perturbation step creation, likelihood computation, likelihood comparison, step acceptance/rejection (with appropriate parameter storing), and convergence assessment (for more information on this final step, see Section 3.1.5.2.1). In the diagram the current iteration is defined as iteration k, and its logarithmic likelihood as $\log(L_k)$.

tween samples. They are designed to provide sampling approximations for the posteriors of probabilistic problems. There are many adaptations of Markov Chain Monte Carlo routines. In the context of astrophysics, the most commonly implemented algorithms are based on a specialised version of the Metropolis-Hastings sampler, called Gibbs sampling (Kalos 1986). This algorithm helps the MCMC to avoid settling for local logarithmic likelihood maxima, and to successfully sample the entire parameter space. An example of an MCMC workflow is summarised in Fig. 3.6. On a basic level, any MCMC routine requires the same four steps: perturbation step creation, function probability computation, value comparison, and step acceptance or rejection. For each iteration step, the MCMC extracts random values to assign to the considered parameters. It then calculates the logarithmic likelihood of the model computed with the extracted parameters. It compares this current value with the one of the previous step, calculated for the last accepted iteration with the respective random set of parameters. Finally, based on a criterion defined by the Metropolis-Hastings formulation, it either stores or rejects the set of parameter values sampled. In most cases this acceptance/rejection criterion is the following: if the current logarithmic likelihood is larger than the previous, the iteration is accepted and the sampled values are stored. If the logarithmic likelihood is smaller, it does not directly result in rejection. Instead a number between 0 and 1 is randomly selected. If the ratio between the current likelihood and the one from the previous step is larger than this random number the iteration is accepted, if it is smaller the iteration is rejected and the set of parameter values of the previous iteration is stored instead. This means that in some cases, a function that fits the training dataset worse than the one of the previous step can be preferred, allowing the algorithm to "escape" a local likelihood maxima in order to find the global one. An effective MCMC chain ideally accepts 1/4 of all iterations (Ford 2006): if too few steps are accepted, their sampling will not be representative of the target distribution; if too many steps are accepted, the chain is performing fully random walks and is not able to identify the target maxima.

3.1.5.2.1 Parallel Sampling Every independent run of an MCMC is called a chain. With sufficiently large numbers of iterations a single chain can fully explore the entire parameter space. However, in high-dimensional cases the number of required steps can be prohibitively large. In order to increase computational efficiency, MCMC algorithms often perform multiple independent runs simultaneously, and use multiple chains to more efficiently sample the space and build posteriors. Moreover, with a single run, it is tricky to establish whether the chain has reached stationarity and therefore the global likelihood maximum has been found. Running multiple chains can help diagnose multimodality, and allows for the computation of convergence diagnostics. One of the most common ways of monitoring and assessing whether the MCMC has converged is the Gelman-Rubin statistic (Gelman and Rubin 1992). This convergence test relies on two estimations of the variance of parallel chains. The intra-chain (within the same chain) variance σ_m^2 of each chain m can be estimated as

$$\sigma_m^2 = \frac{1}{N-1}\sum_{i=0}^{N}(\theta_i^m - \hat{\theta}_m)^2, \tag{3.33}$$

in which N is the current number of steps, θ_i^m represents the list of all parameter values explored by the chain m, and $\hat{\theta}_m$ is the posterior mean of the same chain. The inter-chain variance σ_M^2 can be computed as the variance between the means of all chains in the form

$$\sigma_M^2 = \frac{N}{M-1}\sum_{m=0}^{M}(\hat{\theta}_m - \hat{\theta})^2, \tag{3.34}$$

in which M is the total number of chains, and $\hat{\theta}$ is the mean of all chains, calculated as

$$\hat{\theta} = \frac{1}{M}\sum_{m=0}^{M}\hat{\theta}_m. \tag{3.35}$$

The average variance parameter value over all chains can be computed as

$$W = \frac{1}{M}\sum_{m=0}^{M}\sigma_m^2. \tag{3.36}$$

Under certain stationarity conditions, the pooled variance estimated as

$$V = \frac{N-1}{N}W + \frac{M+1}{NM}\sigma_M^2 \tag{3.37}$$

is an unbiased test of the true variance of the MCMC. With fully converged chains, so is W. We can therefore assess the convergence by comparing V and W as

$$R = \sqrt{\frac{V}{W}}. \tag{3.38}$$

As R decreases to approach 1, the chains are approaching full convergence. Different works use different minimum R thresholds to define acceptable convergence level, generally in the range of $R = 1.1$ and $R = 1.01$.

3.1.5.2.2 Affine Invariance The speed and efficiency at which the MCMC reaches convergence is dependent on the way the parameter space is explored. The "stepping" algorithm with which the chains randomly select their new values is therefore important. The simplest method to generate the new chain position is to randomly extract a parameter value from its prior distribution. An affine invariant sampler is a more efficient stepping algorithm. Affine invariant methods automatically work in the best set of variables that can be achieved via linear transformation. This specific stepping formalism is commonly referred to as the "stretch move", and can be summarised in the form

$$X_{k,i} \longrightarrow X_{k,i+1} = X_{j,i} + z \cdot (X_{k,i} - X_{j,i}). \tag{3.39}$$

In less mathematical terms, the new parameter $X_{k,i+1}$ of chain k at step $i+1$ depends on its parameter value in the previous iteration $X_{k,i}$ and the value of the same parameter $X_{j,i}$ of a different randomly selected chain j. z is a multiplication constant shared between all parameters that is randomly extracted between -1 and +1 at every new iteration. With this sampling algorithm, the size of the step changes at every iteration depending on the separation in parameter space between chains. When chains are diverging in order to more widely explore the space, the steps at each iteration will generally be larger. When chains are converging and a maximum in the logarithmic likelihood is identified, the steps will be smaller. This algorithm allows for dynamical self-tuning of the step function, making it more efficient in the context of computational expense, and less in need of human supervision.

3.1.5.2.3 Burn-in Phase The MCMC is generally initialised with a set of "starting" positions for the analysed parameters. These initial "guesses" are ideally based on previous knowledge about the parameters that are being investigated, but that cannot always be ensured. Assuming enough steps are taken and convergence has been reached, these starting points will not significantly affect the final posterior distribution, but they can partially bias the results. To compensate for this, a starting burn-in phase is usually discarded. This initial set of iterations is selected to be large enough that the chains have reached a more stable status and are properly exploring the posterior of interest. On average, this phase is considered to be between the first 10 to 25% of the total number of iterations, but the number of discarded steps will change depending on the random behaviour of each run. A common way to assess whether the chains have travelled from the initial "guess" to their stationary regime is to plot their mixing, or their parameter value at each MCMC iteration. This burn-in rejection can be done because Markov Chains have no memory outside their final distributions. Their future positions are only dependent on where they are at the present step (and where other chains are, in the case of affine invariant algorithms). This means that discarding the first 10 steps is equivalent to simply starting from the 10^{th} iteration. In summary, burn-in phases allow the MCMC analysis to "start" from the ideal parameter values instead of relying on initial "guesses". Burn-ins are not useful in cases in which the MCMC is already initialised at parameter positions within the high-probability region of the posterior.

At the end of the algorithm, after convergence has been reached, the MCMC outputs the list of the parameter values of all accepted steps. These output distributions for each analysed parameter represent the posteriors of the GP regression. By finding the mean value and the dispersion of each marginalised distribution, the hyperparameters values that best fit the training data and their uncertainties can be identified. Thus, the best-fit GP function can be computed.

3.2 MAGPY-RV

Modelling Activity with Gaussian processes in Python in the Radial-Velocity regime (MAGPY-RV) is a Python-based analysis pipeline for the modelling of time series with Gaussian process regression and affine invariant Markov Chain Monte Carlo parameter space searching algorithm. I first employed this pipeline in Rescigno et al. (2023b). A basic version of MAGPy-RV has since been made public[1] and has been registered at the Astrophysical Source Code Library (ASCL) and at NASA's Exoplanet Modelling and Analysis Centre (EMAC) (Rescigno et al. 2023a). I developed this code for planetary detection, but MAGPy-RV is flexible and can be adapted to any GP regression need. The code requires no complex dependencies (excluding the basic Python libraries) and it is fully self-contained. MAGPy-RV focuses on ease of use and personalisation, with adaptable model and kernel functions. A detailed API with step-by-step tutorials can be found at https://magpy-rv.readthedocs.io/en/latest/. An extended version of the code also includes parallelised Markov Chain handling and simultaneous transit photometry. The various versions of the pipeline have been employed in a variety of works. For example, in Rescigno et al. (2023b), I use MAGPy-RV to model spectral and photometric data to detect and characterise a two-planet system (as covered in Chapter 4 Section 4.1). In Dalal et al. (2024), I use the pipeline to detect a three-planet system (as addressed in Chapter 4 Section 4.2). In Rescigno et al. (2024), I employ MAGPy-RV in the study of the mean longitudinal magnetic field (as explained in Chapter 5 Section 5.2).

My private version of MAGPy-RV includes multiple-core parallelisation as a performance enhancer. Although the Gaussian process workflow is not straightforward to parallelise, the MCMC algorithm can be sped up by performing part of the independent chain calculations simultaneously. Instead of computing the results from each chain in sequence, they can instead be sectioned in different smaller groups and the same calculations can be run at the same time on different computer cores. The chains cannot run fully independently, as the affine invariant formulation requires information from all the chains at the previous MCMC step, and convergence monitoring needs to look at the entire ensemble of iterations and chains. The pipeline workflow is then split between multiple cores as follows: firstly the MCMC is initialised on a single core. The perturbation step creation, the probability computation, value comparison and acceptance and rejection steps are performed simultaneously over the number of cores assigned. Before a new iteration is started, all information from all the chains is merged back on a single core, the new step

---

[1] Code available at: https://github.com/frescigno/magpy_rv

is initialised, and the chain calculations are once again split. This formalism significantly reduces the time taken for the computation as a function of the number of chains selected and the number of cores assigned.

3.2.1 Simultaneous Signal Modelling

MAGPy-RV includes simultaneous signal modelling. It essentially allows to contemporaneously find the best-fit solution of both the parameters used to define the mean model and the hyperparameters that describe the covariance of the GP function. In planet hunting when modelling stellar observations, the variability induced by stellar processes on the surface of the star is usually accounted for by the Gaussian process. The chosen mean function then traces the relative RV motion induced by the gravitational pull of a planet on a star, while the Gaussian process is used to mitigate stellar variability. In the GP formalism this model is referred to as μ, and can be as complex or as simple as necessary. In the following Sections, I describe the most relevant mean functions implemented in MAGPy-RV.

3.2.1.1 Zero Mean

The simplest model is the zero mean function or

$$\vec{\mu}(x) = 0 \qquad \forall \qquad x. \tag{3.40}$$

It essentially assumes that the considered data can be fully described by the Gaussian process.

3.2.1.2 Constant Offset

This mean function includes an offset described by a single constant in the form

$$\vec{\mu}(x) = a \qquad \forall \qquad x. \tag{3.41}$$

It is most useful when taking into consideration two separately-derived datasets X_A and X_B which may not share a common mean point. In this case the constant offset is only applied to a subsection of the data as

$$\vec{\mu}(x) = \begin{cases} a \quad \forall \qquad x \in X_A \\ \\ b \quad \forall \qquad x \in X_B \end{cases}$$

3.2.1.3 Radial-Velocity Keplerian

The RV Keplerian model aims to describe the signal imparted by the orbit of a planet on the radial-velocity of a star. Perryman (2018) solved the equations of motion of the multiple-body problem and described the path taken by a planet orbiting a larger body. In this work, I focus on the radial-velocity equations employed in MAGPy-RV. The stellar orbit around the common centre of mass can be derived with trigonometry to be

$$z = r(t) \cdot \sin(i) \sin(\nu(t) + \omega_\star), \tag{3.42}$$

where $r(t)$ is the distance of the star from the barycentre, i is the inclination of the orbit with respect to the observer (in the range of $0 \le i \le 2\pi$), ν is the true anomaly and $\omega_\star$ is the argument of periastron of the stellar orbit (related to the argument of periastron of the planetary orbit as $\omega_\star = \omega_{\text{pl}} + \pi$). The radial velocity of the star can thus be found by computing the first derivative of z as

$$\text{RV} = \frac{\partial z}{\partial t} = \sin(i)[\dot{r} \sin(\nu + \omega_\star) + r\dot{\nu}(\nu + \omega_\star)]. \tag{3.43}$$

This equation can then be simplified and the radial-velocity signal imprinted by the planet RV_{pl} can be modelled as

$$\mu(t) = \text{RV}_{\text{pl}} = A[\cos(\nu + \omega_\star) + e\cos(\omega_\star)], \tag{3.44}$$

where A represents the semi-amplitude of the signal and e is the eccentricity of the planetary orbit. The true anomaly ν is dependent on the eccentric anomaly ϵ as

$$\nu = 2\arctan\left[\sqrt{\frac{1+e}{1-e}} \cdot \tan\left(\frac{\epsilon}{2}\right)\right]. \tag{3.45}$$

The eccentric anomaly is itself dependent on the mean anomaly M as

$$M = \epsilon - e \sin(\epsilon). \tag{3.46}$$

While M can be directly computed as a function of the period of the planetary orbit P and its time of periastron passage T_p as

$$M = \frac{2\pi}{P}(t - T_p), \tag{3.47}$$

the relationship between ϵ and M cannot be solved analytically for all cases with $e \neq 0$. The eccentric anomaly is most often derived instead with either series expansions, or iterative methods (as is the case in MAGPy-RV).

The semi-amplitude A of the sinusoidal signal depends on the orbital characteristics of the planet as

$$A = \frac{2\pi}{P} \frac{a_\star \sin(i)}{\sqrt{(1-e)^2}}. \tag{3.48}$$

Since the orbital semi-major axis of the star $a_\star$ can be related to the orbital period of the planet P using an extension of Kepler's third law

$$P^2 = \frac{4\pi^2}{G \frac{M_{\rm pl}^3}{(M_\star + M_{\rm pl})^2}} a_\star^3, \tag{3.49}$$

Equation 3.48 can be rewritten in the form

$$A = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_{\rm pl} \sin(i)}{(M_\star + M_{\rm pl})^{2/3}} (1-e)^{-1/2}. \tag{3.50}$$

Therefore finally, the minimum mass of the planet can be approximated to

$$M_{\rm pl} \sin(i) \approx A \left(\frac{P}{2\pi G}\right)^{1/3} M_\star^{2/3} \sqrt{1-e^2}, \tag{3.51}$$

in the realm of $M_{\rm pl} \ll M_\star$.

3.2.1.4 Transit Curve

MAGPy-RV also includes a model for simultaneous transit photometry fitting. It relies on the BATMAN Python package[2] (Kreidberg 2015), which allows for fast computation of exoplanet transit light curves. Kreidberg (2015) computed the fraction δ of stellar luminosity obscured by the planet during transit as the integration of the sky-projected normalised intensity of the star I over the area blocked by the planet S:

$$\delta = \int\int IdS. \tag{3.52}$$

While it is common to approximate circular shape for both bodies, the brightness of the star cannot be considered equal in all positions of the stellar disc. Since a full numerical evaluation for complex stellar luminosity maps is slow and often cannot reach the required precision, BATMAN approximates the stellar intensity profile to be radially symmetric. They therefore can describe δ as

$$\delta = \sum_{i=1}^{n} I(x)\left(\frac{x_i + x_{i-1}}{2}\right)[A(x_i, r_{\mathrm{pl}}, d) - A(x_{i-1}, r_{\mathrm{pl}}, d)], \tag{3.53}$$

in which x is the radial coordinate normalised to be between 0 and 1, $I(x)$ is the one-dimensional stellar luminosity profile, r_{pl} is the radius of the planet and d is the separation between the centre of the circles representing star and planet. $A(x_, r_{\mathrm{pl}}, d)$ is a piece function that describes the area of intersection between the planet and star. An in-depth description of this function, and the approximations required for a fast algorithm can be found in Kreidberg (2015).

In summary, Chapter 3 focused on the description of Gaussian processes as a tool for data analysis, and in particular in the context of simultaneous planetary RV signal modelling and stellar activity mitigation. Section 3.1 provided a mathematical description of GPs, their covariance functions and regression methods. Section 3.2 introduced the pipeline I have developed, and described the relevant signal modelling functions used in exoplanet hunting.

---

[2] Code available at: https://github.com/lkreidberg/batman

Chapter 4

Detecting and Characterising Exoplanetary Systems

> *"A philosopher once asked, "Are we human because we gaze at the stars, or do we gaze at them because we are human?" Pointless, really..."Do the stars gaze back?" Now, that's a question"*
>
> — *Neil Gaiman, Stardust*

In this Chapter, I present the work I have done towards the detection and characterisation of two planetary systems. The first system orbiting the bright K-dwarf TOI-2134 is comprised of two planets, a short-orbit mini-Neptune and a long-orbit highly-eccentric sub-Saturn, detected with TESS, LCOGT, NEOSSat and WASP photometry data, and HARPS-N and SOPHIE radial velocities. The inner planet has since also been observed with Keck/NIRSPEC and escaping helium has been detected in its atmosphere (Zhang et al. 2023). In the second system, I detected three non-transiting super-Earths orbiting the nearby K-dwarf HD 48948 with long-baseline HARPS-N observations. The outermost planet was found to reside within the temperate habitable zone of its system, and has since been identified as a promising target for high-contrast direct imaging and high-resolution spectroscopic studies.

4.1 TOI-2134: a Mini-Neptune and a Temperate Eccentric Sub-Saturn[1]

4.1.1 Introduction

Since the discovery of the first exoplanet circa 30 years ago, more than 5000 have been detected and confirmed. Radial-velocity surveys performed with instruments such as the High Accuracy Radial-velocity Planet Searcher (HARPS) coupled with the Kepler photometric mission started discovering a sub-population of small exoplanets in short (under 100 days) orbits (Mayor and Udry 2008; Lovis et al. 2009; Fressin et al. 2009; Borucki et al. 2011; Batalha et al. 2013). Given their abundance in our galaxy (Chabrier et al. 2000; Winters et al. 2015), and their low mass and size, K and M dwarf stars are prime candidates for small-exoplanet searches and demographic-focused studies (Dressing and Charbonneau 2013; Crossfield et al. 2015; Astudillo-Defru et al. 2017; Pinamonti et al. 2018; West et al. 2019; Rice et al. 2019; Burt et al. 2020).
The transition point between rocky super-Earths and gaseous Neptunes is still debated (Fulton et al. 2017; Luque et al. 2021). Otegi et al. (2020) showed that this transition range is between 5-25 $M_\oplus$ and 2-3 $R_\oplus$, but several factors play into the composition of these planets. Some studies report that all planets under 1.6 $R_\oplus$ must be rocky (Rogers 2015; López-Morales et al. 2016). Others give more importance to the effects of irradiation: less irradiated planets are more likely to maintain a gaseous envelope, while more irradiated ones are typically rocky (Hadden and Lithwick 2014; Jontof-Hutter et al. 2016). Owen and Adams (2019) explored how planetary magnetic fields can also decrease their mass-loss rates and therefore alter the composition of the planetary cores. A continuous effort in the detection of small planets, and in the precise characterisation of their masses and sizes is therefore vital to reach a consensus on which parameters affect planetary composition.

On the other hand, our understanding of long-period planets is also lacking. The great majority of transit-detected exoplanets have periods shorter than 75 days (Jiang et al. 2019). Longer-period planets are harder to detect and determining their masses can be challenging. Moreover, the baselines of most photometric surveys also limit their detec-

---

[1] This Section, alongside all Figures and Tables have been taken from Rescigno et al. (2023b).

tion. This "missing" population hampers studies of planet demographics, of planet formation, and of how planetary characteristics depend on the host star (Winn 2011; Johnson et al. 2010). Temperate giants are located in a period valley, between 10 and 100 days, where gas planets are less frequent (Udry et al. 2003; Wittenmyer et al. 2010). Although more challenging to study, these cooler planets are valuable sources of information. For starters, temperate giant planets represent the middle step between the short-period Hot Jupiters and the gas giants of our own Solar System. They therefore can serve as bridges between their respective formation and migration theories (Huang et al. 2016). The composition of giant planets depends not only on the composition of the protoplanetary disk, but also on their location at birth and migration history. Consequently, studying their metal enrichment levels can constrain the processes driving core formation and envelope enrichment (Thorngren et al. 2016; Mordasini et al. 2016). Recent studies have also shown that long-period planets are correlated to and influence the dynamical evolution of the short-period planets within their systems (Zhu and Wu 2018; Bryan et al. 2019). Moreover, theoretical models predict that the formation of inner Earth-like planets is significantly dependent on the presence of quickly-accreted cold giants (Morbidelli et al. 2022). Due to their lower effective temperatures, the atmospheres of temperate giants produce entirely different molecular abundances and potentially can contain disequilibrium chemistry by-products (Fortney et al. 2020), making long-period gas planets valuable targets for atmospheric characterisation. Their atmospheres are less affected by temperature-induced inflation, which in turn allows us to use cooling models of planet evolution to constrain atmospheric metallicity (Ulmer-Moll et al. 2022). Additionally, there is a clear split in the eccentricity distribution of long-period planets. They are divided into a first group of objects with significantly high eccentricities and a second group with consistently nearly-circular orbits (Petrovich and Tremaine 2016). No clear cause of this bimodality has been found yet.

The numerous and highly varied scientific interests in exoplanet detection and characterisation have in the years motivated many space-based missions and ground-based instruments, including the second-generation HARPS-N (Cosentino et al. 2012) and the SOPHIE (Perruchot et al. 2008) spectrographs. Paired with space photometric missions (e.g., Ricker et al. 2015), the combination of transit photometry and radial velocity (RV) makes the determination of precise planetary masses and radii possible. The precision of

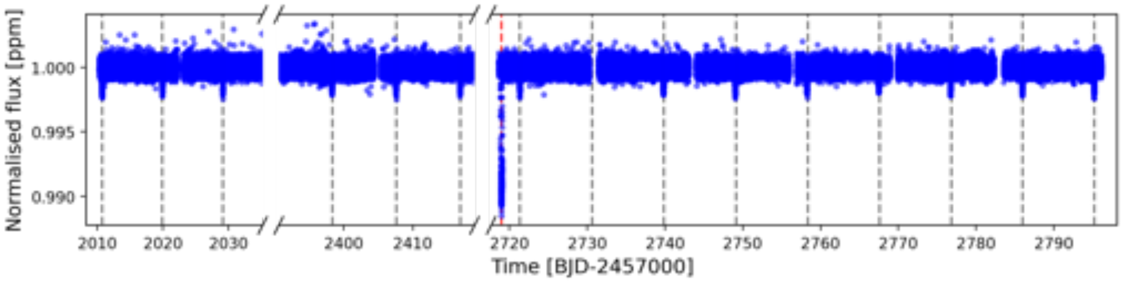

Figure 4.1: TESS normalised lightcurve over 5 sectors. 14 transits of an inner planet and a mono-transit of an outer planet can be seen and are indicated by the grey and red dashed lines respectively.

RV surveys has been steadily improving and the current uncertainty level reaches down to the tens of centimetres per second (Jurgenson et al. 2016; Thompson et al. 2016; Pepe et al. 2021), but the biggest obstacle remains stellar variability (Fischer et al. 2016; Crass et al. 2021), as addressed in Chapter 1. Great care is required when accounting for and modelling stellar activity in order to obtain accurate orbital solutions and to accurately and precisely determine planetary masses.

In the following Sections, I characterise the high proper-motion, bright (m_v=8.9 mag) K5-dwarf TOI-2134 and its planetary system. I detect a multi-transiting mini-Neptune in a short circular orbit and an outer temperate sub-Saturn planet. I also propose these targets for Rossiter-McLaughlin effect (Rossiter 1924; McLaughlin 1924; Queloz et al. 2000) follow-up and for atmospheric characterisation.

4.1.2 Observations

4.1.2.1 TESS Photometry

TOI-2134, also known as TIC 75878355 in the TESS Input Catalog (Stassun et al. 2018), was observed by NASA's Transiting Exoplanet Survey Satellite (TESS: Ricker et al. 2015) mission in 2-minute cadence mode over five sectors (Sectors 26, 40, 52, 53 and 54) for a total of 88,431 datapoints between BJD 2,459,010 and 2,459,035 (2020 June 9 to July 4), BJD 2,459,390 and 2,459,418 (2021 June 24 to July 22), and BJD 2,459,718 and 2,459,797 (2022 May 18 to August 5). The data were originally processed by the TESS Science Processing Operation Centre (SPOC) pipeline based at NASA Ames Research Center (Jenkins et al. 2016). However, Sector 40 showed strong residual systematics after the SPOC correction, so further systematics correction of the SPOC Simple Aperture Photometry (SAP) light curves (Twicken et al. 2010; Morris et al. 2020) was required. The resulting light curve was similar to the SPOC light curve (with slightly lower scatter) in most sectors, and yielded a major improvement in the problematic sector 40.

The transit signature of a TOI-2134b candidate was initially identified in a transit search conducted by the SPOC of Sector 26 on 24th July 2020 with an adaptive, noise-compensating matched filter (Jenkins 2002; Jenkins et al. 2010). Diagnostic tests were also conducted to help make or break the planetary nature of the signal (Twicken et al. 2010). The transit signatures for the TOI-2134b candidate were also detected in a search of Full Frame Image (FFI) data by the Quick Look Pipeline (QLP) at MIT (Huang et al. 2020b, 2020a) for Sector 40. A larger transit was detected by both QLP and the SPOC in searches including Sector 52. This transit was attributed to a second planetary candidate in the system, TOI-2134c. It appears to be a mono-transit and it did not re-occur in the following 75 days. The TESS Science Office (TSO) reviewed the vetting information and issued an alert on 7th August 2020 for TOI-2134b and on 28th July 2022 for TOI-2134c (Guerrero et al. 2021). The signal for the candidate TOI-2134b was repeatedly recovered as additional observations were made in sectors 26, 40, 52, 53, and 54, and the transit signatures passed all the diagnostic tests presented in the Data Validation reports. The difference image centroiding figure and difference images for the multi-sector Sector 26 - Sector 55 run for candidate TOI-2134b show that the centroid of the transit source is consistent with the target star of interest. The host star is located within 3.2±3.7 arcsec of the source of the transit signal for candidate TOI-2134b and within 0.98±2.59 arcsec of the source of the transit signal for candidate TOI-2134c. The light curve was flattened by simultaneously fitting transit models for the two planets along with a basis spline to model long-term trends, and then subtracting the long-term variations (a strategy similar to Vanderburg et al. 2016, except without a simultaneous systematics model; see also Pepper et al. 2020).

The systematics-corrected and flattened TESS data are shown in Fig. 4.1. To better constrain the characteristics of the mono-transiting long-period planet candidate, a ground- and space-based photometric observing campaign was launched to catch a second transit.

4.1.2.2 LCOGT Photometry

The Las Cumbres Observatory Global Telescope (LCOGT: Brown et al. 2013) network observed the star between BJD 2,459,808 and 2,459,818 (2022 August 17 to 27), when preliminary ephemeris prediction suggested the outer planet would re-transit. Due to an unfortunate combination of bad weather and low visibility, only a possible egress was detected. However, the LCO 0m4 SBIG detectors are very susceptible to strong systematics and several combinations of comparison stars and aperture sizes need to be examined to assess the overall reliability of a lightcurve feature, especially for ingress- or egress-only events. When using a different choice of comparison stars, a convincing egress was no longer present in the data. The apparent egress was, in fact, proven to be highly dependent on the choice of comparison star set. For this reason, I could not claim this egress as a detected transit on its own and I do not include this data in the analysis.

A TRansiting Planets and PlanetesImals Small Telescope (TRAPPIST) North (Barkaoui et al. 2017) observation of the outer planet was also attempted on 22$^{\text{nd}}$ August 2022, but it was unsuccessful.

4.1.2.3 NEOSSat Photometry

The position in the sky of TOI-2134 is such that it is not observable after late-October, which precluded the chance of a second ground-based campaign to detect a third transit of the outer planet candidate since the TESS detection. I therefore turned to space observations. TOI-2134 is outside of the CHEOPS field of view, but it is visible to the agile space telescope Near Earth Object Surveillance Satellite (NEOSSat: Hildebrand et al. 2004; Fox and Wiegert 2022). NEOSSat is a Canadian microsatellite orbiting the Earth in a Sun-synchronous orbit of approximately 100 minutes. It was originally deployed to study near-Earth satellites, but it also performs well for follow-up observations of large exoplanets transiting bright stars. It carries a 15cm f/6 telescope, with spectral range between 350 and 1050 nm and a field of view of 0.86×0.86 degrees.

NEOSSat observed TOI-2134 unevenly between BJD 2,459,898 and 2,459,910 (2022 November 14 to 26) with a 70s cadence for a total of 3,364 datapoints. Multiple sets of observations through the run show significant unpredictable offsets that are usually corrected with

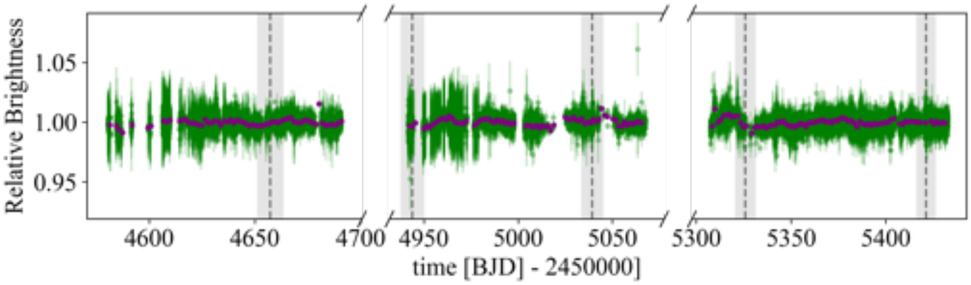

Figure 4.2: WASP normalised flux against Julian Date over the three years of coverage. All datapoints are plotted in green with errorbars, daily averages are plotted in purple. The predicted transits of TOI-2134c are plotted as grey dashed lines, while their uncertainties are plotted as grey shaded areas. As addressed in Section 4.1.8, no transit was detected.

calibration on reference stars. In these orbits, however, the reference stars behave differently from each other and the correction is less precise. This is probably due to image artefacts, as the detector and readout process have quite noticeable imperfections. These high-variance orbits have been flagged in the dataset and appear often enough to prevent a clear confirmation of a transit.

4.1.2.4 WASP Photometry

TOI-2134 was also observed over 3 years by the Wide Angle Search for Planets (WASP: Pollacco et al. 2006; Wilson et al. 2008) with coverage of about 120 nights per year. The data cover similar three-month spans between BJD 2,454,580 to 2,454,690 (2008 April 23 to August 11), BJD 2,454,941 to 2,455,067 (2009 April 19 to August 23) and BJD 2,455,307 to 2,455,432 (2010 April 20 to August 23). A total of 2,3097 datapoints were obtained and reduced with the SuperWASP pipeline (Pollacco et al. 2006). No planetary transit was detected. However, the long baseline, over three years long, allows for long-term monitoring of the stellar activity and of the rotation period of the host star, as shown in Section 4.1.4.1. All data are shown in Fig. 4.2.

4.1.2.5 HARPS-N Spectroscopy

A total of 111 radial-velocity observations of TOI-2134 were collected over two seasons with the High Accuracy Radial-velocity Planet Searcher for the Northern hemisphere spectrograph (HARPS-N: Cosentino et al. 2012; Cosentino et al. 2014) installed on the

3.6m Telescopio Nazionale Galileo (TNG) at the Observatorio del Roque de Los Muchachos in La Palma, Spain. HARPS-N is an updated version of HARPS at the ESO 3.6-m (Mayor et al. 2003). The spectrograph covers the wavelength range of 383-691 nm, with an average resolution R = 115,000. The first 32 spectra were collected between BJD 2,459,417 and 2,459,515 (2021 July 21 to October 27), and the next 79 were collected between BJD 2,459,638 and 2,459,890 (2022 February 27 to November 6). All data were observed under the Guaranteed Time Observations (GTO) programme with the standard observing approach of one observation per night. The average exposure time for TOI-2134 was 900s with an average signal-to-noise ratio (SNR) at 550 nm of ~100. RVs and activity indicators were extracted using the 2.3.5 version of the Data Reduction Software (DRS) adapted from the ESPRESSO pipeline (see Dumusque et al. 2021) and computed using a K6-type numerical weighted mask. The radial-velocity data show a peak-to-peak dispersion of 35 $\mathrm{m\,s^{-1}}$, with standard root mean squared (RMS) scatter of 7.3 $\mathrm{m\,s^{-1}}$ and mean uncertainty of 0.7 $\mathrm{m\,s^{-1}}$.

Several proxies are extracted by the standard DRS pipeline, including (but not limited to) the full-width at half-maximum (FWHM) and the contrast of the cross-correlation function (CCF), and the S-index. The mentioned data are plotted in purple in Fig. 4.3. The reasoning behind the selection of plotted proxies is addressed in Section 4.1.4.2.

4.1.2.6 SOPHIE Spectroscopy

113 radial-velocity observations of TOI-2134 were also obtained with the Spectrographe pour l'Observation des Phénomnes des Intérieurs stellaires et des Exoplanètes (SOPHIE: Perruchot et al. 2008) between BJD 2,459,082 and 2,459,894 (2020 August 20 to 2022 November 10). SOPHIE is a stabilized échelle spectrograph dedicated to high-precision RV measurements in the optical wavelengths (387 to 694 nm) on the 193cm Telescope at the Observatoire de Haute-Provence, France (Bouchy et al. 2009). SOPHIE high resolution mode (resolving power $R = 75,000$) and the fast mode of the CCD reading were used for the observations. The standard stars observed at the same epochs using the same SOPHIE mode did not show significant instrumental drifts. Depending on the weather conditions, the exposure times for TOI-2134 ranged from 4.5 to 30 minutes (average of 11 minutes) and their SNR per pixel at 550 nm ranged from 21 to 77 (average of 54). Five exposures showed

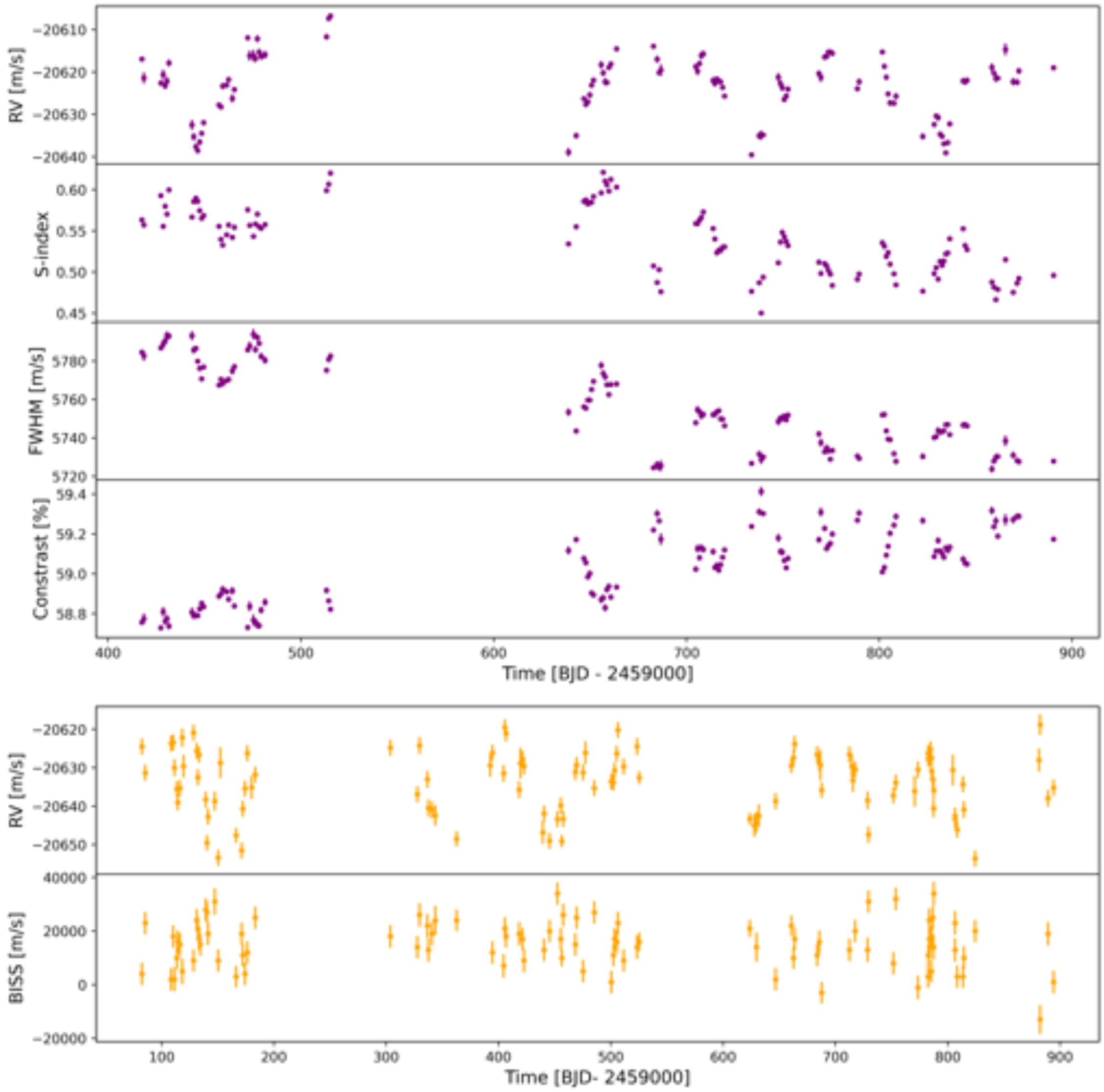

Figure 4.3: Plots of the HARPS-N and SOPHIE radial-velocity data alongside the chosen activity proxies for each dataset (see Section 4.1.4.2). From the top: HARPS-N RVs, S-index, full-width at half-maximum (FWHM) and contrast in purple, followed by SOPHIE RVs and their bisector span (BISS) in orange. Notice the different time axes. All error bars are plotted, but some are too small to be clearly visible in HARPS-N data.

a SNR below 40 and were removed. The final dataset therefore includes 108 epochs.

The radial-velocity data were extracted with the standard SOPHIE pipeline using CCFs (Bouchy et al. 2013) and including the CCD charge transfer inefficiency correction. The cross-correlations were made using several numerical masks, characteristic of different types of stars. All produced similar results in terms of RV variations. Finally, a K5-type mask was chosen, which provided the least dispersed results.

Following the method described e.g. in Pollacco et al. (2008) and Hébrard et al. (2008), the sky background contamination (mainly due to the Moon) was estimated and corrected using the second SOPHIE fibre aperture, which is targeted 2′ away from the first one pointing toward the star. 14 of the 108 exposures were significantly polluted by sky background. The final SOPHIE RVs show variations with a dispersion of 8.2 $\mathrm{m\,s^{-1}}$ (35 $\mathrm{m\,s^{-1}}$ peak to peak), significantly larger than their typical 2 $\mathrm{m\,s^{-1}}$ precision. The FWHM, bisector span and contrast of the CCF were also derived for every observation. The data are plotted in orange in Fig. 4.3 (for more information on proxy selection see Section 4.1.4.2).

4.1.3 Stellar Characterisation

TOI-2134 is a bright, high-proper motion, mid K-dwarf. As the star falls into a parameter space that is not optimal for several of the common stellar characterisation pipelines, the system was characterised with multiple separate and independent methods: a spectral energy distribution (SED) analysis adapted from Mann et al. (2015), Mann et al. (2016), and Mann et al. (2019); the ARES+MOOG method with ISOCHRONE fitting described in Mortier et al. (2014); the Stellar Parameter Classification pipeline (SPC: Buchhave et al. 2012; Buchhave et al. 2014); and a second SED technique developed by Morrell and Naylor (2019, 2020). For more information on the specifics of the stellar characterisation steps, see Rescigno et al. (2023b).

Overall, all analyses agree with each other within their uncertainties. For the scope of this work, I present the mean of all the computed values weighted by the inverse of their errors, as compiled in Table 4.1. Their uncertainties are computed as the standard deviation between measurements in each method, to avoid improper averaging down of systematic effects.

Table 4.1: Stellar parameters of TOI-2134.

Parameter	Value	Source
Name	TOI-2134	TESS Project*
	TIC 75878355	Stassun et al. (2019)
	G204-45	Giclas et al. (1979)
RA [h:m:s]	18:07:44.52	Gaia Collaboration (2020)
DEC [d:m:s]	+39:04:22.54	Gaia Collaboration (2020)
Spectral type	K5V	Stephenson (1986)
m_V [mag]	8.933±0.003	TESS Project*
m_J [mag]	6.776±0.023	TESS Project*
m_K [mag]	6.091±0.017	TESS Project*
$(B-V)$ [mag]	1.192±0.033	TESS Project*
Parallax [mas]	44.1087± 0.0144	Gaia Collaboration (2020)
Distance [pc]	22.655±0.007	Rescigno et al. (2023b)
Proper motion [mas/yr]	288.257±0.016	Gaia Collaboration (2020)
$L_\star$ [$L_\odot$]	0.192±0.008	Rescigno et al. (2023b)
F_{bol} [erg cm^2 s^{-1}]	1.198±0.048	Rescigno et al. (2023b)
$T_{\rm eff}$ [K]	4580±50	Rescigno et al. (2023b)
log(g) [cm s^{-1}]	4.8±0.3	Rescigno et al. (2023b)
[Fe/H]	0.12±0.02	Rescigno et al. (2023b)
Mass [$M_\odot$]	0.744±0.027	Rescigno et al. (2023b)
Radius [$R_\odot$]	0.709±0.017	Rescigno et al. (2023b)
Density [$\rho_\odot$]	2.09±0.10	Rescigno et al. (2023b)
Age [Gyr]	$3.8^{+5.5}_{-2.7}$	Rescigno et al. (2023b)
$v$sin(i) [km s^{-1}]	0.78±0.09	Rescigno et al. (2023b)
$< \log R'_{\rm HK} >$	-4.83±0.45	Rescigno et al. (2023b)
P$_{\rm rot}$ [days]	$45.78^{+5.56}_{-5.31}$	Rescigno et al. (2023b)

*See ExoFOP: https://exofop.ipac.caltech.edu/tess/target.php?id=75878355

4.1.4 Stellar Activity Signal Analysis

I conducted a thorough preliminary analysis of the available data in order to search for and to best characterise the stellar activity-induced signals in both the photometric and the spectroscopic observations.

The projected rotational velocity $v\sin(i)$ of TOI-2134 was determined to be <2 $\mathrm{km\,s^{-1}}$ from the HARPS-N spectra, and 1.5±1.0 $\mathrm{km\,s^{-1}}$ from the SOPHIE cross-correlation functions (following the method in Boisse et al. 2010). No more precise measurement could be derived from the spectra. I therefore calculated the minimum stellar rotation period $P_{\mathrm{rot,min}}$ associated to the lower maximum limit of $v\sin(i)$ as:

$$P_{\mathrm{rot,min}} = \frac{2\pi R_\star}{v\sin(i)} \approx 23 \text{ days}. \tag{4.1}$$

Using the method described in Noyes et al. (1984), I computed the average $\log(R'_{\mathrm{HK}})$ to be -4.83±0.45 dex from the S-index measurements taken by HARPS-N. There was significant scatter in the these data which degraded the quality of the results, but the empirical relations of Noyes et al. (1984) yielded a stellar rotation period of ~42 days. To better identify the stellar rotation period I performed a periodogram analysis.

4.1.4.1 Photometry

I computed the Bayesian Generalised Lomb-Scargle (BGLS) periodograms (Mortier et al. 2015) for both the WASP and the TESS photometric data, shown respectively in green and blue in the first and second rows of Fig. 4.4. The same periodograms in frequency space, alongside their window functions are plotted in Fig. 4.5. The TESS data showed a forest of peaks at ~9.2 days (highlighted by a black dashed line), which is generated by the repeated transits of the inner planet. As expected given the detection of no transits due to lower precision, the WASP periodogram had no power around this period. It instead showed two significant forests of peaks centred around ~29 and ~58 days (highlighted by blue bands in Fig. 4.4), which were originally attributed to the stellar rotation period, but could also be generated by the moon cycle. To further investigate this, I also plotted the BGLS periodograms of each yearly season of WASP, as shown in the first row of Fig. 4.4 as blue, red and purple dashed lines. The BGLS periodograms of the two later years also presented a significant peak at 58 days, but the 2008 data did not. Instead, its most significant peak was at 29 days. A peak at ~29 days was also present in the 2010 data, but not in 2009. While some of the discrepancies could be attributed to differing coverage, these results hinted at either a different lunar contribution over the different seasons, or at evolving surface inhomogeneities structure trends over the years, possibly related to a stellar magnetic cycle. After alias analysis, I found that the 29-days forest of peaks in the full periodogram can be explained as the extended aliases generated by the 1-year systematic. The WASP data span over ~850 days. SOPHIE radial velocities (taken 10 years later) also cover a similar stretch of time. Therefore, assuming these signals are of stellar origin, the structure of the stellar variability that allows us to detect the rotation period of the star can be expected to also evolve during the three years of radial-velocity data. This evolution could be the reason behind the difficulties at constraining the stellar rotation period in the further RV analyses.

4.1.4.2 Radial-Velocity Data and Proxies

Although the proxy time series present significant scatter and the average $\log(R'_{\mathrm{HK}})$ also classified the star as moderately active, both sets of RVs had little to no correlation to their activity indicators. The specific reason for this lack of correlation is ultimately beyond the scope of this work, as the activity indicators were only used as a starting point to the analysis, but I propose some possible origins. As a first most likely option, the Keplerian signals introduced by the planets in the system are large enough to "muddle" the correlation to activity indicators. As the amplitude of the stellar activity computed in the next Sections is comparable to the amplitude of the RV osciallations generated by the planets, it is likely that these signals are significant enough to prevent a clean correlation between RVs and activity indicators (which only map the variations induced by stellar activity). To test this, I also computed the correlation between the activity indicators and the RVs after subtracting the best-fit Keplerian models derived in Section 4.1.8. While the correlation did improve by a factor of 2, it still remained low. So other reasons may be considered. As an example, the stellar rotation axis inclination angle with respect to the observer can influence the strength of the proxy-RV correlation, weakening it for unfavourable line-of-sights: as the the stellar rotational axis becomes parallel to the observer line-of-sight, the

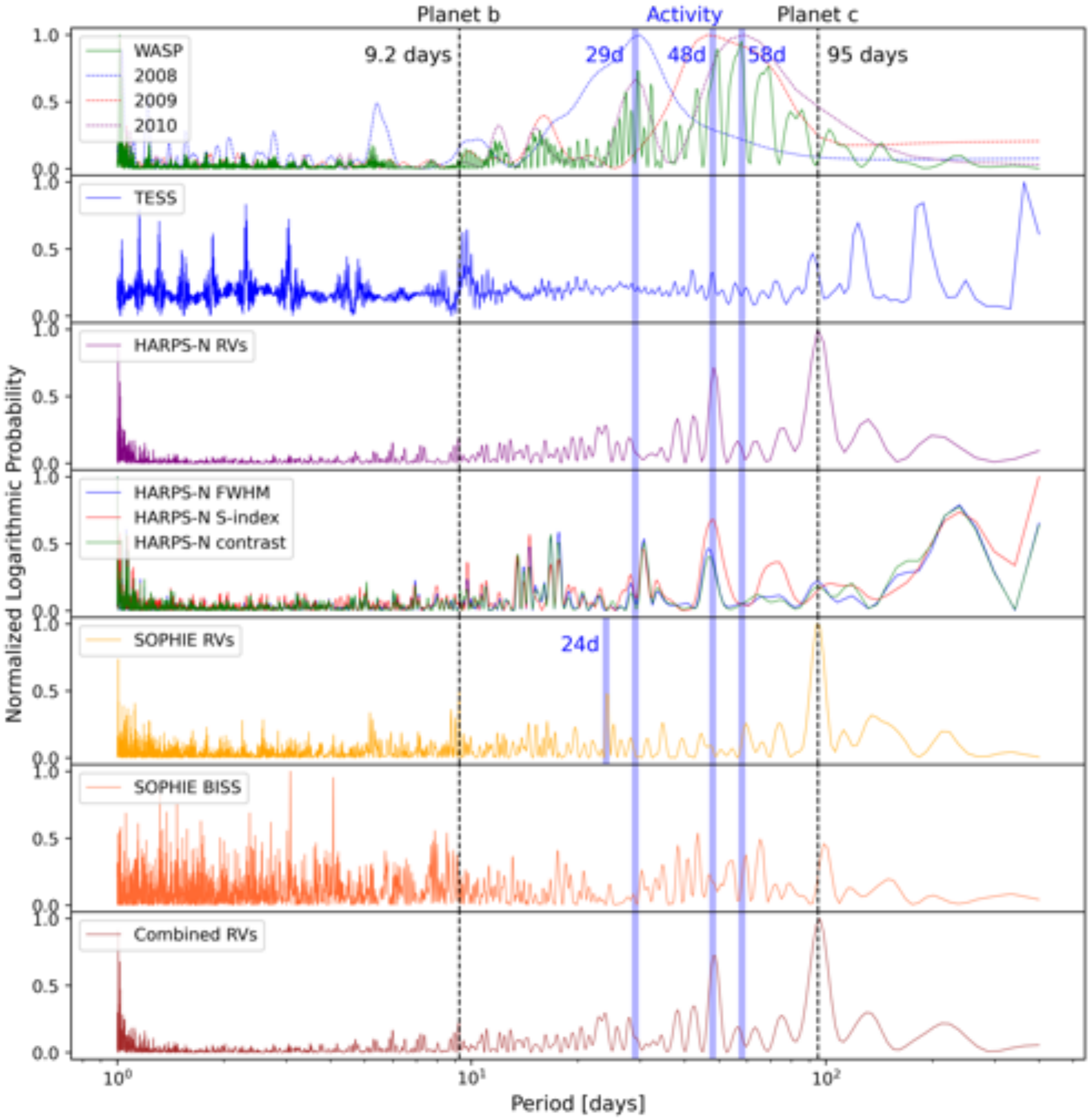

Figure 4.4: Set of BGLS periodograms of the acquired data plotted as period vs logarithmic probability normalised to 1. From the top, WASP photometry in the solid green with yearly seasons in blue for 2008, red for 2009 and purple for 2010 as dashed lines, TESS photometry, HARPS-N RVs, HARPS-N activity proxies (FWHM, S-index and contrast in respectively blue, red and green), SOPHIE RVs, SOPHIE activity proxy (BISS), and the combined SOPHIE and HARPS-N RVs. The dashed black lines represent the periods of the two planets at 9.2 and 95 days. The blue bands indicate the possible stellar rotation signals at 29, 48 and 58 days.

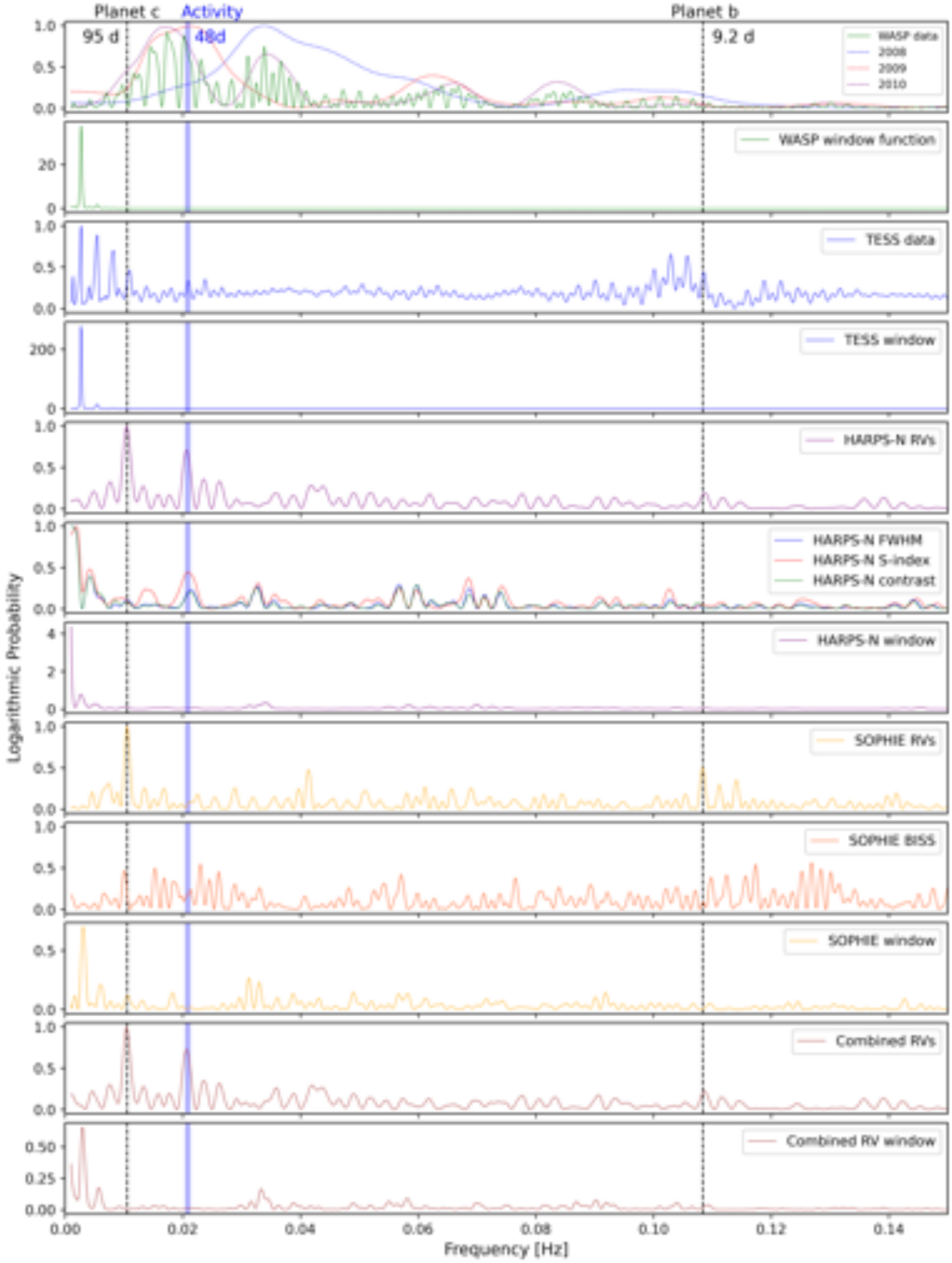

Figure 4.5: Same set of BGLS periodograms as Fig 4.4 in frequency space. The window functions for each dataset are also included. The dashed black lines indicate the periods of the two planet candidates. The blue band shows the stellar rotation period.

signal from active regions coming in and out of view becomes less rotationally modulated. At the same time, in late K-dwarfs convective redshift may in some cases prevail against blueshift. This can happen either due to an opacity effect (like in M-type stars), or if most of the photospheric absorption lines used for RV measurements form in regions of convective overshoot (Norris et al. 2017). Costes et al. (2021) notes that a possible explanation for low correlation between radial velocities and activity proxies, as is the case for our target, is that the convective blue- and redshifts are "cancelling" out one another. The possibility of a temporal lag (Collier Cameron et al. 2019) between the radial velocities and the proxies was also considered, but a visual inspection of their timeseries did not strongly support this possibility.
For the analysis, I nevertheless selected and plotted the indicators with the strongest correlation to their RVs. For HARPS-N I selected the S-index, the FWHM and the contrast. Their Spearman's rank correlation coefficients with their RVs were computed to be 0.15, 0.11 and -0.12 respectively. For SOPHIE I selected only the bisector span, with correlation coefficient of -0.16, as the FWHM and contrast seem to be affected by instrumental systematics.

I conducted a full periodogram analysis of the spectroscopic data. The last five rows of Fig. 4.4 show the BGLS periodograms of, in order, the HARPS-N RVs, the HARPS-N derived proxies (FWHM, S-index and contrast), the SOPHIE RVs, the SOPHIE-derived activity indicator (bisector span, or BISS), and the combined RV data. I was able to combine the RVs with a simple offset, as they are derived from similar wavelength windows and therefore are probing the same section of the stellar spectra. The same periodograms in frequency space, alongside their window functions, are once again shown in Fig. 4.5. While the BGLS periodograms of the radial-velocity datasets did not show clear peaks for the inner planet, there was a strong periodic signal at ~95 days (shown as a black dashed line) shared between the HARPS-N and the SOPHIE RVs that was not present in any of the HARPS-N stellar activity proxies. The SOPHIE bisector does have a peak at ~100 days, but its normalised logarithmic probability is comparable to most other peaks in the periodogram and therefore does not have a strong significance. This preliminary analysis suggested a period of ~95 days for the mono-transiting planet detected by TESS. The only major peak shared between the HARPS-N radial velocities and of all its activity

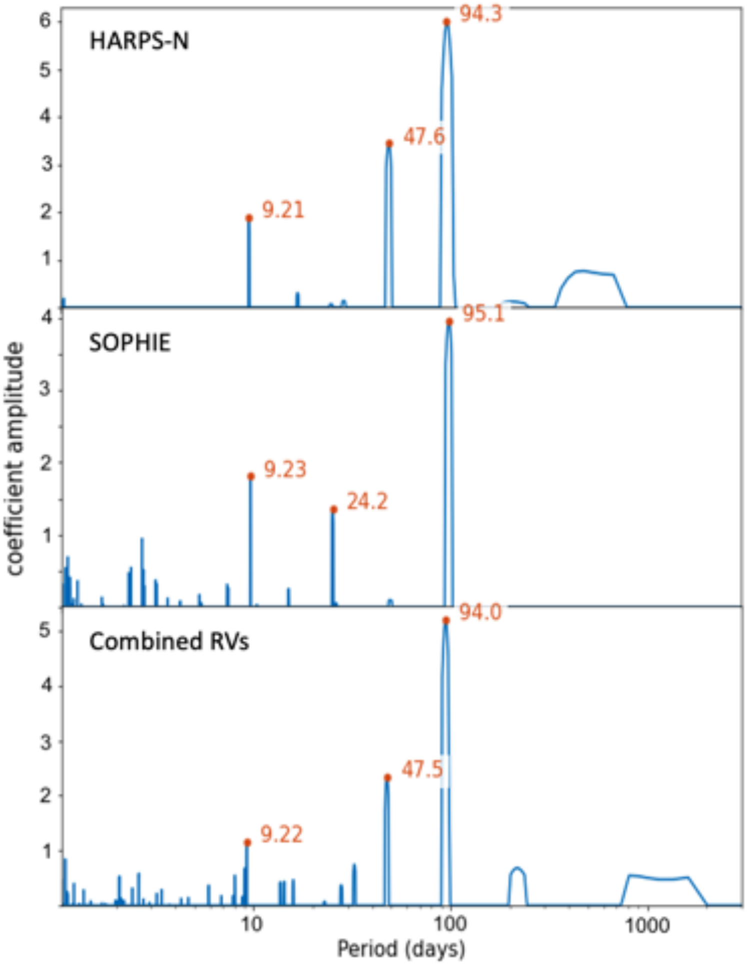

Figure 4.6: $\ell 1$ periodograms of from top to bottom HARPS-N, SOPHIE and combined RVs. The periods of the major signals identified are highlighted in red.

indicators was centred around 48 days (shown as a blue band), with yearly aliases at 42 and 38 days. No such signal could be found in either the SOPHIE RVs or their indicator. However, SOPHIE data presented a minor peak at ~24 days, half of the HARPS-N value. The 48-day signal as a tentative stellar rotation period, although not in perfect agreement, was compatible with the longer modulation of the WASP data, especially given the fact that their baseline in each season only spans just more than twice this period.

To further understand the signals within the spectroscopic datasets, I also performed an $\ell 1$ periodogram[2] analysis with correlated noise (Hara et al. 2017; Hara and Mari 2021), as shown in Fig. 4.6. This periodogram formulation was first devised to overcome the distortions in the residuals that arise when fitting planets one by one, as described in Chapter 2 Sections 2.3.1.2. Once again, HARPS-N and SOPHIE radial velocities on their

---

[2] Available at https://github.com/nathanchara/l1periodogram

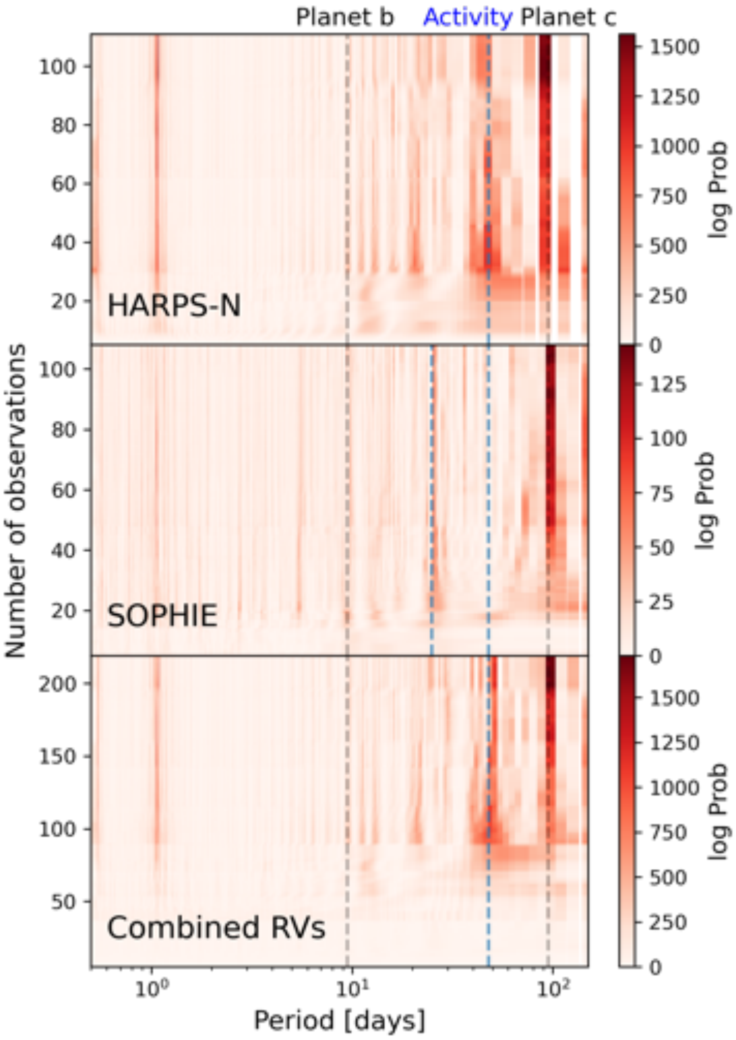

Figure 4.7: From the top, stacked BGLS periodograms of HARPS-N, SOPHIE and combined RVs. The blue dashed blue lines identify 48 days (and 25 days for SOPHIE data). The dashed grey lines show 9.2 and 95 days, the proposed periods of the two planets.

own, as well as their combination, showed a clear peak at ∼95 days. Similarly, the $\ell1$ periodograms of both HARPS-N and SOPHIE also peaked at ∼9.2 days. The $\ell1$ periodogram is also able to isolate the signal of the inner planet in the combined RV dataset. Regarding the possible stellar rotation period, HARPS-N data again showed a clear modulation at ∼48 days, while the strongest peak in SOPHIE not attributed to planetary signals was at half that value. The $\ell1$ periodograms therefore re-confirmed the previous results from the BGLS analysis.

Finally, to probe the coherence of these signals, I plotted the Stacked Bayesian Generalised Lomb-Scargle periodograms (Mortier et al. 2015; Mortier and Collier Cameron 2017) of the three sets of RV data in Fig. 4.7. The Stacked BGLS periodogram was developed to better identify the signals that are generated by stellar activity, as described in Chapter 2 Section 2.3.1.1.2. Planetary signals are coherent in nature, meaning their probability should consistently increase with increasing number of observations. Signals produced by stellar activity are incoherent, meaning that their probability will change and oscillate. Fig. 4.7 clearly showed that the signals indicated by the blue vertical lines (respectively 24 and 48 days, as identified by the $\ell 1$ periodograms) were incoherent. They therefore could not be attributed to planets and were more likely generated by stellar activity. The 9.2 and 95 days signals, highlighted by the grey dashed lines, showed more coherent trends. With the exception of a 1 day alias, no other major signals could be identified.

4.1.5 Transit Photometry

An analysis to determine the best-fit transit parameters and uncertainties for the two planet candidates orbiting TOI-2134 was then performed. The TESS photometry (after systematics correction and flattening as described in Section 4.1.2.1) was modelled with a Mandel and Agol (2002) transit model, which included three parameters describing the host star (its mean density, and both linear and quadratic q_1 and q_2 limb-darkening coefficient parametrisations sampled following Kipping 2013). The inner planet TOI-2134b was described by six parameters (its orbital period, time of transit, orbital inclination, the logarithm of the planet/star radius ratio $\log R_p/R_\star$, and combinations of the eccentricity and argument of periastron of the planet $\sqrt{e}\cos\omega_\mathrm{p}$ and $\sqrt{e}\sin\omega_\mathrm{p}$, which will be further explained in Section 4.1.6). The transit of the outer planet TOI-2134c was described by four parameters (time of transit, transit duration, impact parameter, and the logarithm of the planet/star radius ratio). Finally, two parameters characterising the dataset itself (a constant flux offset and the white noise level) were included.

4.1.5.1 Selection of Priors

An informative Gaussian prior was imposed on the stellar density based on the analysis of the stellar parameters. All other parameters were bound by uniform priors.The inclination of planet b was restricted to be less than 90° and the impact parameter of planet c to be greater than 0 (to avoid the degeneracy for transit configurations with inclinations greater 90°). $\sqrt{e}\cos\omega_\mathrm{p}$ and $\sqrt{e}\sin\omega_\mathrm{p}$ were bound to be in the interval [-1,1] (as necessary as per

Table 4.2: Results and uncertainties of the planetary parameters for the photometry analysis described in Section 4.1.5.

Parameter	Value
Radius ratio ($R_b/R_\star$)	0.03475± 0.00038
Orbital period P_b [days]	9.2292005±0.0000063
Time of transit $t_{0,b}$ [BJD]	2459407.54493±0.00027
Orbital inclination i_b [deg]	89.49± 0.37
Transit impact parameter b_b	0.21±0.14
Radius ratio ($R_c/R_\star$)	0.09404± 0.00078
Time of transit $t_{0,c}$ [BJD]	2459718.96939±0.00020
Transit impact parameter b_c	0.464± 0.042

their definition), and the impact parameters (in the case of TOI-2134b after conversion from inclination) to be in the range [0,1+$R_p/R_\star$] (requiring the planets to transit the star). log($R_p/R_\star$) was allowed to vary in the range $[-\infty, 0]$ (planets must be smaller than the host star), and q_1 and $q2$ in the range [0,1] following Kipping (2013). All other parameters with uniform priors were allowed to explore the range $[-\infty, \infty]$.

4.1.5.2 Transit Results

The parameter space was explored using a Markov Chain Monte Carlo (MCMC) algorithm with a Differential Evolution sampler (Ter Braak 2006). 100 chains were evolved simultaneously for 100,000 steps each, discarding the first 30,000 as burn-in. The convergence was assessed by calculating the Gelman-Rubin statistic. The best-fit models are phase-folded and plotted in Figure 4.8 and the results of the planetary fit are given in Table 4.2. Initially eccentricity, angle of periastron and period for the outer planet candidate were not derived, given the mono-transit. Those parameters will be extracted in a second step, as discussed in Section 4.1.6.2. The multiple transits of the inner planet allow to precisely measure its period and planet-to-star radii ratio. The radii ratio of TOI-2134c is also constrained to over 100σ.

4.1.6 RV Analysis

To analyse the radial velocities I used the MAGPy-RV pipeline introduced in Chapter 3 Section 3.2. I modelled the RV data as a combination of two planetary signals in the form

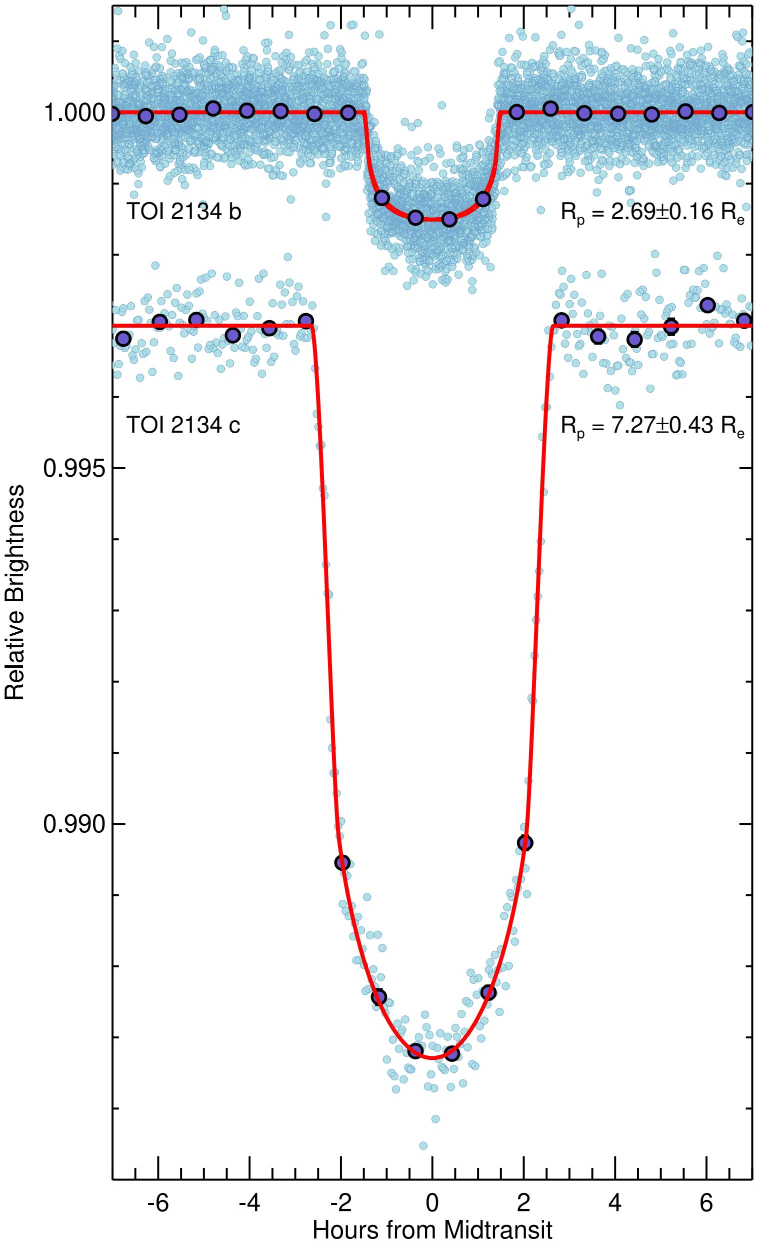

Figure 4.8: Phase-folded TESS light curves of TOI-2134b and c. Faint blue points are individual TESS two-minute cadence measurements, bold darker blue points are data binned in orbital phase, and the red curves are the best-fit transit models. The error bars on the binned points are smaller than the symbols. For the transit of TOI-2134c, an artificial offset to the out-of-transit flux measurements is included for improved visibility. Figure taken from Fig. 8 of Rescigno et al. (2023b) and produced by A. Vanderburg.

of Keplerians (for the two transiting objects), and the stellar activity with a GP function described by a Quasi-periodic kernel. We selected the QP Kernel formulated in Haywood et al. (2014) and defined in Chapter 3 Section 3.1.3 with the inclusion of a white noise "jitter" term, in the form

$$k(t_n, t_m) = A^2 \cdot \exp\left[-\frac{|t_n - t_m|^2}{l^2} - \frac{\sin^2\left(\frac{\pi \cdot |t_n - t_m|}{P_{\mathrm{rot}}}\right)}{h^2} \right] + \delta_{n,m}\beta^2, \tag{4.2}$$

in which t_n and t_m are two datapoints, the four hyperparameters are in order the maximum amplitude A, the timescale over which the quasi-periodicity evolves l, the period of the periodic variation P_{rot} (mapping the stellar rotation), and the "smoothness" of the fit h (its amount of high-frequency structure) also often referred to as the harmonic complexity. The "jitter" term is represented by the delta function, and β can be thought of as the contribution of the uncertainties of the RV measurements.

While eccentricity e, and planetary angle of periastron ω_{p} were used within the Keplerian model, when iterating in the MCMC algorithm I instead took steps in a different set of variables S_k and C_k, defined as

$$\begin{aligned} S_k &= \sqrt{e}\sin\omega_{\mathrm{p}}, \\ C_k &= \sqrt{e}\cos\omega_{\mathrm{p}}. \end{aligned} \tag{4.3}$$

As explained in Eastman et al. (2013), this reparameterisation avoids a boundary condition at zero eccentricity, allowing for a better sampling around zero while maintaining the overall prior flat over eccentricity.

The Keplerian models also depended on time of periastron passage t_{p}, rather than the time of transit t_0, derived by transit photometry. However, the two variables are linked via the following equation

$$t_{\mathrm{p}} = t_0 - \frac{P}{2\pi} \cdot \left[E_{\mathrm{tr}} - e \cdot \sin\left(E_{\mathrm{tr}}\right)\right], \tag{4.4}$$

in which P is the orbital period of the considered planet, e its eccentricity and the eccentric anomaly E_{tr} is computed from the argument of periastron and the eccentricity as described in Chapter 3 Section 3.2.1.3.

I conducted the investigation on the combined HARPS-N and SOPHIE time series, as well as on the two datasets separately. Once again, I was able to combine the two RV time series with a simple offset parameter and could use a single GP to describe both because they have comparable jitters and they are extracted from similar spectral windows in the optical range. Therefore, they are expected to map the same physical processes and to be sensitive to Doppler-shifts in the same way.

4.1.6.1 Selections of Priors

In this Section, I describe the choices of priors for the analysis of the RV data. The same priors are used for all three analyses. They are also summarised in Table 4.3.

Starting with the Keplerians, I imposed a strict 1σ Gaussian prior on the orbital period of the inner transiting planet, P_{b}, derived from the posterior distribution of the same variable in the transit photometry analysis. Similarly, I imposed a strict Gaussian prior to the time of periastron passage, $t_{\mathrm{p,b}}$, inflating the σ to account for the uncertainties in the eccentricity of the planet. The period of the outer planet was bound by a uniform prior between [75,150], derived from the minimum period allowed by consecutive TESS photometry and the information derived from the periodogram analysis. Given the inability to derive a period from transit photometry, the time of periastron passage of the outer planet $t_{\mathrm{p,c}}$ was bound by a uniform prior in the range [2,459,678.5, 2,459,773.5], determined by the preliminary P_{c} from the periodogram analysis. S_k and C_k for both planets were also bound by uniform priors in the range [-1, 1]. The SOPHIE-HARPS-N offset was allowed to vary only in the [-5,5] $\mathrm{m\,s^{-1}}$ interval. The rest of the parameters are left with wide positive (larger than zero) uniform priors.

Regarding the kernel hyperparameters, I applied a strict Gaussian prior to h (the "smoothness" of the fit) centred on 0.5±0.1, as recommended by Jeffers and Keller (2009). This choice is grounded in the fact that even highly complex active-region distributions average out to just two or three large active regions per rotation. I set a wide Gaussian prior on the stellar rotation period P_{rot} derived from the periodogram analysis centred in 48 days with σ=10 days, as wide as the forest of peaks in the WASP BGLS periodogram. The evolution timescale l is bound by a wide Jeffreys prior. A wide positive (larger than zero) uniform prior was applied to the GP amplitude A, and the jitter was only allowed to vary in the interval [0,2] $\mathrm{m\,s^{-1}}$.

4.1.6.2 The Eccentricity of TOI-2134c

Initial analysis of the radial-velocity data showed a significant trimodality in the distribution of the eccentricity of the outer 95 day-orbit planet, e_c. After further investigation, I found that multiple fully-converged models with different outer planet eccentricities existed. The RVs allowed for eccentricities of TOI-2134c equal to $0.002^{+0.029}_{-0.002}$, 0.45 ± 0.05 and $0.67^{+0.05}_{-0.06}$. All the models agreed within their uncertainties for most other parameters. Significantly large eccentricities have been detected before for temperate gas planets and stability can be reached within this system, so I could not a priori exclude any of the models. The stellar rotation period derived from the analysis is close to half the period of TOI-2134c. I therefore postulated that an interaction between the fit of the Keplerian model and the stellar activity-induced signal could be the reason behind the multiple models. While the flexibility of GPs are what makes them valuable tools to model stellar activity, in this case this flexibility allowed the Keplerian to take different accepted forms, while absorbing any "left-over" signal into the activity model. To further compare the final likelihoods of the three solutions, I computed the corrected Akaike Information Criterion, AICc, (Sugiura 1978) for all converged models:

$$\mathrm{AICc} = \quad \mathrm{AIC} + 2\left(\frac{N_{\mathrm{free}}(1 + N_{\mathrm{free}}\ )}{N_{\mathrm{data}} - N_{\mathrm{free}} + 1}\right), \tag{4.5}$$

where N_{free} is the number of free parameters and N_{data} is the number of data points. The original Akaike Information Criterion, AIC, (Akaike 1983) is calculated as

$$\mathrm{AIC} = -2\log(\mathcal{L}) + 2N_{\mathrm{free}}, \tag{4.6}$$

where $\log(\mathcal{L})$ is the logarithmic likelihood maximised after the MCMC analysis. The larger the AICc the less likely the model. The AICcs of the combined (HARPS-N + SOPHIE) RV data for the low-, medium- and high-eccentricity models were respectively 1224.0 and 1195.7 and 1196.7. As a further check, and to test whether this system would significantly benefit from a simpler analysis, I also computed the Keplerian-only best-fit model to the data. For this analysis I only included the planetary model with a jitter term

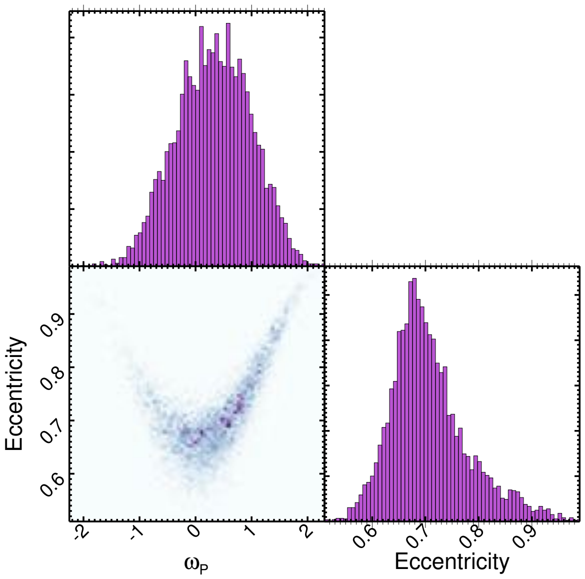

Figure 4.9: Posterior distribution corner of the eccentricity and the argument of periastron ω_P of the outer planet c derived after MCMC model optimisation on the deep mono-transit present in the TESS data, as explained in Section 4.1.6.2. Most notable, the eccentricity of TOI-2134c converges to a high ~0.7 value.

and no stellar activity or GP component. This last model struggled to converge and its AICc was 1253.2. The AICc analysis thus strongly disfavoured the Keplerian-only model and the circular-orbit model (with AICc difference from the best model larger than 7). However, the AICc values for the medium and the high eccentricity cases were similar enough that no single model was significantly favoured and no significant statistical preference could be reached.

I then turned to the obtained photometric data. The orbital period of the singly-transiting planet candidate was estimated using only the TESS light curve, following the procedure of Vanderburg et al. (2018). This method did not take into consideration the results from radial velocity, and derived the planetary period directly from the photometric mono-transit. The impact parameter b_c, planet-star radius ratio $R_c/R_\star$, and total transit duration of the single transit candidate were extracted from the MCMC posteriors of the two-planet transit fit. I then solved for the orbital period assuming the stellar parameters reported in Table 4.1 and an eccentricity probability distribution from Kipping (2014). The constraint that a second transit was not observed by TESS was also imposed, which required the orbital period to be longer than about 75 days. The short duration of the transit and minimum period allowed by TESS ruled out circular orbits for this planet with periastron passage happening near the time of transit (as expected from geometric arguments), as the RV models comparison also had found. The eccentricity e_{c} and argument of periastron $\omega_{\mathrm{p,c}}$ required to reproduce the transit data were then estimated, assuming the orbital period larger than 75 days. The posterior probability distributions of e_{c} and $\omega_{\mathrm{p,c}}$ are shown in Fig. 4.9. The eccentricity was required to be high (~0.7), and the argument of periastron was broadly identified to be near the conjunction of the orbit of the planet. This eccentricity value derived from transit photometry was then used to constrain the RVs. Given the high-eccentricity preference, I added a Gaussian prior centred in 0.7 with a σ of 0.1 to e_{c}.

4.1.6.3 RV Results

A summary of the final results of the RV analyses can be found in Table 4.3. For this MCMC analysis I simultaneously evolved 100 chains for 100,000 iterations each, discarding a burn-in phase of 20,000 steps. I assessed the health and convergence of the chains by computing the Gelman-Rubin statistic and all parameters reached values under the 1.1 convergence cut. As mentioned in the previous Section, I tested a series of models. For each set of HARPS-N only, SOPHIE only and combined RVs I evolved Keplerian-only models with no stellar activity (which overall struggled to converge or did not converge), forced circular-orbit models, medium-eccentricity models, and finally high-eccentricity models bound with an eccentricity prior derived by the photometry analysis. In this work I only present the last set.

The HARPS-N only data could constrain the amplitude and period of the inner TOI-2134b better than the SOPHIE data could, but conversely the SOPHIE RVs were able to better identify the signal of the outer planet, especially its period. A combined analysis allowed to more robustly constrain both planets with a single model. Since all three of the GP models fully converged and reached final values consistently within 1σ of each other, I only discuss the results of the combined RV analysis.

The periods of the two planets are well-defined. Their RV amplitudes are constrained to 12σ for planet b and to 6σ for planet c. The MCMC struggled to constrain the stellar activity evolution timescale l, as expected from the low correlation with activity indicators and the weak overall rotational modulation (see Section 4.1.4.2). The stellar rotation period is derived to be $45.78^{+5.56}_{-5.31}$ days.

Table 4.3: Results from the three Gaussian Process regression analysis. I include the priors applied to each parameter. In order the HARPS-N RVs only, and the SOPHIE RVS only results, followed by the combined HARPS-N and SOPHIE data results (used for all further analysis). Uniform priors are abbreviated as $\mathcal{U}$, Gaussian priors as $\mathcal{G}$ and Jeffreys' priors as $\mathcal{J}$. Only the results obtained for the high eccentricity case are shown, as addressed in Section 4.1.6.2.

Parameter	Prior	HARPS-N RVs	SOPHIE RVs	Combined RVs
GP Amplitude A [$\mathrm{m\,s^{-1}}$]	$\mathcal{U}[0, 20]$	$4.24^{+0.81}_{-0.59}$	$5.52^{+0.98}_{-0.68}$	$5.52^{+0.67}_{-0.71}$
GP Timescale l [days]	$\mathcal{J}[0, 100]$	$31.84^{+9.93}_{-10.36}$	$10.15^{+22.51}_{-7.99}$	$25.05^{+8.48}_{-8.53}$
GP Period P_{rot} [days]	$\mathcal{G}[48, 10]$	$45.85^{+4.89}_{-4.84}$	$38.89^{+13.99}_{-14.17}$	$45.78^{+5.56}_{-5.31}$
GP Smoothness h	$\mathcal{G}[0.5, 0.1]$	$0.48^{+0.05}_{-0.05}$	$0.48^{+0.06}_{-0.05}$	$0.48^{+0.05}_{-0.05}$
Jitter [$\mathrm{m\,s^{-1}}$]	$\mathcal{U}[0, 2]$	$0.69^{+0.13}_{-0.12}$	$0.82^{+0.24}_{22}$	$0.91^{+0.14}_{-0.13}$
SOPHIE - HARPS-N Offset [$\mathrm{m\,s^{-1}}$]	$\mathcal{U}[-5, 5]$			$2.30^{+0.44}_{-0.45}$
Orbital period P_b [days]	$\mathcal{G}[9.2292004, 0.0000063]$	$9.22923^{+0.00004}_{-0.00003}$	$9.2292^{+0.0002}_{-0.0001}$	$9.22923^{+0.00004}_{-0.00004}$
RV Amplitude K_b [$\mathrm{m\,s^{-1}}$]	$\mathcal{U}[0, 20]$	$3.01^{+0.32}_{-0.32}$	$4.13^{+0.84}_{-0.87}$	$3.40^{+0.28}_{-0.29}$
$S_{\mathrm{k,b}}$	$\mathcal{U}[-1, 1]$	$-0.04^{+0.06}_{-0.08}$	$0.21^{+0.08}_{-0.09}$	$-0.07^{+0.06}_{-0.09}$
$C_{\mathrm{k,b}}$	$\mathcal{U}[-1, 1]$	$0.22^{+0.06}_{-0.09}$	$0.21^{+0.07}_{-0.09}$	$0.21^{+0.06}_{-0.08}$
Time of periastron $t_{\mathrm{peri,b}}$ [BJD]	$\mathcal{G}[2459408.22, 0.50]$	$2459407.71^{+0.46}_{-0.38}$	$2459407.55^{+0.43}_{-0.44}$	$2459407.89^{+1.52}_{-1.23}$
Orbital period P_c [days]	$\mathcal{U}[75, 150]$	$94.71^{+1.17}_{-1.11}$	$94.86^{+1.13}_{-0.83}$	$95.50^{+0.36}_{-0.25}$
RV Amplitude K_c [$\mathrm{m\,s^{-1}}$]	$\mathcal{U}[0, 20]$	$11.92^{+1.82}_{-1.82}$	$10.28^{+2.99}_{-2.94}$	$9.74^{+1.60}_{-1.63}$
$S_{\mathrm{k,c}}$	$\mathcal{U}[-1, 1]$	$-0.65^{+0.11}_{-0.07}$	$0.69^{+0.29}_{-0.10}$	$-0.57^{+0.19}_{-0.13}$
$C_{\mathrm{k,c}}$	$\mathcal{U}[-1, 1]$	$0.42^{+0.13}_{-0.21}$	$0.41^{+0.66}_{-0.32}$	$0.59^{+0.13}_{-0.16}$
Time of periastron $t_{\mathrm{peri,c}}$ [BJD]	$\mathcal{U}[2459678.5, 2459773.5]$	$2459724.33^{+3.27}_{-2.53}$	$2459731.05^{+4.07}_{-8.86}$	$2459721.20^{+1.52}_{-1.23}$

4.1.7 Joint RV and Transit Analysis

Finally, I also modelled the TESS photometry and the radial-velocity data jointly, to more robustly test whether the high eccentricity model was still favoured. This more complex analysis allowed for simultaneous modelling of the orbital solutions for both planets. I once again used the code MAGPy-RV. I modelled the RVs similarly to Section 4.1.6, as two Keplerian signals for the planet candidates, with a GP function defined by a Quasi-periodic kernel describing the stellar activity and an offset parameter to match the zero-line of the HARPS-N and the SOPHIE datasets. For the TESS data, I described the transits of both planets with six parameters each (period, time of transit, $S_{\rm k}$, $C_{\rm k}$, planet to stellar radius ratio, and orbital inclination). The photometric model also included five parameters to describe the host star (its mean density, q_1, q_2, photometric jitter and offset). In this analysis, I was jointly modelling the periods, time of transits, eccentricities and angles of periastron of both planet.

4.1.7.1 Selection of Priors

I imposed similar priors on the GP hyperparameters as described in Section 4.1.6.1: Gaussian priors on the stellar rotation period and the harmonic complexity, uniform priors on amplitude and RV jitter, and a Jeffreys prior on the evolution timescale. The RV offset between SOPHIE and HARPS-N data was also similarly bound by a uniform prior between [-5,5] $\mathrm{m\,s^{-1}}$. The period of the inner planet, $P_{\rm b}$, was bound by a Gaussian prior centred on 9.2 days with σ of 0.2 days derived from preliminary transit analysis. The time of transit $t_{\rm tr,b}$ was also similarly bound by a Gaussian prior. The period of the outer planet, $P_{\rm c}$, was bound by a uniform prior between [75, 150] days, as it was in the original RV analysis. The RV amplitude of both planets were as before bound between [0,20] $\mathrm{m\,s^{-1}}$. $S_{\rm k}$ and $C_{\rm k}$ of both planets were only allowed to vary in the interval [-1,1] by definition. For the photometry, the stellar density was bound by a Gaussian prior centred on the derived stellar density with σ equal to its uncertainty (see Table 4.1). I allowed both planet-star radii ratios, $R_{\rm b}/R_\star$ and $R_{\rm c}/R_\star$, to only vary between [0,1] (the planets are expected smaller then the star), q_1 and q_2 between [0,1] as per their definition, and I required both inclinations i to be less than 90°. All other priors were flat uninformative priors.

Table 4.4: Results and uncertainties of the planetary parameters for the joint photometry and RV analysis described in Section 4.1.7

Parameter	Value
GP Amplitude A [m s^{-1}]	$4.59^{+1.38}_{-1.29}$
GP Timescale l [days]	$28.01^{+21.31}_{-22.15}$
GP Period $P_{\rm rot}$ [days]	$53.87^{+3.14}_{-3.02}$
GP Smoothness h	$0.44^{+0.08}_{-0.06}$
Jitter [m s^{-1}]	$0.85^{+0.95}_{-0.59}$
SOPHIE HARPS-N Offset [m/s]	$2.64^{+0.09}_{-0.06}$
Orbital period $P_{\rm b}$ [days]	$9.229209^{+0.000006}_{-0.000004}$
Radius ratio $(R_{\rm b}/R_\star)$	0.02±0.01
Orbital inclination $i_{\rm b}$ [deg]	$89.91^{+0.05}_{-0.06}$
RV Amplitude $K_{\rm b}$ [m s^{-1}]	$3.51^{+0.33}_{-0.41}$
Eccentricity $e_{\rm b}$	$0.05^{+0.03}_{-0.03}$
Argument of periastron $\omega_{\rm p,b}$ [rad]	$-0.75^{+0.47}_{-0.88}$
Time of periastron $t_{\rm p,b}$ [BJD]	$2459407.82^{+0.09}_{-0.06}$
Orbital period $P_{\rm c}$ [days]	$94.98^{+0.95}_{-1.02}$
Radius ratio $(R_{\rm c}/R_\star)$	0.09±0.01
Orbital inclination $i_{\rm b}$ [deg]	$89.91^{+0.02}_{-0.03}$
RV Amplitude $K_{\rm c}$ [m s^{-1}]	$9.83^{+0.85}_{-0.89}$
Eccentricity $e_{\rm c}$	$0.62^{+0.09}_{-0.02}$
Argument of periastron $\omega_{\rm p,c}$ [rad]	$1.41^{+0.49}_{-0.48}$
Time of periastron $t_{\rm p,c}$ [BJD]	$2459432.39^{+3.11}_{-3.01}$

4.1.7.2 Joint Analysis Results

I simultaneously evolved 100 chains for 100,000 iterations each, discarding once again a burn-in phase of 20,000 steps, and I tested for convergence with the Gelman-Rubin statistic. The results of the joint analysis for the combined SOPHIE and HARPS-N RVs and the TESS photometry are listed in Table 4.4. All parameters agreed within 1σ uncertainty with the results from the previous less complex transit and RV analyses, shown in Tables 4.2 and 4.3. These results once again confirmed the high-eccentricity model for the outer planet TOI-2134c.

Overall, both planet candidates and their periods were fully recovered. Their RV amplitudes were constrained to 10σ for the inner planet and 11σ for the outer one. The joint photometry and RV analysis is minorly less effective in the retrieval of the RV signal of inner planet than the radial-velocity data on their own, but it performed better for TOI-

Table 4.5: System Parameters for TOI-2134b. The transit and radial-velocity parameters are computed in Sections 4.1.5 and 4.1.6. Derived parameters are addressed in Section 4.1.8 and its subsections alongside the necessary assumptions.

Parameter	Value
TOI-2134 b	
Transit and Radial-Velocity Parameters	
Orbital period P_b [days]	9.2292005±0.0000063
Time of transit $t_{0,\mathrm{b}}$ [BJD]	2459407.54493±0.00027
Radius ratio ($R_\mathrm{b}/R_\star$)	0.03475± 0.00038
Orbital inclination i_b [deg]	89.49± 0.37
Transit impact parameter b_b	0.21±0.14
Transit duration τ_b [hours]	2.995±0.047
RV Amplitude K_b [$\mathrm{m\,s^{-1}}$]	$3.40^{+0.28}_{-0.29}$
Eccentricity e_b	$0.06^{+0.03}_{-0.04}$
Argument of periastron $\omega_{\mathrm{p,b}}$ [rad]	$1.91^{+0.32}_{-0.34}$
Time of periastron $t_{\mathrm{p,b}}$ [BJD]	$2459407.89^{+0.45}_{-0.49}$
Derived Parameters	
Radius R_b [$\mathrm{R_\oplus}$]	2.69±0.16
Mass M_b [$\mathrm{M_\oplus}$]	$9.13^{+0.78}_{-0.76}$
Density ρ_b [$\mathrm{kg\,m^{-3}}$]	2607±516
Density ρ_b [$\rho_\oplus$]	0.47±0.09
Scaled semi-major axis ($a_\mathrm{b}/R_\star$)	23.66±0.52
Semi-major axis a_b [AU]	0.0780±0.0009
Incident Flux $F_{\mathrm{inc,b}}$ [$F_{\mathrm{inc},\oplus}$]	32±2
Equilibrium temperature $T_{\mathrm{eq,b}}$ [K]	666±8

2134c. Once again, the stellar activity evolution timescale is not very well-constrained. The stellar rotation period was here derived to be slightly longer ($54.27^{+3.27}_{-3.23}$) but it was still consistent with the previous analysis. Both planet radius ratios were fully retrieved to 2 and 9σ for TOI-2134 b and c respectively.

4.1.8 Results and Discussion

The results of the joint photometry and RV analysis fully agreed within their 1σ uncertainties with the results from the separate transit and RV analyses. While the joint method successfully retrieved and characterised both planet candidates, I primarily focused the results of the less complex, separated analyses undertaken in Sections 4.1.5 and 4.1.6. All the final results are compiled in Tables 4.5 and 4.6. The covariance hyperparameters of

Table 4.6: System Parameters for TOI-2134c. The transit and radial-velocity parameters are computed in Sections 4.1.5 and 4.1.6. Derived parameters are addressed in Section 4.1.8 and its subsections alongside the necessary assumptions.

Parameter	Value
TOI-2134 c	
Transit and Radial-Velocity Parameters	
Orbital period P_c [days]	$95.50^{+0.36}_{-0.25}$
Time of transit $t_{0,c}$ [BJD]	2459718.96939±0.00020
Radius ratio $(R_c/R_\star)$	0.09404± 0.00078
Transit impact parameter b_c	0.464± 0.042
Transit duration $\tau_{,c}$ [hours]	5.267±0.028
RV Amplitude K_c [m s^{-1}]	$9.74^{+1.60}_{-1.63}$
Eccentricity e_c	$0.67^{+0.05}_{-0.06}$
Argument of periastron $\omega_{p,c}$ [rad]	$02.32^{+0.22}_{-0.32}$
Time of periastron $t_{p,c}$ [BJD]	$2459721.20^{+1.52}_{-1.23}$
Derived Parameters	
Radius R_c [R$_\oplus$]	7.27± 0.42
Mass M_c [M$_\oplus$]	$41.89^{+7.69}_{-7.83}$
Density ρ_c [kg m^{-3}]	599±152
Density ρ_c [$\rho_\oplus$]	0.11±0.03
Scaled semi-major axis $(a_c/R_\star)$	112±2
Semi-major axis a_c [AU]	0.371±0.004
Incident Flux $F_{inc,c}$ [$F_{inc,\oplus}$]	1.4±0.1
Equilibrium temperature $T_{eq,c}$ [K]	306±4

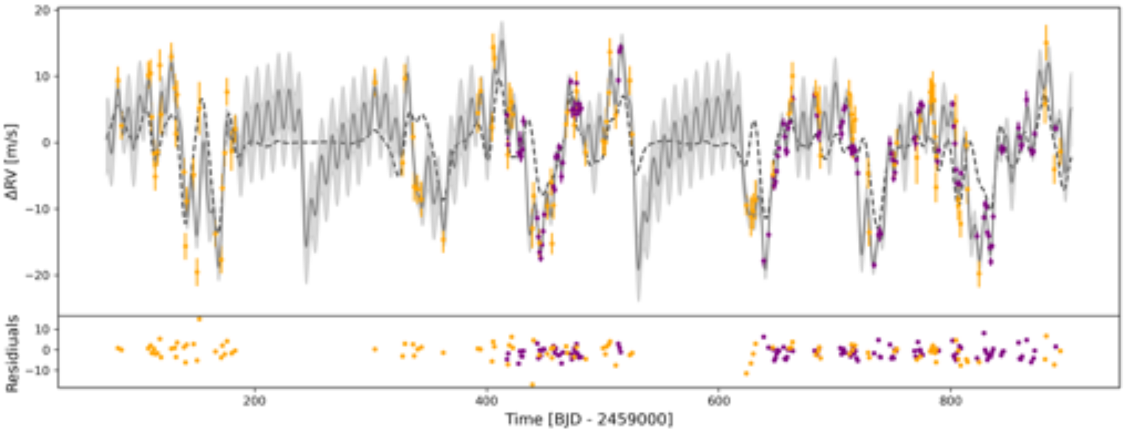

Figure 4.10: Combined SOPHIE (orange) and HARPS-N (purple) radial-velocity data plotted with errorbars (HARPS-N errorbars are too small to be clearly visible). The complete model, which includes two Keplerians and the predicted activity, is plotted in grey, with its uncertainties as the gray shaded area. The dashed black line represents the GP activity prediction only. On the bottom, the residuals between the data (in the corresponding colour) and the complete model are plotted.

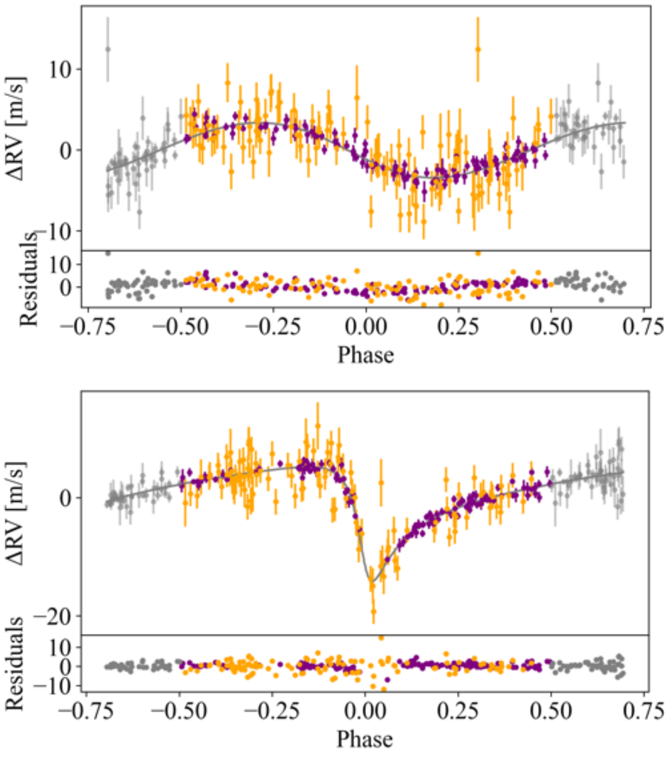

Figure 4.11: Phase folded activity model-subtracted plots for the inner (top) and outer (bottom) planets. In orange are the SOPHIE RVs and in purple the HARPS-N ones with respective errorbars (some HARPS-N errorbars may be too small to be visible). The Keplerian model is plotted as a gray line, with the residuals shown on the bottom. The phase has also been extended on both sides.

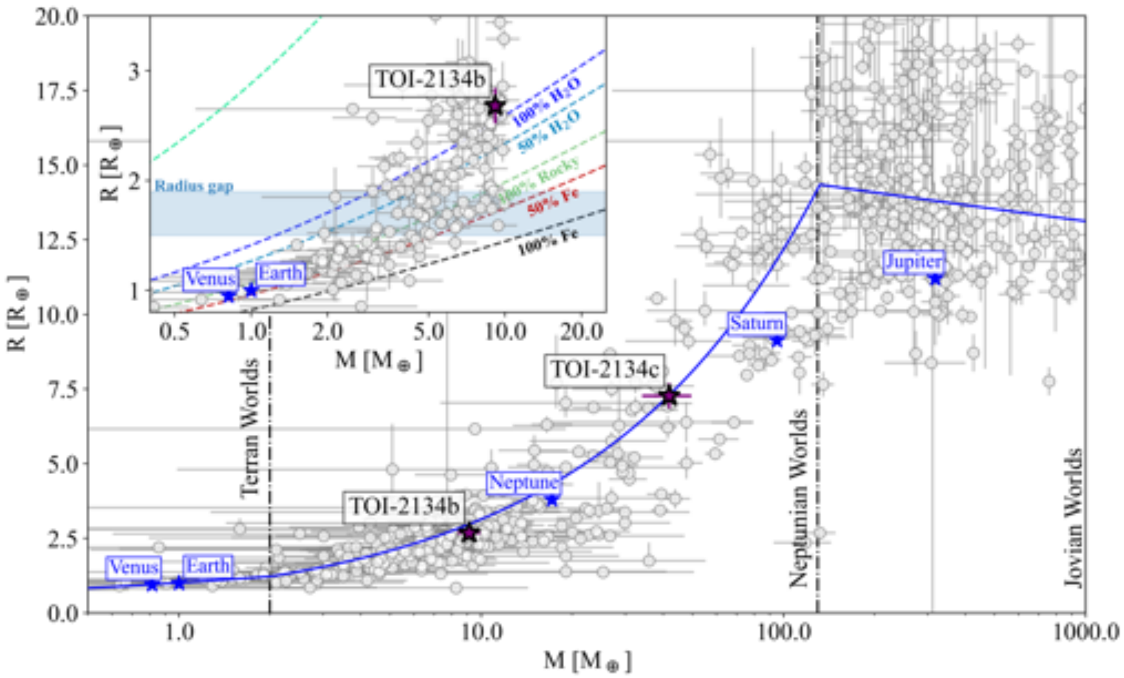

Figure 4.12: Mass-radius diagram with zoom-in for sub-Neptunian planets. The data are taken from the EU Exoplanet catalogue at exoplanet.eu/catalog/ on 17 Feb 2023. The solid blue line shows the mass-radius relation developed by Chen and Kipping (2017), with its categorisation of Terran ($M<2M_\oplus$), Neptunian ($2M_\oplus<M<0.4M_J$) and Jovian worlds ($M>0.4M_J$). The zoomed-in plot includes composition lines taken from Zeng et al. (2016), and the Radius Valley band. Solar system planets are included for scale.

the GP function used to model the stellar variability can be found in the last column of Table 4.3. In Fig. 4.10 I plot the combined SOPHIE and HARPS-N dataset alongside the complete best-fit model in grey, as well as the GP-predicted activity as a black dashed line. Fig. 4.11 shows the phase folded, best-fit Keplerian orbital models, after subtracting the stellar activity-induced signal modelled by the GP, and their residuals.

As a result of the investigation, I established the presence of an inner planet TOI-2134b, and an outer planet TOI-2134c. All derived planetary characteristics are listed in Table 4.5 and 4.6. Fig. 4.12 shows the two planets in a mass-radius diagram.

I computed for the inner planet TOI-2134b a mass M_b of $9.13^{+0.78}_{-0.76}$ $M_\oplus$ and a radius of 2.69±0.16 $R_\oplus$, for an orbital period of 9.2292005±0.0000063 days. Combining mass and radius yielded a bulk density of 0.47±0.09 $\rho_\oplus$. In the mass-radius diagram TOI-2134b falls in a parameter space significantly degenerate in composition. Planet b could be a 100% water-planet (Zeng et al. 2016). At the same time, it could also have a rocky core, a water (or other heavy volatile elements) layer and a low-mass H/He envelope. Overall, it is not possible to distinguish the composition of planet b without additional information.

For more information about the atmospheric characteristics of TOI-2134b see Zhang et al. (2023).

The outer planet TOI-2134c has mass M_c of $41.89^{+7.69}_{-7.83}$ $M_\oplus$ and a radius of 7.27±0.42 $R_\oplus$, with an orbital period of $95.50^{+0.36}_{-0.25}$ days. The derived mass from the RVs and radius from photometry are well-matched and further justify the association of the mono-transit and the detected radial-velocity signal. The presence of a third planet with similar mass to TOI-2134c that could instead explain the transit would have been detected in the radial-velocity. The bulk density of TOI-2134c is calculated to be 0.11±0.03 $\rho_\oplus$ (similar to the density of Saturn). It can therefore be considered a long-orbit mini-Saturn. Given its derived period, I also went back to the other photometric data and computed when transits would have occurred. The derived transit times are plotted in Fig. 4.2 as black dashed lines, and their uncertainty windows as gray shaded regions. TOI-2134c transited 5 times over the 3 years of WASP coverage, but none of those transits was originally detected. The possible explanation for this is twofold. On one hand, WASP is a ground instrument and therefore only observes during dark hours; given the transit duration of ~5 hours, the event could have easily been missed. At the same time, the precision of the WASP data fluctuates significantly and a 0.01 flux deficit (as it is for TOI-2134c) is often too shallow for WASP to reliably detect.

4.1.8.1 System Orbital Stability

As a preliminary test of the stability of the system given the high eccentricity of TOI-2134c, I calculated the radius of the Hill Sphere (Hamilton and Burns 1992) of the outer planet and compared it to the closest approach distance between the two planets. If the orbit of the inner TOI-2134b at any point falls within the Hill Sphere of TOI-2134c, the two bodies are expected to gravitationally interact enough to de-stabilise their orbits. If a body of mass m is orbiting a larger body of mass M at semi-major axis a with an eccentricity e, the Hill Radius R_{Hill} of the smaller body can be approximated to be

$$R_{\text{Hill}} \approx a(1-e)\sqrt[3]{\frac{m}{3M}}. \tag{4.7}$$

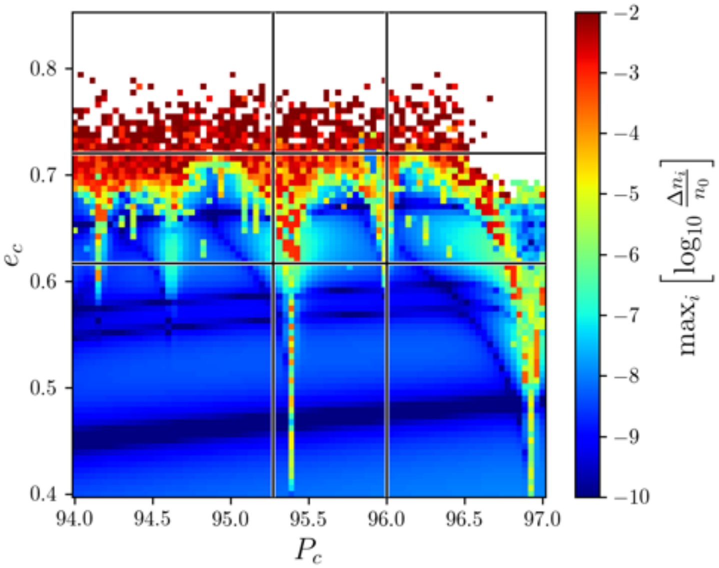

Figure 4.13: Chaos map for the outer planet TOI-2134c. The period P_c and eccentricity e_c are explored on a 81x81 grid of different system configurations. After numerical integrations the NAFF indicator is computed and plotted as a colour-scale. Blue regions correspond to weakly chaotic (more stable) planetary systems, while red areas refer to strongly chaotic (more unstable) systems. The best-fit system position in this space together with its 1σ uncertainties indicate that both stable and unstable solutions are compatible with our high-eccentricity fit. Figure taken from Rescigno et al. (2023b) and produced by M. Stalport.

For planet c, I computed a $R_{\mathrm{Hill,c}}$ of 0.0047±0.0008 AU. The closest approach between the outer and inner planets is 0.048±0.026 AU. Therefore, the orbit of planet b at no point intersects with the Hill Sphere of TOI-2134c.

To further assess the stability of the system under the high eccentricity e_c model, the chaos map in the neighbourhood of the best-fit solution to the high-eccentricity model were also computed, shown in Fig. 4.13. A grid of 81x81 system configurations that vary between each other based on period P_c and eccentricity e_c was created. All other parameters were fixed to their values derived from the MCMC best-fit estimation. Each system defined a unique set of initial conditions that was then used for 50 kyr numerical integrations with `REBOUND`[3] (Rein and Liu 2012) with the 15th order adaptive time-step integrator `IAS15` (Rein and Spiegel 2015). After the simulations, the Numerical Analysis of Fundamental Frequencies (NAFF: Laskar 1990; Laskar 1993) was computed. The NAFF indicator informs about the amount of chaos in a planetary orbit by precisely estimating its main frequency via a technique called frequency analysis (Laskar 1988). The main frequency of

---

[3] REBOUND is an open-source software package dedicated to N-body integrations: http://rebound.readthedocs.org

a planetary orbit corresponds to its mean-motion, which does not drift over time in non-chaotic dynamics, but does drift if the system is chaotic. Therefore, frequency analysis is applied on the two halves of each simulation, and for each planetary orbit, to estimate the amount of drift in the mean-motions. Weakly chaotic (hence stable) orbits should only show small differences in mean motions between the two integration halves. In this work, the NAFF of the system is defined as the logarithmic maximum value of this drift in the form

$$\mathrm{NAFF} = \max_i \left[\log_{10} \frac{\Delta n_i}{n_0}\right], \tag{4.8}$$

in which i refers to the chosen planet, Δn_i is the difference in the mean-motion of planet i between its estimation on the first and second halves of the integrations, and n_0 is the initial mean motion of that planet i. In Fig. 4.13, blue regions have lower NAFF, and are weakly chaotic. Red regions correspond to systems that undergo strong chaos, and likely lead to rapid instability[4]. White regions refer to those systems which had an escape or a close encounter between two bodies, and for which the simulation was stopped. The area of 1σ limit uncertainties on the estimates of P_c and e_c is also overplotted. Inside this square, both chaotic and regular systems can exist. In other words, the high eccentricity model is not incompatible with a stable system.

4.1.8.2 Planetary Incident Flux and Equilibrium Temperature

The incident flux of a planet F_{inc} is computed from stellar luminosity $L_\star$ and planetary semi-major axis a with the following formula:

$$F_{\mathrm{inc}} = \frac{L_\star}{4\pi a^2} = \frac{4\pi R_\star^2\, \sigma_{\mathrm{SB}}\, T_{\mathrm{eff}}^4}{4\pi a^2}, \tag{4.9}$$

where T_{eff} and $R_\star$ are the stellar effective temperature and radius and σ_{SB} is the Stefan-Boltzmann constant. This same incident flux can be expressed in Earth units as

$$\frac{F_{\mathrm{inc}}}{F_{\mathrm{inc},\oplus}} = \left(\frac{T_{\mathrm{eff}}}{T_\odot}\right)^4 \left(\frac{R_\star}{R_\odot}\right)^2 \left(\frac{1}{a}\right)^2, \tag{4.10}$$

---

[4] See Stalport et al. (2022) for details on the link between NAFF and orbital stability.

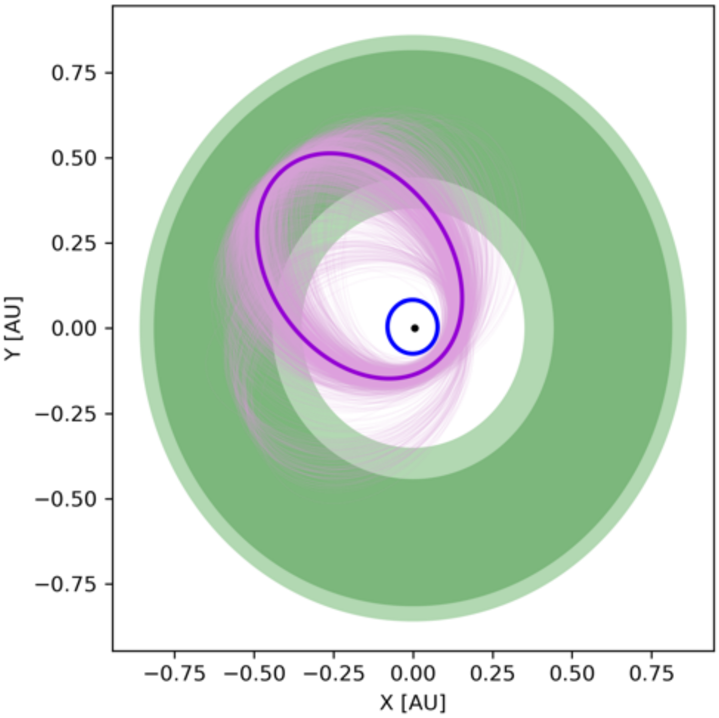

Figure 4.14: Depiction of the configuration of the TOI-2134 system. We include the inner planet with a circular orbit of 9.2292004 ± 0.0000063 days in blue, and the outer planet with an eccentric ($e_c = 0.67^{+0.05}_{-0.06}$) orbit of $95.50^{+0.36}_{-0.25}$ days in purple. Their uncertainties are depicted as lighter orbits. The habitable zone boundaries are indicated as green shaded regions: the empirical HZ is plotted in lighter green, while the narrow HZ is overplotted in darker green. The boundaries are computed as described in Section 4.1.8.2 based on results from Kopparapu et al. (2014).

in which $T_\odot$ and $R_\odot$ are the solar effective temperature and radius and a is expressed in AU. Given semi-major axes a_b and a_c of 0.0780±0.0009 and 0.371±0.004 AU respectively, I computed incident fluxes of 33±2 and 1.4±0.1 $F_{\mathrm{inc},\oplus}$ for planet b and c.

The planets' equilibrium temperatures T_{eq} can be derived as

$$T_{\mathrm{eq}} = T_{\mathrm{eff}}\sqrt{\frac{R_\star}{2a}}[f(1-A_B)]^{1/4}, \tag{4.11}$$

where A_B is the Bond albedo of the considered planet and f represents the effectiveness of atmospheric circulation. Assuming isotropic re-emission and a uniform equilibrium temperature over the entire planet (therefore $f = 1$), an upper limit on T_{eq} can be derived from Equation 4.11 by setting $A_\mathrm{B} = 0$. I, therefore, calculated the upper limit of the equilibrium temperature of planet b to be 666±8 K, and of planet c to be 305±4 K.

From this analysis, the upper limit of the equilibrium temperature of the sub-Saturn object would be compatible with liquid water. Planet c is a gas giant, but could host potentially temperate rocky moons. However, the orbit of TOI-2134c is highly eccentric and the distance of the planet from the star changes significantly during its orbit, as shown in purple in Fig. 4.14. The boundaries of the habitable zone (HZ) of the system, $r_{\mathrm{HZ},\star}$, can be derived from the solar luminosity $L_{\odot}$ and the stellar luminosity as:

$$\frac{L_{\odot}}{r_{\mathrm{HZ},\odot}} = \frac{L_{\star}}{r_{\mathrm{HZ},\star}}, \tag{4.12}$$

where $r_{\mathrm{HZ},\odot}$ is the radius of the boundaries of the solar HZ. The boundaries were here determined following the two models for narrow and empirical habitable zones described in Kopparapu et al. (2014). The narrow HZ is bound by an inner Runaway Greenhouse limit and an outer Maximum Greenhouse limit. The boundaries of the empirical HZ are defined by the Recent Venus and Early Mars limits. The narrow and empirical HZs for the TOI-2134 system are shown in Fig. 4.14 respectively in dark and light green. As Fig. 4.14 clearly shows, TOI-2134c only spends less than half of its orbit within the HZ boundaries. In fact, I also computed the incident flux and upper limit of the equilibrium temperature of planet c at periastron to be 13±4 $F_{\mathrm{inc},\oplus}$ and 533±8 K respectively.

4.1.9 Suggested Follow-Up Observations

4.1.9.1 Long-term RV Observations and Transit Detection for TOI-2134c

This system would foremost benefit from long-term radial-velocity observations to better constrain the period and eccentricity of the outer planet. Both HARPS-N and SOPHIE plan on continuing observing the star sporadically. A second photometric observing campaign aimed at detecting another transit of the outer planet candidate would also be valuable. In the current mission plan, TESS will re-observe TOI-2134 in Sectors 74, 79 and 80 in 2024. A transit of planet c should occur in Sector 80 (late June to early July 2024). Given the brightness of TOI-2134 and the larger radius ratio between planet c and its host star, transits of the outer planet can also be observed with ground-based telescopes. Another firm detection of a transit would re-confirm its period and further inform the eccentricity model choice. I include a list of the times of transit between the original detection and the end of 2025 in Table 4.7. The uncertainties on the times of transit σ_{tr} increase with

Table 4.7: List of times of transits of TOI-2134c between the detected mono-transit and the end of 2025. The uncertainty on the dates computed as shown in Section 4.1.9.1. The transit that should be observed by TESS in Sector 80 is highlighted in bold.

BJD	UT Date (yyyy-mm-dd)	UT Time (hh:mm:ss)
2459814.5±0.3	2022-08-22	23:20:35
2459910.0±0.6	2022-11-26	11:25:12
2460005.5±0.9	2023-03-01	23:29:53
2460101.0±1.2	2023-06-05	11:34:34
2460196.5±1.5	2023-09-08	23:39:11
2460292.0±1.8	2023-12-13	11:43:52
2460387.4±2.1	2024-03-17	23:48:29
2460483.0±2.4	**2024-06-21**	**11:53:10**
2460578.5±2.7	2024-09-24	23:57:50
2460674.0±3.0	2024-12-29	12:02:28
2460769.5±3.3	2025-04-04	00:07:08
2460865.0±3.6	2025-07-08	12:11:46
2460960.5±3.9	2025-10-12	00:16:26

increasing number of "missed" transits as:

$$\sigma_{\mathrm{tr}} = \sqrt{(n\sigma_P)^2 + \sigma_{t_0}} \approx n\sigma_P, \tag{4.13}$$

in which n is the epoch since the observed transit, and σ_P and σ_{t_0} are the uncertainties on respectively the period of the planet and its observed transit time.

4.1.9.2 Rossiter-McLaughlin Analysis

Given the presence of both the inner mini-Neptune and the outer temperate sub-Saturn, TOI-2134 and its planets are scientifically valuable targets for follow-up Rossiter-McLaughlin (RM: Rossiter 1924; McLaughlin 1924) analysis to determine the spin-orbit alignment of the system. The RM amplitude K_{RM} can be computed as

$$K_{\mathrm{RM}} = 52.8\mathrm{ms}^{-1}\frac{v\sin(i)}{5\mathrm{kms}^{-1}}\left(\frac{R_{\mathrm{pl}}}{R_{\mathrm{J}}}\right)^2\left(\frac{R_\star}{R_\odot}\right)^{-2}, \tag{4.14}$$

in which R_{pl} and $R_{\star}$ are the radius of the considered transiting planet and the radius of the star. Instead of using a maximum limit for $v\sin(i)$, I recomputed it starting from the derived stellar rotation period to be 0.78±0.09 $\text{km}\,\text{s}^{-1}$. Since both TOI-2134b and c transit, I calculated the minimum expected RM amplitude for both: $K_{\text{RM,b}} = 0.98 \pm 0.17$ $\text{m}\,\text{s}^{-1}$ and $K_{\text{RM,c}} = 7.2 \pm 1.2$ $\text{m}\,\text{s}^{-1}$. Although the longer transit duration can be an obstacle, RM observations of temperate gas giants as TOI-2134c are valuable to further our understanding of planet migration. A significant fraction of hot giants are shown to have orbits that are misaligned with the rotational axis of their star (Winn et al. 2010; Albrecht et al. 2012). The origin of such misalignment is still unclear, but a leading hypothesis is that high-eccentricity migration tilts the orbit of the planet away from its initial plane via dynamical interactions (e.g., Ford and Rasio 2008; Fabrycky and Tremaine 2007; Petrovich 2015). Unlike hot giants, it is significantly more challenging to form temperate gas planets via high-eccentricity migration (Dong et al. 2013), and it is even less likely in the case of this system due to the presence of an inner small planet. Therefore, if high-eccentricity migration is in fact the driving factor behind the misalignment, the majority of temperate giants should have orbits aligned to spin of their star. However, given their lower transit probabilities, there are only few RM observations of temperate giants. Whether the aim is to observe a whole transit or just the ingress or egress in a shorter summer night, the temperate sub-Saturn planet c has a large peak-to-peak amplitude (7.2 ± 1.3 $\text{m}\,\text{s}^{-1}$) that makes it easily detectable. With a more firmly constrained eccentricity model, TOI-2134c would be a great candidate for RM follow-up.

4.1.9.3 Transmission Spectroscopy

I also assessed the suitability of TOI-2134b and c for follow-up atmospheric characterisation via transmission spectroscopy. Kempton et al. (2018) developed an analytic metric to estimate the expected SNR of transmission-spectroscopy observations based on the strength of the spectral features and the brightness of the star: the Transmission Spectroscopy Metric, or TSM. It can be computed as

$$TSM = \epsilon \cdot \frac{R_{\text{pl}}^3 T_{\text{eq}}}{M_{\text{pl}} R_{\star}^2} \cdot 10^{-m_J/5}, \tag{4.15}$$

in which R_{pl} and M_{pl} are the radius and mass of the considered planet in Earth radii and masses, $R_\star$ is the stellar radius in solar radii, T_{eq} is the equilibrium temperature of the planet computed at zero albedo and full day-night heat redistribution (as in Section 4.1.8.2), and m_{J} is the apparent magnitude of the host star in the J-band. The term ϵ is a normalisation factor to give one-to-one scaling to the JWST/NIRISS 10-hour simulated observations described in Louie et al. (2018). This scaling constant also absorbs the unit conversion factors so that the parameters can be in natural units. ϵ changes depending on the radius of the planet, and is equal to 1.26 for TOI-2134b, and 1.15 for TOI-2134c. I computed a $TSM_{\mathrm{b}} = 172 \pm 42$ and a $TSM_{\mathrm{c}} = 243 \pm 54$. The TSMs of both planets are therefore considered well above the suggested cut-offs for their size bin. It is however important to note that the TSM was developed for targeted JWST effort and therefore it is not optimised for stars with $m_{\mathrm{J}} < 9$ mag, as brighter stars require the bright readout mode and have substantially lower duty cycles. Given its brightness, TOI-2134 is currently only observable without saturation by the JWST with NIRCam in its bright mode, with similar observational strategies as the ones successfully proposed by Dr. Hu for 55 Cancri e (Program ID: 1952) and by Dr. Deming for HD 189733b (Program ID: 1633). However, higher efficiency read modes for JWST observations are being investigated (Batalha et al. 2018) and future dedicated missions such as Ariel, and the ground-based ELTs are suitable for brighter targets such as TOI-2134 (Danielski et al. 2022; Houllé et al. 2021).

4.1.10 Summary and Conclusions

In this work I presented the photometric light curves of five TESS sectors and of three years of WASP monitoring, alongside 219 high-precision radial-velocity measurements obtained with HARPS-N and SOPHIE of the star TOI-2134. After careful periodogram analysis, I performed a transit photometry fit on the photometric data and a Gaussian Process regression analysis on the radial-velocity data to constrain the radii and masses of the planets in the system. To test the statistical strength of the derived models, I also completed a joint analysis of the photometric and the RV data. The resulting planetary parameters fully agree within 1σ uncertainties with the results of the previous investigations. I therefore selected to focus on the results of the less complex, separate analyses for the discussion. As a result, I reached the following conclusions:

- I characterise the new multi-transiting inner TOI-2134b in a 9.2292004±0.0000063 day orbit with M_b = $9.13^{+0.78}_{-0.76}$ $\mathrm{M}_\oplus$ (12σ detection) and R_b = 2.69 ± 0.16 $\mathrm{R}_\oplus$. Its bulk density (ρ_b = 0.47±0.09 $\rho_\oplus$) identifies the planet as either a water-world or a mini-Neptune with a rocky core and a low-mass H/He envelope. I compute the upper limit of the equilibrium temperature of the planet to be 666±8 K.

- I also constrain a second mono-transiting planet TOI-2134c with M_c = $41.89^{+7.69}_{-7.83}\mathrm{M}_\oplus$ (5σ detection) and R_c = 7.27 ± 0.42 $\mathrm{R}_\oplus$ in a $95.50^{+0.36}_{-0.25}$ days orbit, with an upper limit of the equilibrium temperature of 306±4 K. Its bulk density (ρ_c = 0.11±0.03 $\rho_\oplus$) is similar to Saturn's.

- After GP regression, I find three possible orbital architectures for the outer TOI-2134c that model the radial-velocity data, one with low eccentricity ($0.0002^{+0.0025}_{-0.0002}$), one with medium eccentricity (0.45±0.05), and one with high eccentricity ($0.67^{+0.05}_{-0.06}$). While the circular orbit case is disfavoured, the AICc values of the latter two solutions are comparable, and therefore neither can statistically be preferred. I note that in all models the rotation period of the star is half the orbital period of the outer TOI-2134c. I postulate that fitting interactions between the Keplerian model for the planet, and the activity-induced signal that the GP is extrapolating are the reason behind the multiple fully-converged solutions. The flexibility of the GP allows the Keplerian to take different accepted forms while the GP model absorbs the residual signal and attributes it to stellar activity. As described in Section 4.1.6.2, further analysis of the photometry data showed that, given the derived orbital period for planet c, its transit duration time was too short to allow circular orbits. In fact, the mono-transit in the TESS data strongly prefers the high eccentricity case. To further strengthen these results, I also jointly model the photometric and the RV data. This investigation yields a single converged state with an e_c=$0.61^{+0.08}_{-0.03}$. In this work, I therefore present the high-eccentricity model of the separate, less complex RV only analysis and use it for all further studies. I also test the stability of the system given these results and reach the conclusion that the high-eccentricity model is not incompatible with a stable system.

- Since the mass-radius parameter space planet TOI-2134c resides in is not well populated and in order to better constrain its period and eccentricity, I recommend further RV observations and a second photometric observing campaign to detect another transit. To further characterise the architecture of the system, I also recommend Rossiter-McLaughlin follow-up observations. I compute the expected RM amplitude of the temperate sub-Saturn TOI-2134c as $7.2 \pm 1.2\,\mathrm{m\,s^{-1}}$, making it accessible to ground instruments.

- I also compute the Transmission Spectroscopy Metric (TSM) of both planets of the system for possible follow-up atmospheric characterisation via transmission spectroscopy. Although the derived TSMs place the planets well above the recommended cuts, TOI-2134 is close to the bright limits of most instruments on JWST, and is currently only observable with NIRCam in its bright mode. Future missions such as Ariel or ground-based transition spectroscopy will be suited for brighter target such as TOI-2134.

4.2 HD 48948: a Trio of Super-Earth Candidates

4.2.1 Introduction

The primary aim of the Guaranteed Time Observation Program of the stabilised high-precision spectrograph HARPS-N was defined to be the follow-up and mass characterisation of Kepler/K2 (Borucki et al. 2010; Howell et al. 2014) candidates, but was later expanded to include candidates from TESS, and synergies with the CHaracterising ExOPlanets Satellite mission (CHEOPS: Benz et al. 2021). One of the main sub-missions of the HARPS-N GTO is the Rocky Planet Search (RPS). Differently from most of the GTO time, this science mission focuses on the detection of mostly non-transiting, low-mass planets orbiting nearby quiet stars. In order to minimise the magnitude of stellar variability signals, the stars selected for the RPS mission are observed as often as possible, with ideally two observations per night in order to mitigate the granulation signal, and with 15-minute exposure times in order to average-out the effects of p-modes (Dumusque et al. 2011; Chaplin et al. 2019). The first detected system of this program was published in Motalebi et al. (2015): they detected four low-mass planets orbiting the nearby bright star HD 219134. Since then, thanks to a recent upgrade in the Data Reduction Software (DRS) and a renewed effort to understand the instrumental systematics of HARPS-N (Dumusque et al. 2021), multiple other RPS targets have been characterised by the collaboration (e.g., DiTomasso et al. 2023; Stalport et al. 2023; Anna John et al. 2023).

Within this context, in this work I detect three planetary candidates orbiting the nearby bright K-dwarf HD 48948. I summarise the observations used in the analysis in Section 4.2.2, and the characterisation of the host star in Section 4.2.3. To better understand the data employed in the modelling of the signals, I include a brief description of the reduction pipelines used to extract the radial velocities in Section 4.2.4. A Fourier analysis of the data and of the activity indicators (Section 4.2.5), highlights three possible planetary signals. The analysis of the RVs is included in Section 4.2.7. The results are summarised in Section 4.2.8.

4.2.2 Observations

A total of 189 spectra, binned nightly, were obtained with the HARPS-N spectrograph as part of the RPS program. HD 48948 was observed sporadically from 2013-Oct-6 to 2023-Apr-13. After a prilimary analysis of the long-term data, three signals of interests were recovered. During the last observing season, a new intensive observing strategy was implemented to better sample these signals, and thus, roughly 50% of all observations for this target were obtained during this last season. The average signal-to-noise ratio per pixel of the RV observations at 527 nm was 140, and ranged from 110 to 160. Observations were taken with an average exposure time of 15 minutes.

4.2.3 Stellar Characteristics

HD 48948 is a K-dwarf star located at a distance of 52.18 light years (Gaia Collaboration et al. 2023) from Earth. It has a B-V colour of 1.21 mag. The star is moderately active, with a mean $\log(R'_{\rm HK})$ of -4.915 dex. A first estimate of the stellar rotation period $P_{\rm rot}$, was computed to be $48.68^{+7.35}_{-5.68}$ days following the $\log(R'_{\rm HK})$-period relationship of Mamajek and Hillenbrand (2008). To identify the characteristics of the host star, three separate stellar analysis methods were applied: ARES+MOOG (Mortier et al. 2014), SPC (Buchhave et al. 2012; Buchhave et al. 2014) and CCFPams (Malavolta et al. 2017).

Table 4.8: Stellar Parameters for HD 48948. Table taken from Table 1 of Dalal et al. (2024).

Parameter		Value	Source
RA	[h m s]	06 49 57.57	Gaia DR3
DEC	[d m s]	+60 20 14.53	Gaia DR3
Spectral type		K3V	Simbad
B		9.79 ± 0.14	Gaia DR3
V		8.58 ± 0.01	Gaia DR3
J		6.33 ± 0.02	2MASS
H		5.73 ± 0.02	2MASS
K		5.61 ± 0.02	2MASS
$W1$		5.64 ± 0.13	AllWISE
$W2$		5.51 ± 0.06	AllWISE
$W3$		5.56 ± 0.02	AllWISE
Parallax	[mas]	59.393 ± 0.025	Gaia DR3
Distance	[pc]	16.837 ± 0.007	Gaia DR3
T_{eff}	[K]	4593 ± 60	Dalal et al. (2024)
[Fe/H]	[dex]	-0.21 ± 0.03	Dalal et al. (2024)
ξ_t	[km s^{-1}]	0.14 ± 0.03	Dalal et al. (2024)
$v \sin i$	[km s^{-1}]	< 2	Dalal et al. (2024)
$\log g_{\mathrm{spec}}$	[dex]	4.59 ± 0.08	Dalal et al. (2024)
$\log g_{\mathrm{iso}}$	[dex]	4.61 ± 0.01	Dalal et al. (2024)
Mass	[$M_\odot$]	$0.686^{+0.020}_{-0.013}$	Dalal et al. (2024)
Radius	[$R_\odot$]	$0.679^{+0.004}_{-0.004}$	Dalal et al. (2024)
ρ_* [$\rho_\odot$]		$2.18^{+0.069}_{-0.033}$	Dalal et al. (2024)
Age	[Gyr]	$11.48^{+1.93}_{-4.67}$	Dalal et al. (2024)
$\log R'_{\mathrm{HK}}$	[dex]	-4.915	Dalal et al. (2024)
P_{rot}	[days]	$48.68^{+7.35}_{-5.68}$	Dalal et al. (2024)

Gaia DR3 - Gaia Collaboration et al. (2023), Simbad - Grieves et al. (2018), 2MASS - Skrutskie et al. (2006), AllWISE - Wright et al. (2010) and Cutri et al. (2013)

Table 4.9: Table listing the correction applied by each pipeline to produce different RV datasets. Table taken from Table 2 of Dalal et al. (2024).

RV Dataset	YV2	YVA	TWEAKS
Input	Spectrum	Spectrum	CCF
Use	Section 4.2.7.1	Sections 4.2.7.2, 4.2.7.3	Section 4.2.7.4
Magnetic Cycle	Yes	No	Yes
Rotation Modulations	Yes*	No	Yes*
Instrumental Systematic	Yes†	Yes†	Yes†
Reference	C21	C21	A23

*: There might still be residual rotational modulation at the harmonics of stellar rotation periods.

†: The dataset may still exhibit long-term instrumental systematic effects. C21: Cretignier et al. (2021), A23: Anna John et al. (2023)

The results of all three techniques were independently derived, and were all compatible with each other within uncertainties. The final stellar parameters were therefore derived via a weighted average of the results from the three methods, with weights assigned according to their inverse variance. Finally, mass, radius and age of the host star were derived with the code `isochrones` (Morton 2015). The final stellar parameters are listed in Table 4.8. For more information on the specifics of the stellar characterisation method, see Dalal et al. (2024).

4.2.4 Radial Velocity Extraction

In order to better understand the signals in the data and to differentiate between instrumental, stellar and planetary variability, the same set of spectral observations were reduced, and the RVs were extracted with a variety of methods.

Firstly, the data were reduced with the 2.3.5 version of the HARPS-N DRS pipeline. These RVs were extracted employing the CCF technique (Baranne et al. 1996; Pepe et al. 2002) described in Chapter 2 Section 2.1.1.1. The pipeline also estimated the common activity indicators, including the full-width at half-maximum, the contrast, and the bisector span inverse slope (BIS) of the CCF. Additionally, the DRS provided measurements of S-index, Hα, Na, and $\log(R'_{\mathrm{HK}})$ derived from the spectra. A preliminary Fourier analysis highlighted the fact that the DRS RVs still showcased a significant signal with a periodicity at roughly 1,400 days, a pattern of uncorrected instrumental systematics found also in the RVs of other RPS targets (as observed in Anna John et al. 2023). Instead of a basic RV extraction with the DRS pipeline, two specialised post-processing tools were employed independently to correct for these unwanted effects: YARARA (Cretignier et al. 2021; Cretignier et al. 2023) and TWEAKS (Anna John et al. 2023). Both techniques are described in Chapter 2 Sections 2.1.2.4 and 2.1.3.2 respectively.

The radial velocities obtained from the YARARA corrected spectra are referred to as YARARA Version 1 (YV1) RVs. Since all corrections implemented in the pipeline can be reinserted at spectrum-level, a second Version 1 dataset, with re-injected stellar activity correction, was also produced. This dataset was called YARARA Version 1 Activity (YVA), and is expected to only include planetary signals and stellar variability. Further corrections on the radial velocities can be implemented using shell time-series coefficients (Cretignier et al. 2023) and principal component analysis. In the case of HD 48948, only six basis vectors are considered: three related to the shell components and three associated with instrumental-related effects. The limited SNR of the data in fact prevents optimal extraction of time-domain vectors by PCA. These six basis vectors were then used to extract the new RVs starting from the YV1 radial velocity time series. These data were called the YARARA Version 2 (YV2) RVs. The YV2 dataset is preferred over the YV1 time series because the this approach was proven to outperform the original YARARA formalism in the correction of systematics and stellar variability, especially in the case of large datasets such the HD 48948 one. However, differently than for YARARA Version 2 data, not all corrections are applied contemporaneously on the spectra and therefore the equivalent YARARA Version 1 Activity re-injected time series cannot be produced.

Finally, a final set of radial velocities was computed using the TWEAKS pipeline (Anna John et al. 2022; Anna John et al. 2023). All different datsets used in this work, alongside their applied corrections are listed in Table 4.9. The analyses undertaken in this work only employ the three datasets plotted in Fig. 4.15: YV2, YVA and TWEAKS.

4.2.5 Periodogram Analysis

In order to investigate the possible planetary signals present in the stellar-activity corrected YV2 dataset, an extensive periodogram analysis was completed. In summary, two main signals were are identified above the 0.01% False Alarm Probability level at periods

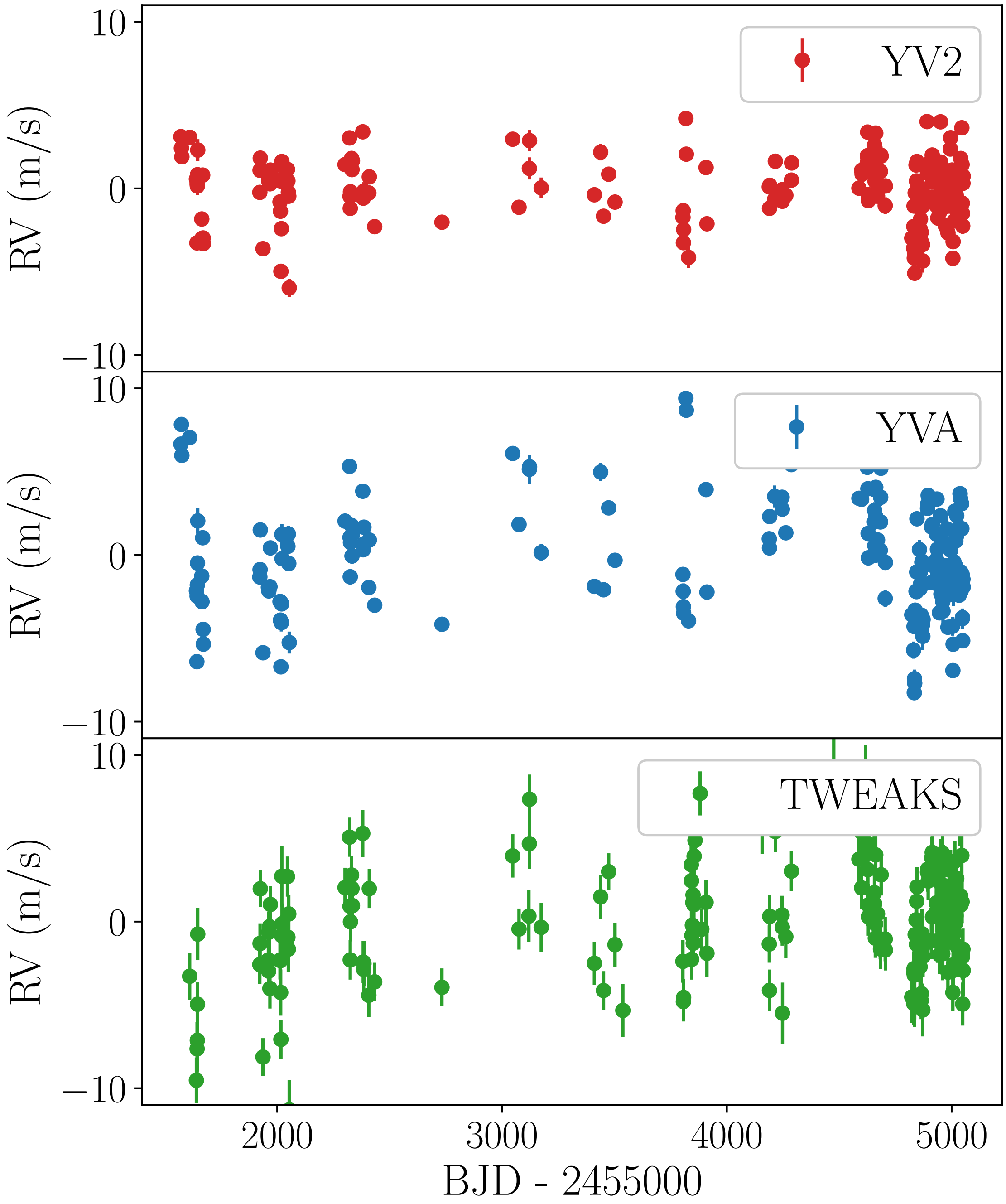

Figure 4.15: Plot of the three RV time series analysed in this work. The YV2 RVs are represented by solid red circles, the YVA RVs by solid blue circles, and the TWEAKS RVs by solid green circles. Uncertainties are plotted as errorbars, but may be too small to be visible. Figure taken from Fig. 2 of Dalal et al. (2024).

of 7.3 days and 38 days, with a third prominent peak arising after whitening at roughly 151 days, as shown in Fig. 4.16. A forest of peaks was present in the periodogram of the residual RVs after removing the three signals of interest at around 20 to 22 days, which was consistent with first harmonic of the stellar rotation period.

In order to assess whether the three recovered signals could possibly be attributed to planet candidates, the GLS periodograms of the derived activity indicators were also computed, as shown in Fig. 4.17, since signals recovered in both the RVs and the activity proxies are more likely to be generated by stellar activity, instead of being of Keplerian nature. The most significant peak for all activity indicators was around 42 to 43 days, as highlighted by the orange shaded region. No power was retrieved at the periods of two of the planet candidates at 7.3 and 151 days. The middle planet, at roughly 38 days was harder to confirm, as its period was very close to the main stellar rotation peak. For this reason, further model testing to assess the presence of this planet candidate was undertaken in Section 4.2.7.

4.2.6 Gaussian Process Regression for Stellar Period Detection

In order to more precisely isolate the rotational period of HD 48948, I completed a Gaussian process regression analysis on the activity indicator that best correlated to the stellar activity time series: the $\log(R'_{\mathrm{HK}})$. The signal in the activity proxy was modelled with the MAGPʏ-RV pipeline with a GP function described by a Quasi-periodic kernel with added jitter to the diagonal of the covariance matrix, as defined in Eq. 4.2 in Section 4.1.6. The best-fit kernel hyperparameters were identified via affine invariant MCMC optimisation. I evolved 2,000 chains over 1,000 iterations each, discarding a burn-in phase of 200 steps. All hyperparameters were constrained with forced positive uniform priors. Additionally, I bound the harmonic complexity to be between 0 and 1, the stellar rotation period to range from 0 to 60 days, and the jitter to vary between 0 and the highest uncertainty value of the $\log(R'_{\mathrm{HK}})$ time series.

Fig. 4.18 shows the posterior distribution of the GP period. The results of the GP analysis pointed to a stellar rotation period of $41.8^{+2.2}_{-1.5}$ days. The evolution timescale for the $\log(R'_{\mathrm{HK}})$ was not well-defined, resulting in wide uncertainties, as 171^{+128}_{-66} days. The

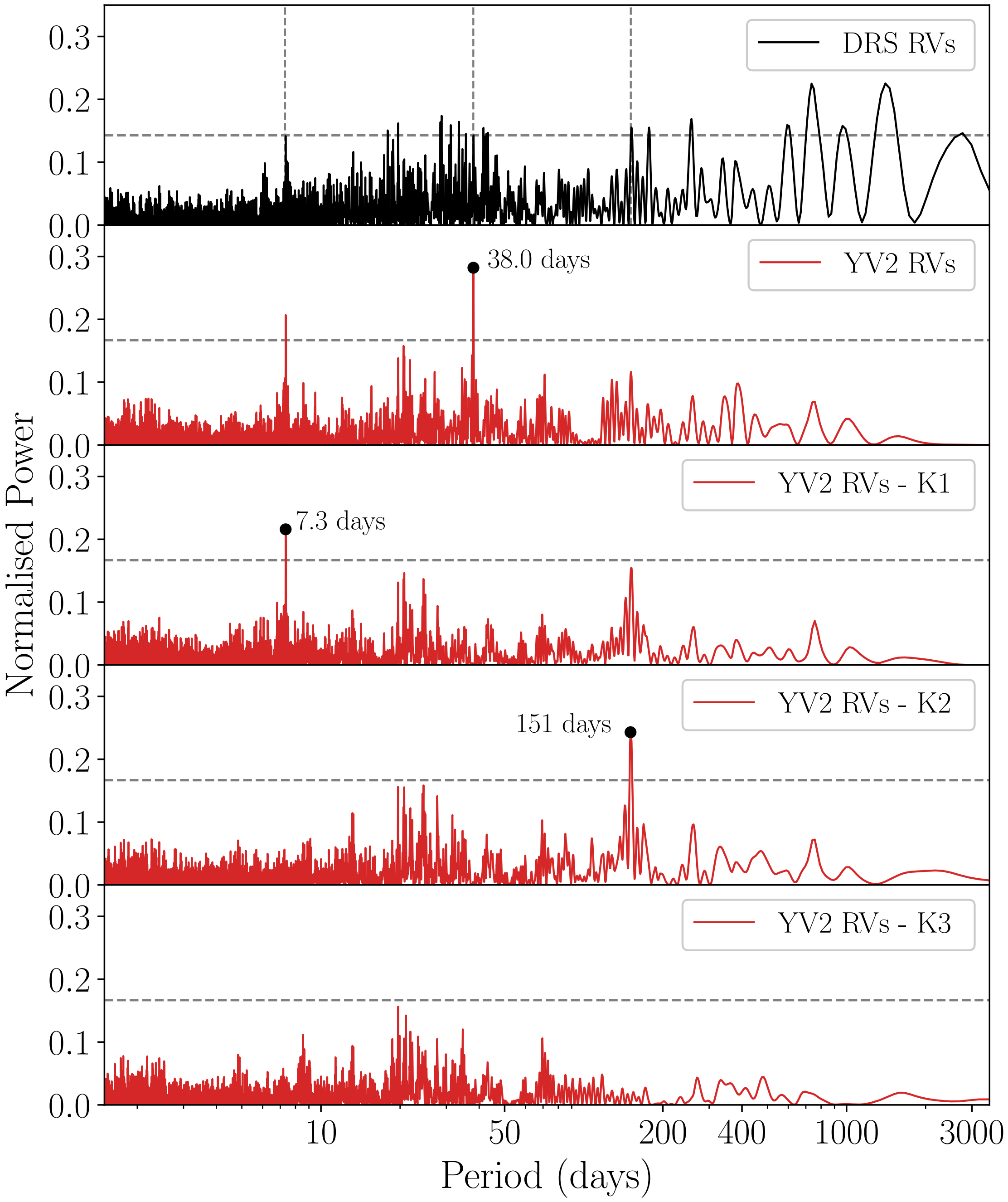

Figure 4.16: *Top panel:* The GLS periodogram of the DRS (uncorrected for activity) RVs of HD 48948 is plotted in black. The 0.01% False Alarm Probability is indicated by a horizontal dashed line. Vertical dashed lines identify the orbital periods of the three planetary candidates. Panels 2-5: GLS periodograms of the YV2 RVs. The most significant signal, identified by a label, is iteratively subtracted from the radial velocities, and the periodogram of the remaining variability is plotted until no signal surpasses the 0.01% False Alarm Probability level plotted as horizontal dashed lines. The dominant Keplerian peaks are found at 38 days, 7.3 days and at 151 days. Figure taken from Fig. 3 of Dalal et al. (2024).

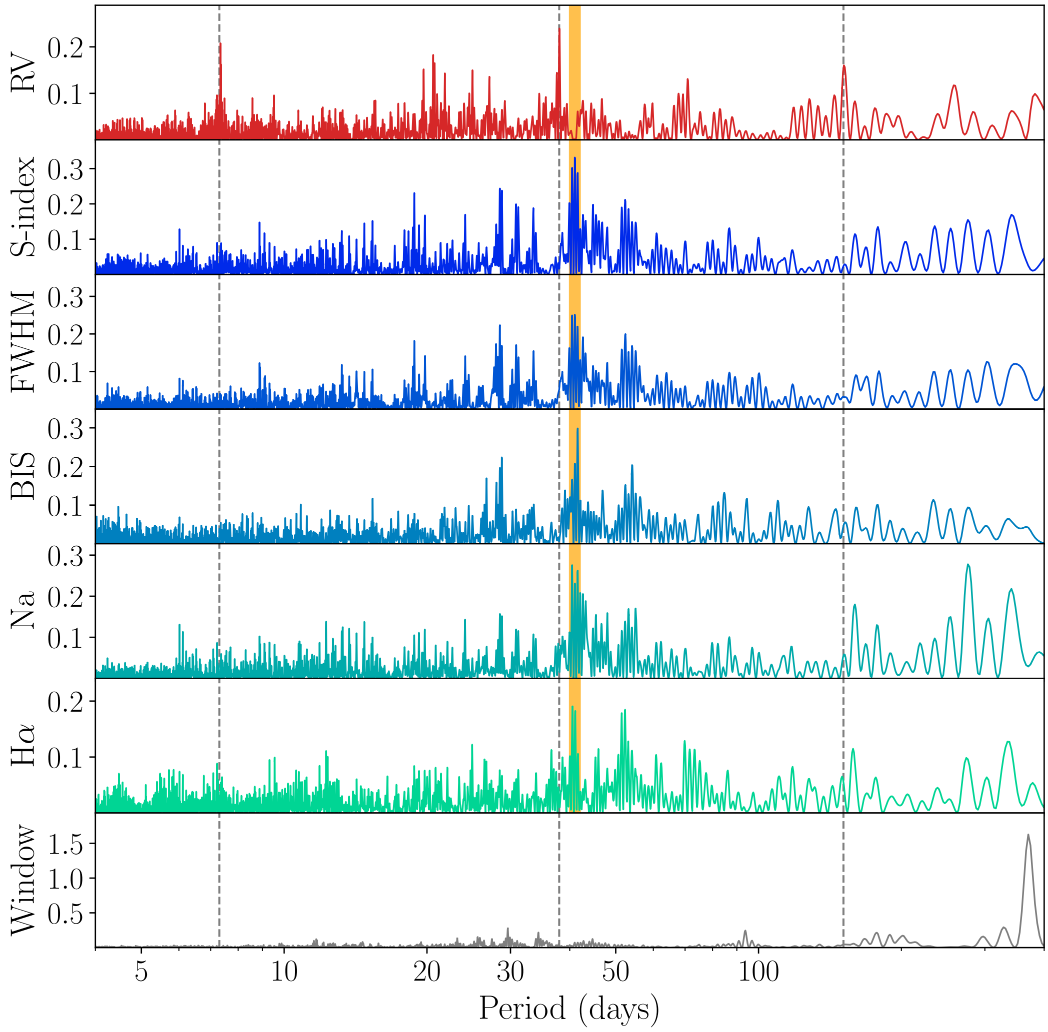

Figure 4.17: The GLS periodograms of, in order, the YV2 RVs, the S-index, the FWHM, the bisector spane (BIS), the Nα and the Hα indexes are plottted. The window function of the epochs is also included in the bottom panel. The orbital periods of the three planetary candidates are identified by black dashed lines at 7.3, 38, and 151 days. The most significant peak shared by all indicators is highlighted by the orange band at roughly 42 days. Figure taken from Fig. 6 of Dalal et al. (2024).

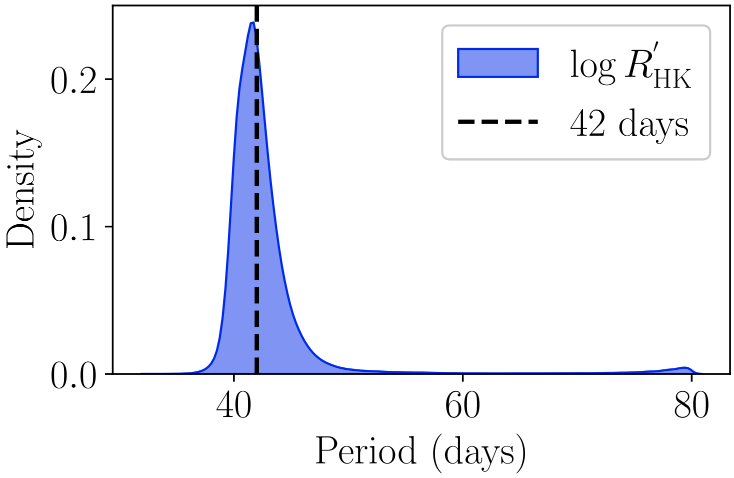

Figure 4.18: Posterior distribution of the period of the Quasi-periodic kernel after GP regression of the $\log(R'_{\mathrm{HK}})$ time series. Figure taken from Fig. 8 of Dalal et al. (2024).

harmonic complexity was found to be 0.73 ± 0.19. Overall, the GP best-fit covariance function for $\log(R'_{\mathrm{HK}})$ agreed with the results obtained with the periodogram analysis, confirming the stellar rotation period to be roughly equal to ~ 42 days.

4.2.7 Radial Velocity Analysis

The YV2 radial velocities represent the best activity-corrected dataset extracted in this work, and are meant to only be modulated by the presence of planets. Thus, they could be modelled simply as the sum of Keplerian signals. However, it is important to underline that the corrections applied by YARARA are not fully understood, and are not based on physically-motivated models. It is therefore inadvisable to solely rely on the analysis of the YV2 dataset. In this Section, multiple RV time series were in fact analysed with different methods, in order to more robustly characterise the planetary system. This was especially relevant for HD 48948 due to the complexity of the stellar signal and the fact that the predicted orbital period of the middle 38-day planet candidate falls on one of the closest yearly aliases of the derived stellar rotation. In this work, I modelled the YV2 data as the sum of Keplerian oscillations (addressed in Section 4.2.7.1), and the YVA time series with a one-dimensional Gaussian process (shown in Section 4.2.7.2). Although only briefly described in this thesis, the YVA RVs were also modelled simultaneously with their activity indicators in a multi-dimensional GP approach. TWEAKS was also employed for a further independent analysis to confirm the masses derived with the previous methods.

In order to ascertain the presence of the middle planet with orbital period close to the stellar rotation, I examined three potential system configurations: one with three Keplerians, each having orbital periods of 7.3, 38, and 151 days, and two models with two-Keplerian setups, with periods at 7.3 and 38 days, and at 7.3 and 151 days, respectively. As described in Section 4.1.6.2, I compared the different models by comparing their derived corrected Akaike Information Criterion, known as AICc (Sugiura 1978; Akaike 1983). In this analysis, I also calculated the percentage likelihood, Φ, as

$$\Phi(\mathrm{AICc}, \mathrm{AICc}_{\mathrm{ref}}) = 100 \times \left[1 - \exp\left(-\frac{\mathrm{AICc} - \mathrm{AICc}_{\mathrm{ref}}}{2}\right)\right]. \tag{4.16}$$

This percentage quantifies the relative support for the model under consideration (AICc) compared to the reference model ($\mathrm{AICc}_{\mathrm{ref}}$). Higher values indicates stronger evidence in favour of the considered model. I defined there being enough evidence to prefer one model over another when a 95% confidence level was achieved, i.e., when $\mathrm{AICc} > \mathrm{AICc}_{\mathrm{ref}} + 6$ or $\Phi > 95$. Differences in AICc less than 4.5 (less than 90% confidence) are typically insufficient to reliably differentiate between models. Ideally a difference of $\mathrm{AICc} > \mathrm{AICc}_{\mathrm{ref}} + 10$ was preferred for a firm statistical model choice.

4.2.7.1 Keplerian-only Analysis

I began the analysis by assuming that the YV2 dataset had been fully and successfully cleaned of all stellar activity signals, and it only mapped the coherent motion due to gravitational interactions between the star and the planets in the system. Consequently, I modelled the YV2 radial-velocity dataset with MAGPy-RV by summing Keplerian signals, with each representing an individual planet. As for Section 4.1.6, I re-parametrised eccentricities, e, and arguments of periastron, ω, in the form $\sqrt{e}\cos\omega$ and $\sqrt{e}\sin\omega$ (Eastman et al. 2013). In terms of prior information, I applied uniform priors to each planetary orbital period based on the results of the periodogram analysis, with the width of each prior related to the width of the corresponding peak such that wider peaks corresponded to wider priors. Additionally, I imposed a wide uniform prior equivalent to its preliminary orbital period on the time of periastron for the innermost planet, in order to prevent multiple global maxima results. All other parameters were bound by wide positive uniform priors. For each of the three system configurations, I evolved 700 chains for 5,000 iterations each, with the first 1,000 steps disregarded as burn-in. I also computed the likelihood and AICc values for each model, as presented in Table 4.10. It was evident from both the likelihood values and the AICcs that the three-Keplerian model was strongly preferred. The RV semi-amplitudes of the three planet candidates were extracted to be $1.35^{+0.18}_{-0.17}$ m s^{-1} (7σ) for Keplerian b, $1.71^{+0.20}_{-0.22}$ m s^{-1} (8σ) for Keplerian c, and 1.09 ± 0.19 m s^{-1} (6σ) for Keplerian d.

The semi-amplitudes calculated with this method were smaller than expected when compared to the results of the other analyses, in particular the one for planet d, leading to smaller overall masses. YARARA's primary focus is the unveiling of signals obscured by effects such as stellar variability and instrumental systematics. It is very successful as a preliminary test, but in some cases YARARA tends to overcorrect. For an accurate mass determination, information on the oscillations induced by planetary Doppler-shifts should thus be included in the RV extraction and spectra correction. This analysis step was first developed for correcting for the presence of binaries or very large planetary signals (Cretignier et al. 2021), but it has since been extended to include smaller planetary variability. Therefore, the preliminary orbital periods of the three Keplerians were included in the pre-processing of the observational spectra, and the YV2 RVs were re-computed. I then re-produced the same analysis as described in this Section with the newly derived YV2 radial velocities. After MCMC parameter optimisation, I extracted the orbital parameters of the three planet candidates. The periods of the three Keplerians were derived to be to be $7.34023^{+0.00067}_{-0.00065}$, $39.003^{+0.020}_{-0.017}$ and $150.83^{+0.42}_{-0.37}$ days, and all planet orbits were computed to be circular. Their RV semi-amplitudes were found to be $1.85^{+0.18}_{-0.19}$ m s^{-1} (10σ) for Keplerian b, $2.12^{+0.26}_{-0.27}$ m s^{-1} (8σ) for Keplerian c, and 1.74 ± 0.22 m s^{-1} (8σ) for Keplerian d. These semi-amplitudes were in better agreement with all other results, and the new YARARA corrections seemed to better preserve the planetary signals. I thus derived minimum masses of 4.33 ± 0.46 $M_\oplus$, 8.55 ± 0.97 $M_\oplus$, 10.91 ± 1.39 $M_\oplus$ for b, c and d respectively. The results of this analysis are detailed in Table 4.11, under the section "Three Keplerians-only Analysis".

Table 4.10: Likelihood and AICc comparison between the tested models (3 Keplerians bcd, 2 Keplerians bc, 2 Keplerians bd) in Sections 4.2.7.1, 4.2.7.2. The results of Section 4.2.7.3 are not included. The Table can be found in its entirety as Table 5 in Dalal et al. (2024). The models highlighted in bold are the most favoured in each analysis based on their likelihood and AICc. Table taken from Table 5 of Dalal et al. (2024).

Parameter	3 Keplerians bcd	2 Keplerians bc	2 Keplerians bd
	Keplerian-Only Analysis		
$\ln\mathcal{L}$	**-379.89**	−482.17	−550.62
AICc	**792.55**	985.58	1122.48
	One-dimensional GP Analysis		
$\ln\mathcal{L}$	**-350.95**	−361.28	−352.66
AICc	746.90	755.33	**738.07**

4.2.7.2 Gaussian Process Regression Analysis

I modelled the YVA radial velocities as a combination of multiple planetary signals, in the form of Keplerians, and stellar variability, in the form of a GP function described by a Quasi-periodic kernel which also included a white noise component as defined by Eq. 4.2.

I once again tested three distinct models combined with a stellar activity prediction in order to better assess the structure of the system. Initially, I modelled the RVs as the GP function added to three Keplerians, namely b, c, and d, with respective periods of approximately 7, 38, and 151 days. To further evaluate the presence of the candidate d and to validate the mass and orbital period of the 38-day planet, I also modelled the signal with a GP and only two Keplerians, including only planets b and c, and only planets b and d, respectively.

I used the same pipeline and a similar MCMC structure as in Section 4.2.7.1, with 700 chains over 5,000 iterations each and a discarded 1,000-steps burn-in. I imposed priors derived from both the periodogram and the activity proxy GP analysis. Starting with the kernel hyperparameters, I applied an informed uniform prior between 40 and 47 days on the stellar rotation period, P_{rot}, primarily to prevent any signal "bleeding" between the modelled stellar activity and the Keplerian c. In other words, I aimed to prevent the GP from modelling the planetary signal instead of the stellar activity, or from absorbing part of the periodic signal at the orbital frequency of the planet candidate yielding a lower planetary mass than expected. The harmonic complexity h was constrained with a uniform prior to be between $[0, 1]$. I also bound the jitter term with a uniform prior between $[0, 2]$ $\mathrm{m\,s^{-1}}$. All prior information for the GP is also included in Table 4.13.

Table 4.11: The table displays the results of three Keplerian models, as discussed in Sections 4.2.7.1, 4.2.7.2. Results from the analyses in Sections 4.2.7.3 and 4.2.7.4 are not included. The table can be seen in its entirety in Dalal et al. (2024) as Table 3. The values and uncertainties of each parameter are derived from their respective posterior distributions, with the 50th percentile representing the central value and the 16th to 84th percentile range indicating one-sigma uncertainty. Additionally, the derived parameters are also included in this table. The maximum equilibrium temperature is calculated assuming an albedo (A_B) of 0 and isotropic re-emission and uniform equilibrium temperature.

Parameter	Symbol	Unit	HD48948 b	HD48948 c	HD48948 d
Three Keplerians-only Analysis on YV2 RVs					
Orbital period	P	days	$7.34023^{+0.00067}_{-0.00065}$	$39.003^{+0.020}_{-0.017}$	$150.83^{+0.42}_{-0.37}$
RV Amplitude	K	$\mathrm{m\,s^{-1}}$	$1.85^{+0.18}_{-0.19}$	$2.12^{+0.26}_{-0.27}$	1.74 ± 0.22
Eccentricity	e		$0.08^{+0.09}_{-0.07}$	$0.08^{+0.06}_{-0.07}$	$0.14^{+0.13}_{-0.12}$
Argument of periastron	ω	rad	$1.16^{+1.27}_{-1.28}$	$-1.01^{+1.04}_{-1.01}$	$-0.95^{+0.92}_{-0.56}$
Time of periastron	t_p	BJD	$2456574.23^{+1.27}_{-1.44}$	$2456570.81^{+5.80}_{-3.03}$	$2456568.87^{+17.67}_{-23.89}$
Minimum Mass	$m\sin(i)$	$\mathrm{M}_\oplus$	4.33 ± 0.46	8.55 ± 0.97	10.91 ± 1.39
One-dimensional GP Analysis on YVA RVs					
Orbital period	P	days	$7.34012^{+0.00052}_{-0.00046}$	$37.99^{+0.16}_{-0.22}$	150.28 ± 0.69
RV Amplitude	K	$\mathrm{m\,s^{-1}}$	$2.07^{+0.15}_{-0.17}$	1.06 ± 0.69	$1.55^{+0.35}_{-0.38}$
Eccentricity	e		$0.08^{+0.06}_{-0.05}$	$0.23^{+0.14}_{-0.19}$	$0.15^{+0.17}_{-0.10}$
Argument of periastron	ω	rad	$0.55^{+0.95}_{-0.92}$	$-0.48^{+1.81}_{-1.40}$	$-1.05{-}0.79^{+0.96}$
Time of periastron	t_p	BJD	$2456573.74^{+1.05}_{87}$	$2456571.71^{+10.5}_{-9.80}$	2456571.52 ± 15.3
Minimum Mass	$m\sin(i)$	$\mathrm{M}_\oplus$	$4.86^{+0.35}_{-0.39}$	$3.94^{+3.02}_{-2.85}$	$9.72^{+2.26}_{-2.56}$
Derived Parameters					
Semi-major axis	a	AU	0.0652 ± 0.0005	0.1951 ± 0.0016	0.4894 ± 0.0042
Semi-major axis	a	arcsec	0.0039	0.0116	0.0291
Scaled semi-major axis	$a/R_\star$		20.65 ± 0.21	61.79 ± 0.62	154.97 ± 1.61
Incident Flux	F_inc	$F_{\mathrm{inc},\oplus}$	138.85 ± 6.98	15.50 ± 0.78	2.46 ± 0.12
Equilibrium Temperature	T_eq	K	957 ± 11	553 ± 6	349 ± 4

Finally, I computed the likelihood and the AICc for the three best-fit models, as shown in Table 4.10. Overall, although the two-Keplerian configuration including only planets b and c could be rejected as disfavoured, no strong statistical preference between the other models could be found. The logarithm of the likelihood function, $\ln \mathcal{L}$, showed a preference for the model with three Keplerians, reinforcing the long-period signal as a planetary candidate. However, the difference between the AICcs (which balance the goodness of fit to the complexity of the model) of the two remaining options was less than 4, which made a statistical differentiation between the two hard to reach. The slight preference of the AICc for the two-Keplerian (b and d) model was interpreted as follows: the flexibility of the GP allowed it to absorb the signal of the middle planet in the stellar activity model and to describe the combination of the two variabilities with a single covariance function (Damasso et al. 2018). Each extra Keplerian introduced five new parameters to the model. Thus assuming comparable fits to the data provided, the model without these five extra terms will always be favoured. In this case, instead of being an assessment of the goodness of the fit, the AICc is simply acting as an Occam Razor. Therefore further analyses were required to fully assess the strength of this middle signal, and to confirm whether the lack of statistical preference was in fact due to the flexibility of the GP model.

4.2.7.2.1 GP Mock Data Tests I further evaluated whether it was appropriate to directly compare likelihoods and AICc values for models in scenarios where a planet's orbital period is very similar to the period of the covariance of the GP function, or instead caution should be exercised. To accomplish this, I derived the activity signal removed by the YARARA pipeline (i.e using YVA - YV1 RVs), and then injected a Keplerian signal with a 38-day orbital period. Initially, I set the Keplerian semi-amplitude to match the results obtained from the one-dimensional GP analysis for the three-Keplerian model (with a value of $Kc = 1.06 \ \mathrm{m\,s^{-1}}$). I then also examined the extreme cases by testing semi-amplitude values of $1.75 \ \mathrm{m\,s^{-1}}$ and $0.38 \ \mathrm{m\,s^{-1}}$, derived from the 1σ uncertainties on K_c. I computed the best fit functions to these radial velocities using two distinct models: a GP+Keplerian model that incorporated both an activity term (with a QP kernel) and a Keplerian signal with a 38-day period (as expected in the generated dataset), and a GP-only model that included only the activity component. All three runs that included the Keplerian signal

Table 4.12: Likelihood comparison between the tested models in the injected Keplerian retrieval analysis described in Section 4.2.7.2. The runs are defined by the amplitude of the injected signal. Table taken from Table A1 from the Appendix of Dalal et al. (2024).

	$0.38\ \mathrm{m\,s^{-1}}$	$1.06\ \mathrm{m\,s^{-1}}$	$1.75\ \mathrm{m\,s^{-1}}$
	$\ln \mathcal{L}$		
GP + Keplerian	**-282.99**	**-280.66**	**-279.84**
GP only	−284.49	−289.27	−295.59
	AICc		
GP + Keplerian	587.21	**582.55**	**580.91**
GP only	**579.30**	588.87	601.51

successfully recovered the parameters of the injected planetary oscillations. The semi-amplitudes of the Keplerians for the three cases were determined to be $0.47^{+0.37}_{-0.32}$, $1.12^{+0.33}_{-0.45}$, and $1.87^{+0.28}_{-0.31}\ \mathrm{m\,s^{-1}}$ respectively, and were all consistent within 1σ with the injected signals (0.38, 1.06, and $1.75\ \mathrm{m\,s^{-1}}$). Additionally, the hyperparameters of the GP kernel, including period and amplitude, were found to be consistent within 1σ uncertainties to those recovered in the previous GP analysis, confirming a correct modelling of the activity signal.

I then examined the likelihoods and AICc values for all models in all semi-amplitude cases, detailed in Table 4.12. When comparing runs with and without the Keplerian model, I observed that the AICc strongly favoured the GP-only approach in the case with the lowest injected semi-amplitude of $0.38\ \mathrm{m\,s^{-1}}$, with a difference larger than 10. This means that at the lower bound of the uncertainties retrieved for the semi-amplitude of planet c, a one-dimensional GP analysis would significantly prefer a simpler planet-less model over the “truth" (as we know that a planet is in fact present). For the intermediate semi-amplitude level ($1.06\ \mathrm{m\,s^{-1}}$), I could distinguish between the GP-only and GP+Keplerian models with 95% confidence, but the difference between the two values is still under the preferred 10. Finally, I find that the GP+Keplerian model was more suitable in the case with the highest semi-amplitude of $1.75\ \mathrm{m\,s^{-1}}$, where the planetary signal was most prominent. Only at the upper level of uncertainties I was therefore able to strongly prefer the “correct" model based only on statistical reasons. These findings indicated that when Keplerian orbital periods closely match the activity period, a GP analysis can effectively absorb the Keplerian signal within the activity function (Damasso et al. 2018). This explained the slight preference for the two-Keplerian (b and d) system architecture over the three-planet model

Table 4.13: Stellar activity priors and results from the Gaussian Process regression analyses undertaken in Sections 4.2.7.2 and 4.2.7.3 (upper section and lower section). Priors have been identified as follows: $\mathcal{U}$ is a uniform prior. Table adapted from Table 4 of Dalal et al. (2024).

Parameter	Unit	Prior	Values
One-dimensional GP Analysis			
GP Amplitude A	$\mathrm{m\,s^{-1}}$	$\mathcal{U}[0,20]$	$3.06^{+0.52}_{-0.36}$
GP Timescale l	days	$\mathcal{U}[0,+\infty]$	$110.84^{+29.53}_{-23.23}$
GP Period P_{rot}	days	$\mathcal{U}[40,47]$	$42.87^{+0.99}_{-1.02}$
GP Smoothness h		$\mathcal{U}[0,1]$	$0.59^{+0.10}_{-0.09}$
Jitter β	$\mathrm{m\,s^{-1}}$	$\mathcal{U}[0,2]$	$0.84^{+0.11}_{-0.09}$

found in Section 4.2.7.2. It is therefore important to exercise caution and take into account the astrophysical context when interpreting ΔAICc values in such situations.

Finally, taking into consideration these findings, alongside the periodogram analysis undertaken in Section 4.2.5, I chose to present the three-Keplerian model results. The final parameters for the three planet candidates and the GP kernel are included in Tables 4.11 and 4.13, respectively. The orbital periods of all three planetary signals are found to be $7.34012^{+0.00052}_{-0.00046}$, $37.99^{+0.16}_{-0.22}$ and 150.28 ± 0.69 days and their eccentricities were close to 0. The radial velocity semi-amplitudes for Keplerians b, c, and d were constrained to 13σ, 2σ and 4σ levels respectively. By modelling the activity with a GP, the detection of the innermost planetary candidate is significantly improved with respect to the YV2 analysis, but the constraints on the outer two planets are not as robust. The increased uncertainties in the semi-amplitudes are reflected in the uncertainties of their masses. The GP amplitude is determined to be $3.06^{+0.52}_{-0.36}$ $\mathrm{m\,s^{-1}}$, with a period of $42.87^{+0.99}_{-1.02}$ days. These results are plotted in Fig. 4.19, and the phase-folded RVs for all Keplerians can be found in Fig. 4.20.

4.2.7.3 Multi-dimensional GP analysis

The YVA RV data were also modelled with a multi-dimensional GP (described in Chapter 2 Section 2.3.4.2) simultaneously to the BIS and the $\log(R'_{\mathrm{HK}})$ with the code PyORBIT[5] (Malavolta et al. 2016; Malavolta et al. 2018) through the library TINYGP[6] for improved performances (Nardiello et al. 2022; Mantovan et al. 2024). For a full description of the

---

[5] Code available at: https://github.com/LucaMalavolta/PyORBIT, version 10

[6] Code available at: https://github.com/dfm/tinygp

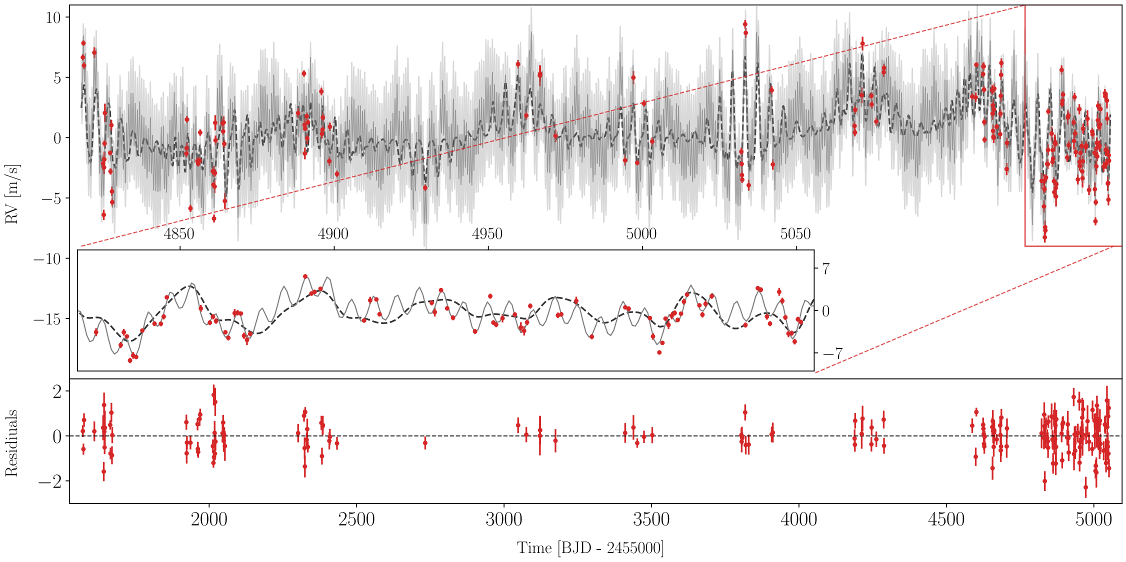

Figure 4.19: *Top Panel:* Complete model derived after one-dimensional GP regression of the HD 48948 data computed as the sum of three Keplerian signals and a GP described by a QP kernel shown as a gray solid line. The gray shaded areas represent the model's uncertainties. The RV time series is plotted in red. Uncertainties on the measurements are included as red errorbars but may be too small to be visible. A zoomed-in version of the last semester of data is also provided, with the GP function describing the stellar variability is included as a black dashed line. *Bottom Panel:* The residuals, or the differences between the radial velocities and the model (grey line), are plotted alongside their uncertainties. Figure taken from Fig. 9 of Dalal et al. (2024).

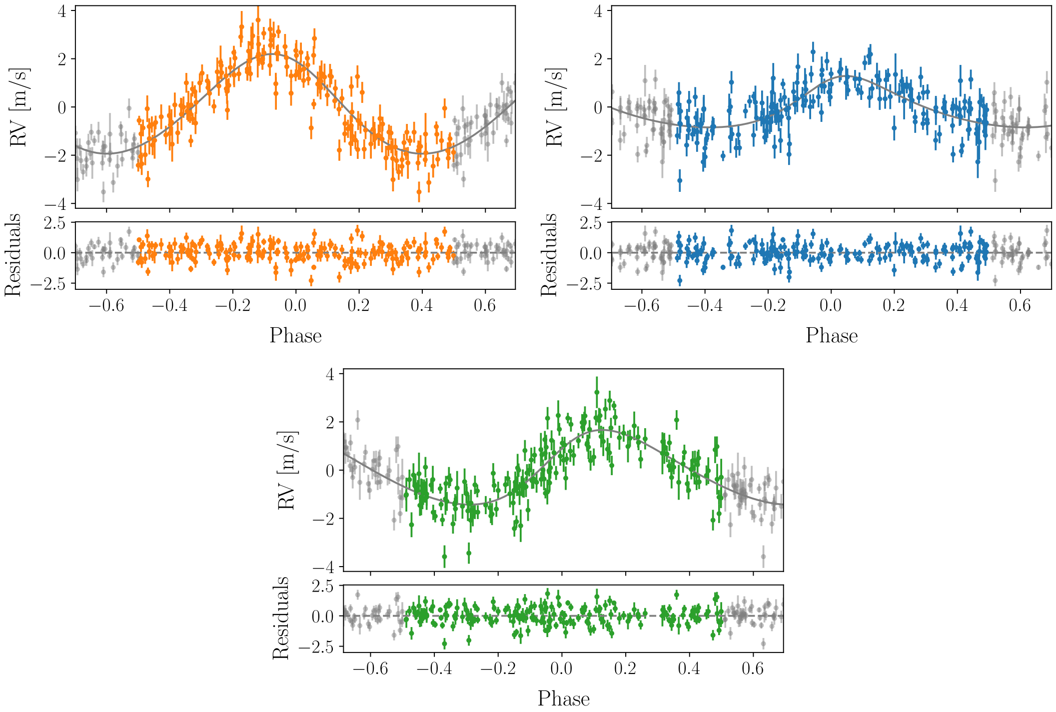

Figure 4.20: Phase foldes plots of the three planetary signals identified by the one-dimensional GP analysis. The radial velocities of each planet are plotted in orange, blue and green for the planets at periods of 7.3, 38, and 151 days respectively. The best fit Keplerian model for each planetary candidate is plotted as a gray line. The phase has been extended in each plot. Figure taken from Fig. 10 of Dalal et al. (2024).

analysis, see Dalal et al. (2024). Multi-dimensional GPs can be particularly useful in cases such as this, as modelling the stellar activity simultaneously across multiple time series prevents the GP from absorbing the signal of a planet close to the stellar rotation period (e.g., Rajpaul et al. 2021). To further test this, the prior boundaries of the rotation period of the star were deliberately extended to include the orbital period of Keplerian c. As for the previous Sections, three system architectures were compared. In this analysis, the three-planet model was consistently preferred over other models with ΔAICc = 18.2 for two-Keplerian (b and c) and ΔAICc = 6.2 for two-Keplerian (b and d).

The best-fit RV model is plotted in Fig. 4.21 in black with its uncertainties as gray shaded areas. Residuals are included. A zoomed-in version of the same plot focused on the last semester of data is also shown in Fig. 4.22. The results of this analysis are not reported in this thesis, but can be found in Tables 3 and 4 of Dalal et al. (2024). In summary, they agreed with the ones derived with in the previous analyses within 1σ uncertainties. Despite the lack of informed priors, the stellar rotation and the orbital period of Keplerian c were well disentangled, with the first yielding $P_{\rm rot} = 43.45^{+1.00}_{-0.71}$ days and the second being $P_{\rm c} = 37.92 \pm 0.03$ days. The RV semi-amplitude of all three Keplerian signals were successfully recovered as $K_{\rm b} = 2.11 \pm 0.13$ m s^{-1} (16σ), $K_{\rm c} = 1.75 \pm 0.25$ m s^{-1} (7σ), $K_{\rm d} = 1.72 \pm 0.22$m s^{-1} ($8\sigma$), corresponding to the minimum masses of $m_{\rm b} = 4.96 \pm 0.32$ M$_\oplus$, $m_{\rm c} = 6.9 \pm 1.0$ M$_\oplus$, and $m_{\rm d} = 11.0 \pm 1.5$ M$_\oplus$.

4.2.7.4 Independent analysis using TWEAKS

A final analysis using the TWEAKS method was also performed. After a blind search for Keplerian signals in the radial velocities using KIMA (Faria et al. 2018), this technique can clearly detect two Keplerian signals at orbital periods 7.3401 ± 0.0004 days and 38.01 ± 0.06 days, with some spurious signals above 300 days, most likely due to instrumental systematics rather than genuine planetary candidates. By constraining the maximum orbital period for the planet search to be less then 270 days, a third planetary candidate with an orbital period of 150.98 ± 0.77 days was found, as shown in Figure 4.23. The following RV semi-amplitudes were obtained after Gaussian Mixture modelling: 2.29 ± 0.13 (18σ) for planet b, 1.48 ± 0.39 m s^{-1} (4σ) for planet c, and 1.55 ± 0.34 m s^{-1} (5σ) for planet d.

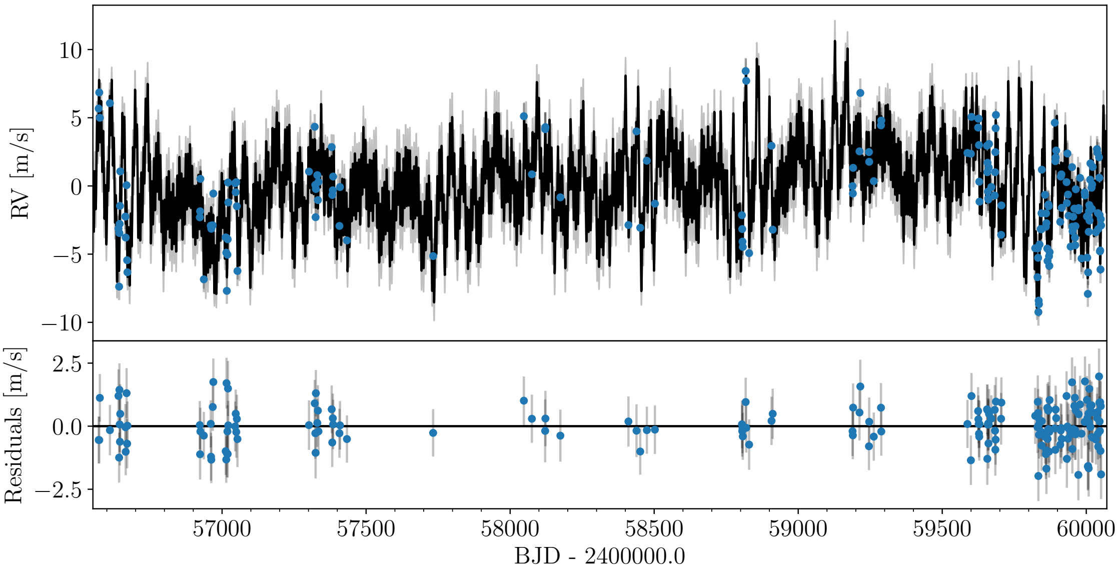

Figure 4.21: Plot illustrating the YVA RVs and the best-fit model obtained through multi-dimensional GP regression. The black line represents the modelled activity with three Keplerians, and the gray shaded region denotes the 1σ uncertainty on the model. The residuals are also displayed. Figure adapted from Fig. A1 of the Appendix of Dalal et al. (2024).

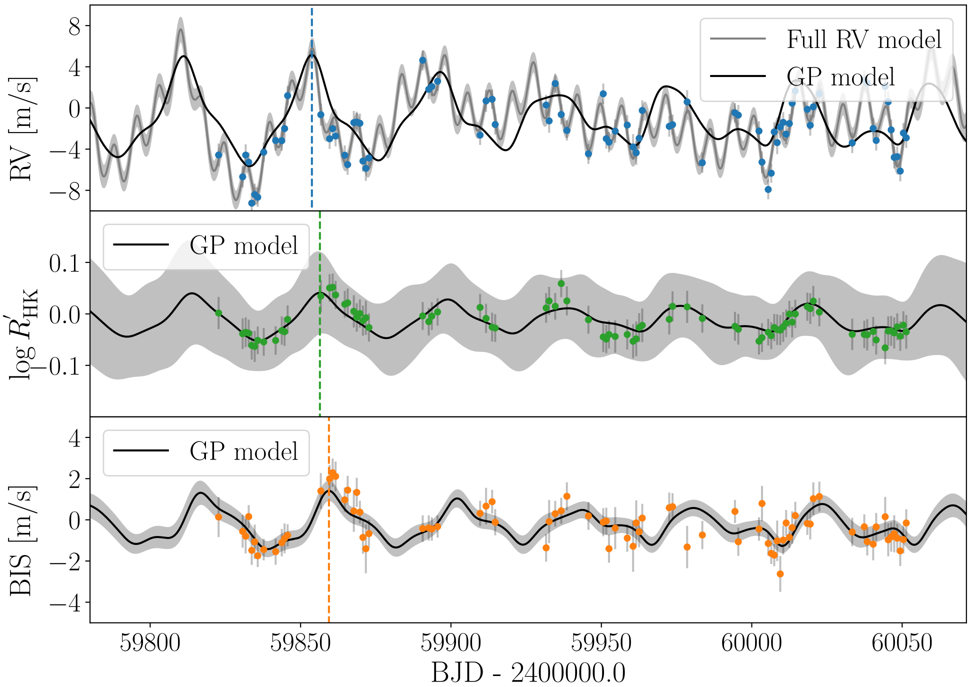

Figure 4.22: Zoomed-in version of the last semester of data and the best-fit model derived from the multi-dimensional GP regression for, from top to bottom, the YVA RVs, the log(R'_{HK}) and the BIS. Here the black lines in all plots represent the best-fit GP function modelling the activity, while the gray solid line in the YVA RVs on the top shows the full model, combining the GP with the three Keplerian signals. Uncertainties on all models are included as gray shaded regions. The dashed blue, green and orange line represent the maxima of corresponding peaks, indicating a time lag between RVs and activity indicators. Figure taken from Fig. 12 of Dalal et al. (2024).

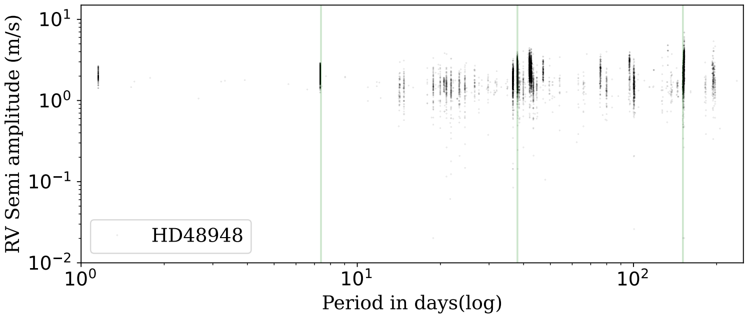

Figure 4.23: Results from the TWEAKS pipeline. The posterior distributions over periods are plotted. The vertical lines depict the three planet detections with significant False Inclusion Probabilities, at 7.34 days, 37.98 days and 150.92 days, respectively. The clustering around the 37.98 day planet signal can be attributed to the stellar rotation period and its one-year aliases. Figure adapted from Fig. 13 of Dalal et al. (2024).

Therefore the minimum masses of all planet candidates were extracted to be 5.43 ± 0.30 $M_\oplus$, 6.33 ± 1.61 $M_\oplus$, and 10.05 ± 2.27 $M_\oplus$ respectively. An extended description of this analysis and its results can be found in Dalal et al. (2024).

4.2.8 Results and Discussion

Through the analyses presented in the previous Sections, I identified three planet candidates with orbital periods of roughly 7.3, 38, and 151 days. The minimum masses derived with the separate analyses are shown in Fig. 4.24. They all agree within their uncertainties. The largest variations were forund for the mass estimates of planet candidate c. This was unsurprising, given the closeness between its orbital period and the stellar rotation period. The results from the one-dimensional GP analysis on the YVA data in particular showed a significantly lower minimum mass with large uncertainties. This behaviour could be explained by the fact that the GP was most likely absorbing some of the power from the Keplerian signal and instead assigning to the stellar variability. On the opposite, YV2 results yielded a larger minimum mass than all other techniques. It is possible the discrepancy arose from either an underfittting of all techniques that model the YVA dataset, or from remaining non-planetary (possibly stellar) variability in the YV2 data which bolstered the planetary signal in the best-fit Keplerian model of Section 4.2.7.1.

All four different techniques presented in this analysis have their advantages and limitations. The modelling of the YV2 data, while the simplest of the four, relies on spectra processing which is not wholly understood. GPs have proven to be successful to model stellar activity on short timescales, but they have not been as successful on longer base-

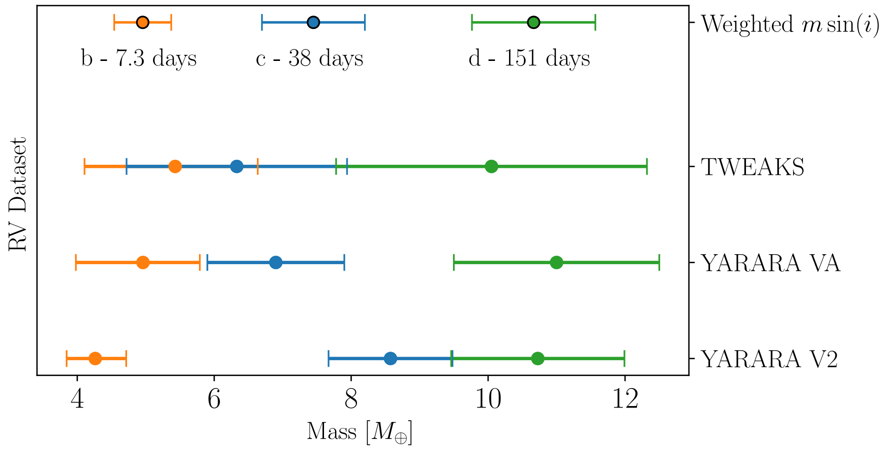

Figure 4.24: The figure shows the minimum masses of three Keplerians derived from various datasets, with the weighted mass combining all three presented at the top. Figure taken from Fig. 14 of Dalal et al. (2024).

lines. Moreover, their large flexibility both allows them to fit the stellar activity in a variety of cases, and makes them prone to overfitting the data, therefore absorbing some of the planetary signals. The multi-dimensional GP framework addresses this overfitting, but is dependent on mathematical relationships between the activity indices and the radial velocities that are approximations. Moreover, it works best for cases in which a good correlation can be found between the RVs and the activity proxies described by the same number of latent GPs (e.g., the $\log(R'_{\mathrm{HK}})$). The highest Person correlation coefficient between the YVA RVs and an activity indicator was computed to be only 0.55 for the $\log(R'_{\mathrm{HK}})$. Finally the TWEAKS method, while it allows for a dynamic modelling of the number of planets, is still significantly sensitive to stellar activity signals and systematics, limiting its precision. I therefore calculated the final results as the weighted mean of the derived minimum masses as 4.96 ± 0.42 $\mathrm{M}_\oplus$ for planet candidate b, 7.45 ± 0.75 $\mathrm{M}_\oplus$ for planet candidate c, and 10.67 ± 0.90 $\mathrm{M}_\oplus$ for planet candidate d respectively.

In order to fully understand the configuration of the planetary system of HD 48948, I calculated the orbital semi-major axis of all planetary candidates. The results of this computation are included in the derived parameter section of Table 4.11. I then derived the incident flux of these three planetary candidates in terms of the incident flux received by Earth from the Sun, $F_{\mathrm{inc},\oplus}$, following the formula described in Eq. 4.10. I also computed the equilibrium temperatures of the three planetary candidates T_{eq} using Eq. 4.11. For simplicity, I assumed isotropic re-emission and uniform equilibrium temperature over

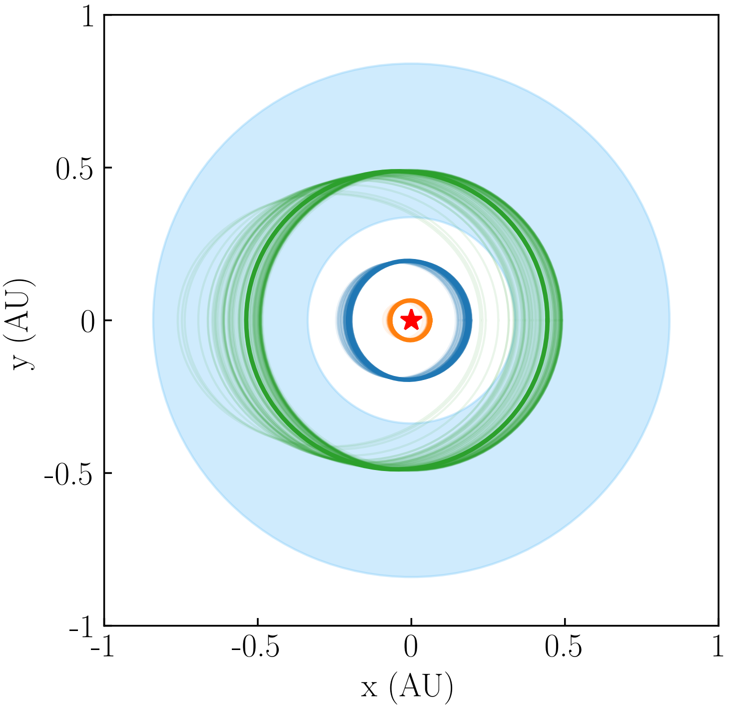

Figure 4.25: Configuration of the proposed HD 48948 planetary system. The inner planetary candidates, depicted in orange and blue, have circular orbits of 7.3 and 38 days respectively. The outer planetary candidate, illustrated in green, has an orbit of 151 days. A selection of 100 random orbits from the MCMC chains for each of the three planetary candidates are represented in lighter shades. The habitable zone boundaries, shown as sky-blue shaded regions, are calculated as outlined in Section 8.2, based on the Kopparapu et al. (2014). Figure taken from Fig. 15 of Dalal et al. (2024).

the entire planet, setting $f = 1$, and black-body absorption, setting $A_{\mathrm{B}} = 0$, and derived an upper limit of T_{eq}. The calculated $F_{\mathrm{inc},\oplus}$ and maximum equilibrium temperatures for HD48948 b, c and d are included in Table 4.11. Finally, I determined the habitable zone around the star. The inner and outer boundaries of the HZ were computed based on the Recent Venus and Early Mars model, as detailed in Kopparapu et al. (2014). The orbits of the system from a top-down view and the calculated HZ are plotted in Fig. 4.25. Planet d falls within the habitable zone of the system, with a orbital semi-major axis of 0.4894 ± 0.0042 AU, incident flux of 2.46 ± 0.12 $F_{\mathrm{inc},\oplus}$ and an upper limit of the equilibrium temperature of 349 ± 4 K.

4.2.9 Conclusions

In this work, I have characterised the planetary system orbiting the nearby, bright K-dwarf HD 48948. I analysed 189 spectra observed by the HARPS-N spectrograph under the Rocky Planet Search mission. The complexity of the system prompted the extraction of multiple versions of the radial velocities with different techniques: DRS, YARARA and TWEAKS. After careful analysis of the activity indicators, I pinpointed the stellar rotation period to be ~42 days. After Fourier analysis, evidence of the presence of 3 planetary candidates with orbital periods of roughly 7.3 days, 38 days and 151 days was found. In order to test and further prove the presence of the middle planet, with orbital period similar to the stellar rotation period, I undertook systematic model testing. I compared the AICcs of a three-Keplerian model against models only including two planetary signals (two configurations: b and c, and b and d). The system was analysed with four independent techniques as follows:

- I modelled the YARARA Version 2 data as the sum of Keplerians and found best-fit parameters with iterative MCMC optimisation. With this modelling technique, I found consistent preference for the three-planet model. I however noted that the YARARA correction for stellar activity and other instrumental effects is not based on physically-motivated models and there is no profound understanding of what it corrects.

- I then performed a one-dimensional Gaussian process regression of the YARARA Version 1 Activity re-injected data. I modelled the stellar activity with a Quasi-periodic kernel and the planets with Keplerians using the pipeline MAGPY-RV. In this case, there was a slight preference for one of the two-planet models (including only planets b and d). I therefore performed mock GP tests in order to understand this behaviour. I used the activity-only radial velocities (computed as the residuals between the Activity re-injected RVs and the YARARA Version 2 data) and manually added a Keplerian signal at the time series level with orbital period equal to the best-fit orbital period for planet c. I did this to test the ability of the GP to retrieve this signal I knew existed. With this analysis I found that it was plausible that the partial (AICc difference under 10) preference for the two-planet system was due to the GP absorbing the signal of the middle planet with orbital period close to the stellar rotation.

- The YARARA Activity re-injected data was also modelled simultaneously to the BIS and the $\log(R'_{\mathrm{HK}})$ activity indicators within a multi-dimensional GP framework, in order address the possible GP overfitting that was limiting the one-dimensional analysis. Lastly, the RVs were also extracted and modelled with TWEAKS. In both analyses, a three-planet model was preferred, and the derived best-fit results were in agreement within 1σ uncertainties with each other and the previous results.

After careful comparison, I finally report the presence of three planetary candidates orbiting HD 48948. I find their orbital periods to be roughly 7.3, 38, and 151 days, and their minimum masses to be $4.96 \pm 0.42\,M_{\oplus}$, $7.45 \pm 0.75\,M_{\oplus}$, and $10.67 \pm 0.90\,M_{\oplus}$ for candidates b, c and d respectively. The outer planet was also found to reside in the habitable zone of the system. Overall this work is a further proof of the challenges posed by stellar variability in the radial-velocity regime and is also an analysis of the limits of current modelling techniques. This work emphasised the need of truly understanding stellar variability for the positive detection of small exoplanets, in particular of those with orbital period close to the stellar rotation.

Chapter 5

Solar Activity

> *"Even*
> *After*
> *All this time*
> *The Sun never says to the Earth,*
> *"You owe me."*
> *Look*
> *What happens*
> *With a love like that,*
> *It lights the whole sky."*
>
> — *Hafiz*

One of the greatest challenges limiting the understanding of stellar activity in the radial-velocity regime is the fact that the great majority of variability is highly degenerate with spatial distribution. At most our resolving capabilities have to rely on maximum entropy assumptions (as in the case of e.g. Doppler Imaging), and even in the reconstruction of simulated stellar surfaces, RV data can at best only recover the large scale structures of the sources of variability. There is, however, a star that we are always able to resolve, and that fits right into the target list of planet hosts that would be selected for a standard exoplanet survey: the Sun. The Sun is a G2V-type main-sequence star. It is generally considered a slow rotator, with a period of roughly 27 days. It formed approximately 4.6 billion years ago and is considered to be generally on the quieter side in the context of

stellar variability in the RVs, with a mean $\log(R'_{H,K}) \sim -5.02$ (e.g., Lubin et al. 2012). The Sun can thus overall be considered a good starting point for modelling and mitigating activity-induced variability for later extrapolation to other stars.
It has been observed since hundreds of years ago. Sporadic observations can be traced back to the 1610s, when the quartet of scientists Galileo Galilei of Italy, Johannes Fabricius of Holland, Christopher Scheiner of Germany, and Thomas Harriott of England all began to observe the presence of dark blemished (sunspots) on the solar surface and record them in the form of sketches. More systematic observations started in 1749, when the Zurich Observatory started recording daily images. With the inclusion of other telescopes at different latitudes, they were able to obtain near-continuous observations starting in 1849. Since then, a variety of solar telescopes and space-based missions have dedicated their time uniquely to the study of the Sun and the wealth of physical processes taking place on its surface, in its atmosphere and in its interior.

Scientists have used solar data in many ways. In the context of radial-velocity surveys the Sun has not only become a testing ground for new mitigation techniques, but it also is the ideal "standard star" to build a full understanding of the profile of the spectrograph, of its instrumental systematics and stability capabilities. RV studies can use Sun-as-a-star data to mimic stellar observations and integrate all the solar flux into a point-like source to scramble its signal. The HARPS-N solar telescope (Dumusque et al. 2015; Phillips et al. 2016; Collier Cameron et al. 2019) has been taking observations since August 2015 with a 5-minute cadence. The HARPS Experiment of Light Integrated Over the Sun (HELIOS) is a copy of the HARPS-N solar telescope in the southern hemisphere that started observing the Sun in September 2018, with a 1-minute cadence. The NN-explore Exoplanet Investigations with Doppler spectroscopy instrument (NEID: Lin et al. 2022) solar telescope followed soon after, with first-light in December 2020 observing a wider range of wavelength with a ~1.4-minute cadence. A solar telescope was also connected to the EXPRES spectrograph (Llama et al. 2022) in 2020, and it has been taking data over a similar wavelength range as the NEID solar feed with a cadence comparable to the HARPS-N solar telescope. Finally, the Birmingham Solar-Oscillations Network (BiSON: Hale et al. 2020) is a collection of ground-based automated solar telescopes that have been collecting continuous data since 1992 and therefore represent the longest-running RV solar dataset readily available to the community. The observations taken with these telescopes have been used in a variety of ways to support stellar activity analysis for the detection of exoplanets (e.g., Haywood et al. 2016; Collier Cameron et al. 2019; Haywood et al. 2022; Zhao et al. 2023). Most crucially, researchers are able to directly assess the RV variability in Sun-as-a-star data by comparing the extracted RVs with resolved observations of the solar surface. There are a number of ground- and space-telescopes that take resolved images of the Sun over a range of wavelengths and with a variety of techniques. The most relevant space mission in the context of this work is the Solar Dynamics Oservatory (SDO: Pesnell et al. 2012). The SDO spacecraft was lauched in February 2010 as a follow-up mission to the Solar Heliospheric Observatory (SOHO: Fröhlich et al. 1995; Fröhlich et al. 1997), and has been observing the solar disc continuously since late 2010. SDO includes three instruments: the Helioseismic and Magnetic Imager (HMI: Schou et al. 2012; Scherrer et al. 2012), the Extreme Ultraviolet Variability Experiment (EVE), and the Atmospheric Imaging Assembly (AIA). Given the long baseline and the high cadence of SDO observations, there has also been interest in the development of systematic methods to derive Sun-as-a-star radial velocities from resolved solar observations comparable to those retrieved by ground spectrographs. In this context, the Python pipeline SolAster was created.

In this Chapter, I make use of resolved and Sun-as-a-star observations to investigate the effects of solar variability on radial-velocity data. Section 5.1 briefly describes the SolAster pipeline for a better understanding of the algorithms employed for the derivation of data used in the following Section. Section 5.2 includes the study of the mean longitudinal magnetic field as a tracer of stellar activity in the RV-domain.

5.1 SolAster: an SDO/HMI Analysis Pipeline

The Solar Dynamics Observatory was launched in 2010 by NASA's Living With a Star Program, a program designed to understand the causes of solar variability and their impacts on Earth. The aim of the SDO mission was specifically to study the solar atmosphere on short timescales over many wavelengths. One of its three scientific instruments, the Helioseismic and Magnetic Imager has been taking continuous full-disc observations of the solar surface with its two cameras of 4096x4096 pixels nearly without interruption since mid-2010. The SDO/HMI instrument has near single-granule resolution (Schou et al. 2012; Pesnell et al. 2012). It takes polarised filtergrams of the visible solar disc in two polarisation states by measuring six wavelengths centred in the 6173.3 Å neutral Fe I line (Couvidat et al. 2016). Observations are taken every 45 seconds, as well as compiled in 12-minutes (720s) integrated exposures. These filtergrams are then reduced with two main pipelines: the Line of Sight Pipeline and the Vector Pipeline (for more information refer to Couvidat et al. 2016 and Hoeksema et al. 2014). In summary, the six observed wavelengths are fitted with a Gaussian profile to calculate the observable characteristics of the solar surface, such as continuum intensity, photospheric Doppler velocity and magnetic field via Stokes profiles. While the fitting loses any line asymmetry generated within the pixel area, larger processes are preserved.

The SolAster pipeline was developed to extract Sun-as-a-star observations from resolved solar SDO/HMI filtergrams. In particular, it allows for direct comparison between space-based measurements and ground-based disc-integrated spectral data. A packaged version of the pipeline was published by Ervin et al. (2022a)[1] based on the work of Haywood et al. (2016) and Milbourne et al. (2019), which themselves built on techniques developed by Meunier et al. (2010). SolAster requires as inputs a series of SDO/HMI images from which it derives all its outputs: the wide-band continuum (the intensitygrams), the line-of-sight longitudinal magnetic field data (the magnetograms), and the map of the line-of-sight velocity (the Dopplergrams). The great majority of these data are freely available at https://sdo.gsfc.nasa.gov/data/aiahmi/ or can be queried from the archive using SunPy (SunPy Community et al. 2015), with the exception of some brief periods of planetary transits which are often kept private for a short time. The pipeline workflow begins with coordinate transformation based on the description in Thompson (2006), so that the images are correctly centred on the solar surface. The image pixels are assigned new position coordinates based on their distance from the centre-point of the solar disc. The "edge" of the Sun is then identified based on these new coordinates, and the fluxes of all pixels defined as "outside" the Sun are set to zero. The next step is to correct the Dopplergram for the motion of the spacecraft itself. The position of SDO with respect to each pixel in the image, carefully recorded for each observation, can be combined with its velocity component

---

[1] Code available at: https://github.com/tamarervin/SolAster

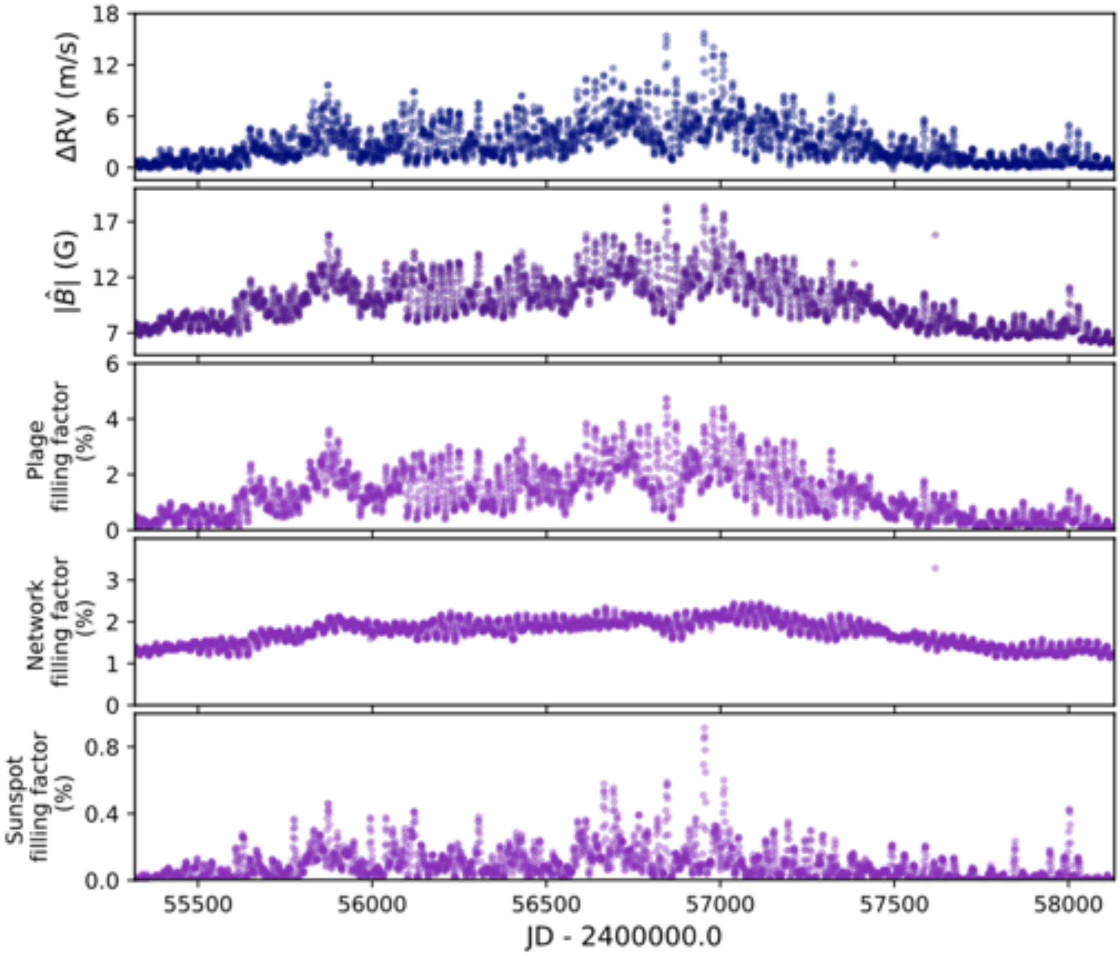

Figure 5.1: From top to bottom, the ΔRVs derived using the physically motivated model described in the text, the unsigned magnetic field and the filling factor of plage (or facuale, large-scale bright active regions), network and sunspots. Figure adapted from Fig. 4 of Haywood et al. (2022).

to properly subtract the velocity of the spacecraft. In this process the Doppler velocities are also corrected for the solar Carrington rotation in order to highlight all line-of-sight velocities that cannot be explained by the spinning of the Sun. The Dopplergrams are thus de-rotated following the parameterisation of Snodgrass and Ulrich (1990). Two further corrections need to be applied to the magnetograms and intensitygrams respectively. The true line-of-sight magnetic field is calculated by including foreshortening effects due to the geometric projection of the disc (for more a specific description of the process, see Section 5.2.2.2). Finally the intensitygrams are corrected for limb-darkening effects using a static fifth-order circularly symmetric polynomial brightness function with scaling coefficients determined by Allen (1973), and a flattened continuum image is then produced.

SolAster includes a wide range of outputs. In this work, I will focus on four families:

the magnetic field, the intensity, the filling factors and the radial-velocity data. A selection of these outputs is plotted in Fig. 5.1. Starting with the magnetic outputs, the pipeline provides both the mean longitudinal magnetic flux and the mean absolute magnetic flux. The first is computed as the sum of the continuum intensity-weighted, line-of-sight magnetic field in all pixels of the resolved disc (for an extended explanation, see Section 5.2.2.2). The second can be computed similarly but considering the absolute magnetic field, instead of its line-of-sight values and is plotted in the second panel of Fig. 5.1. These data have been analysed by Rescigno et al. (2024) (also included in this Chapter) and Haywood et al. (2022) respectively to study the uses of magnetic field measurements in the context of RV surveys. SolAster also outputs the mean solar intensity of each observation as well as the brightness intensity of the "quiet" Sun regions. There are multiple routines that are tailored to the identification of active regions on the solar surface. The SDO/HMI consortium itself provides the geometric information of active regions as they cross the visible disc and evolve. These images are produced by the HMI Active Region Patches (HARPs: Turmon et al. 2012) pipeline (not to be confused with the HARPS telescope) and are available in the SDO archive or via SunPy query. SolAster also identifies magnetically active regions following the threshold method described in Yeo et al. (2013): all pixels with foreshortening-corrected unsigned magnetic field larger than three times the average noise of 8 G are considered active. All other pixels are defined as quiet. Bright areas and dark spots are then differentiated based on a flattened continuum intensity threshold proportional to the mean intensity of all quiet pixels. Milbourne et al. (2019) further extended this formalism by separating bright regions into faculae with areas larger than 60 Mm^2, and network, which is mostly made up of isolated active pixels. SolAster outputs this information for each observation in the form of filling factors, separated in total active region filling factor, spot filling factor, faculae (or plage) filling factor and network filling factor. A subsection of these is included in the bottom three panels of Fig. 5.1. Finally, SolAster extracts disc-integrated radial velocities. The SDO/HMI instrument was not developed with long-baseline RV analysis in mind, and therefore it lacks the stability required to do so. Thus, radial velocities cannot be computed from a simple intensity-weighted summation of the line-of-sight velocity of each pixel, as highlighted by Fig. 5.2. SolAster instead derives the RVs using a physically-motivated model. This model accounts only from stel-

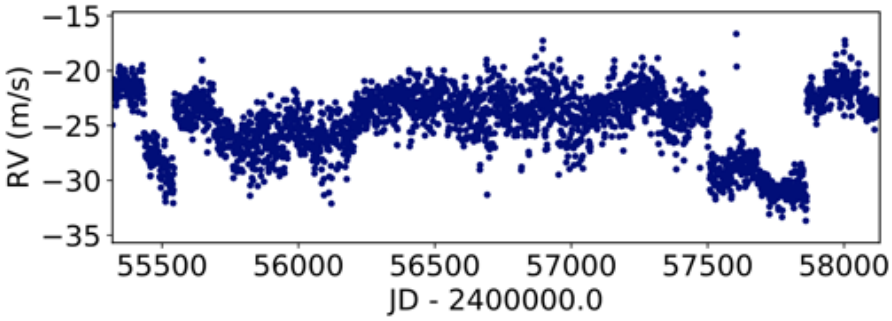

Figure 5.2: Radial velocity time series derived from a simple intensity-weighted sum of all the line-of-sight velocities of the de-rotated solar SDO/HMI observations. The multiple trends and jumps in the data are caused by instrumental systematics, as the HMI instrument is not calibrated for long-term stability. These systematics are not trivial to correct for, and a physically-motivated model is therefore preferred. Figure taken from Fig. 2 of Haywood et al. (2022).

lar variability directly induced by the presence of large active regions (faculae and spots) following the two processes introduced in Chapter 1, Section 1.2.2: suppression of convective blueshift and photometric inhomogeneities. The computed radial velocities do not therefore include all other effects contributing to the stellar variability in ground-based disc-integrated Sun-as-a-star observations. They are calculated as

$$\Delta \mathrm{RV} = A(t)\Delta\hat{v}_{\mathrm{phot}} + B(t)\Delta\hat{v}_{\mathrm{conv}} + v_0, \tag{5.1}$$

in which $A(t)$ and $B(t)$ are weighting coefficients, $\Delta\hat{v}_{\mathrm{phot}}$ is the photometric contribution, $\Delta\hat{v}_{\mathrm{conv}}$ is the suppression of blueshift term, and v_0 represents an instrumental parameter constant with time, an offset of sort that depends on the Sun-as-a-star data the derived ΔRVs are meant to be compared against. Haywood et al. (2016), Milbourne et al. (2019) and Ervin et al. (2022a) all include step-by-step derivation of each term in Eq. 5.1. SolAster provides the full reconstructed ΔRV as well as each component separately. These derived ΔRVs have been compared to HARPS-N and NEID integrated-disc observations, and have since been used in various stellar variability analyses (e.g., Lienhard et al. 2023; Lakeland et al. 2024).

In the remaining part of this Chapter, I use SDO/HMI-derived disc-integrated time series to study solar variability. In particular, I aim to assess whether the mean longitudinal magnetic field can be used as a tracer of the solar magnetic activity.

5.2 The Mean Longitudinal Magnetic Field and its Uses in Radial-Velocity Surveys[2]

5.2.1 Introduction

In the last 20 years, the radial-velocity method has been used to successfully detect and characterise hundreds of exoplanets. With the aim of finding potential Earth analogues that future missions such as the Habitable Worlds Observatory[3] (HWO: Harada et al. 2024) or the Large Interferometer for Exoplanets (LIFE: Quanz et al. 2022) can observe in the search of biosignatures, the community is now more than ever targeting rocky exoplanets in their stellar habitable zone. As it has been re-iterated throughout this thesis, stellar variability is now the greatest challenge in this pursuit, as stellar-induced signals often dominate the RV variations of the observed stars. These signals are still challenging to model as they affect the time series over multiple timescales, from minutes to years. Moreover, the longer baselines required for disentangling Earth-like signals introduce a further source of "noise": magnetic cycles.

As addressed in Chapter 1 Section 1.2.3, over the years stars are expected to undergo similar activity cycles (e.g. Oláh et al. 2009) to what the Sun experiences, with years of maxima, where activity is much stronger and more significantly modulated by stellar rotation, and stretches of minima, where activity-induced variations are much weaker and non-rotationally-modulated effects dominate (e.g., granulation and supergranulation). Understanding and modelling these long-term cycles is often necessary for a comprehensive characterisation of planetary systems, in particular in the case of possible wide companions. A contemporaneous effort towards the confirmation of outer planets may in fact be vital for the detection of Earth analogues. In fact, recent studies have shown that the formation of inner Earths is dependent on the presence of quickly-accreted long-orbit gas giants (Morbidelli et al. 2022).

A common approach to activity mitigation is to use Gaussian processes. GP regression however has its limitations. In particular, its ability to predict stellar variability is strongly

---

[2] This Section, alongside all Figures and Tables have been taken from Rescigno et al. (2024).

[3] Based on the studies for the Large UV/Optical/IR Surveyor (LUVOIR: The LUVOIR Team 2019), and the Habitable Exoplanet Observatory (HabEx: Gaudi et al. 2020)

reliant on an accurate detection of the stellar rotational period (Nicholson and Aigrain 2022). A precise determination of the periodicity of the stellar activity-induced RVs is vital to correctly differentiate them from Keplerian signals (Bortle et al. 2021), and to compute accurate masses (Blunt et al. 2023). In fact, confirming the presence of non-transiting planets can be particularly challenging when the stellar rotational period is similar to the orbital period of the planet candidate (Nava et al. 2022; Dalal et al. 2024), as was the case in Chapter 4 Section 4.2. An inaccurate rotational period can also have significant direct impacts on the derived best results for all the other kernel hyperparameters that are less reliably tied to physical processes and therefore much harder to interpret correctly.

An accurate detection of the stellar rotation period is vital for many other areas of astrophysics beside exoplanet characterisation. As an example, Irving et al. (2023) studied the relationship between stellar rotation periods and magnetic cycle amplitudes, as well as between the ratio of rotation and cycle periods and the stellar Rossby number. McQuillan et al. (2014) highlighted a bimodality in the rotation period-temperature relation of more than 30,000 *Kepler* targets. Amard et al. (2020) studied the contribution of stellar metallicity to the decay of their rotational periods with age. On the whole, accurate and precise measurements of stellar rotational periods are at the basis of multiple current fields of study. The stellar rotation period is often challenging to extract only from RVs (Nava et al. 2020) or photometry (Aigrain et al. 2015), especially in times of low activity. We therefore use activity proxies as extra suppliers of information, as covered multiple times in this work. Nevertheless, even an analysis of the common activity indicators often fails to consistently measure the stellar rotation period (e.g., Nava et al. 2022). Therefore for the analysis of RVs and in particular for the detection of the stellar rotational period, especially over all stages of a star's magnetic cycle, a different tracer of activity is required.

5.2.1.1 The Mean Longitudinal Magnetic Field

Haywood et al. (2022) have shown that the unsigned (absolute) magnetic flux maps the stellar-induced RV variations better than any other activity indicator to date. However, measuring the unpolarised magnetic flux in stars that are not the Sun is extremely challenging (Reiners 2012). Stellar magnetism is usually investigated with polarimetric observations (see Trippe (2014) and similar reviews).

Polarisation of light is fully characterised by four Stokes parameters I, V, Q and U. The first parameter I is the total intensity. Circular polarisation V is the difference of intensities carried by left- and right-hand photons with electric field vector rotating around the direction of propagation. Linear polarisation Q and U are the difference of intensities carried by vertically and horizontally polarised photons in two coordinate systems rotated relative to each other by 45 degrees, with the line-of-sight as the rotation axis. Typically, a single observation yields two Stokes parameters: I and one of polarisations. Radiation formed in the presence of a magnetic field polarises the light through the Zeeman effect. In this case, V is sensitive to the line-of-sight component of the magnetic field, and Q and U to the transversal field component. All are needed to directly detect the absolute magnetic field. The linear polarisation (Q and U) is a second order effect in the wavelength-domain and, especially in the case of weak-fields, it has significantly smaller signals than the circular polarisation (Bagnulo and Landstreet 2015). In fact, in most cases the signals of Q and U are so small and require such high signal-to-noise ratio that they are not measurable with current precisions and reasonable exposure times. Although some newly developed techniques are extracting a proxy of the unsigned magnetic field directly from the spectra without employing Stokes profiles (e.g., Lienhard et al. 2023), it remains a challenge.

I instead turn to the more easily detectable mean longitudinal line-of-sight magnetic field. It only requires I and V Stokes profiles, which have stronger signals. The mean longitudinal magnetic field, B_l (sometimes also referred to as $<B_\mathrm{z}>$), is the line-of-sight projected component of the magnetic field vector averaged over the visible hemisphere of the star. B_l is related to the circular polarisation as

$$\frac{V}{I} = g_\mathrm{eff} C_\mathrm{z} \lambda_0^2 \frac{1}{I}\frac{dI}{d\lambda} B_\mathrm{l}, \tag{5.2}$$

where g_eff is the effective Landé factor, I the intensity at wavelength λ, λ_0 is the average wavelength, and $C_\mathrm{z} = 4.67 \times 10^{-13}$ Å^{-1} G^{-1} (e.g. Landstreet 1982). The mean longitudinal field can then be expressed as the first order moment of the Stokes V parameter as

$$B_\mathrm{l} = -2.14 \times 10^{11} \frac{\int \nu V(\nu) d\nu}{\lambda_\mathrm{av} g_\mathrm{av} c \int [I_\mathrm{c} - I(\nu)] d\nu}, \tag{5.3}$$

where $\lambda_{\rm av}$ and $g_{\rm av}$ refer to the average wavelength and the average Landé factor of the lines used to compute $B_{\rm l}$, and $I_{\rm c}$ is the continuum intensity (Donati and Collier Cameron 1997). The integration limits over frequency ν are somewhat arbitrary and can change between analyses. They are selected wide enough to include all the information of the Stokes profiles but narrow enough to reduce the contribution of noise.

Previous analyses with the near-infrared SPectropolarimètre InfraROUge (SPIRou: Donati et al. 2020) have shown that $B_{\rm l}$ derived from polarimetry can be used to successfully detect rotational periods of chemically peculiar stars (Babcock 1949) and M-dwarfs (e.g., Landstreet 1992; Donati and Landstreet 2009; Klein et al. 2021; Fouqué et al. 2023). With these successes, other polarimeters were also turned to similar analyses (e.g., Hébrard et al. 2016). Donati et al. (2023) also introduced Gaussian process regression to the modelling of $B_{\rm l}$ in M-dwarfs. For Sun-like stars, and most of the stars selected in RV surveys for exoplanet detection, the low observed projected rotational velocity ($v\,\sin(i) < 2$ km s^{-1}) makes the detection of complex magnetic fields difficult due to magnetic flux cancellation between opposite polarities. To test these limits, Petit et al. (2008) observed a small sample of active Sun-like stars, and successfully detected their magnetic fields. On a larger scale, the BCool magnetic survey (Marsden et al. 2014) analysed spectropolarimetric data of 170 solar-type stars (F-, G- and K-type or FGK) collected between 2006 and 2013. With mostly a single observation per star, the survey reached precisions in the $B_{\rm l}$ of 0.2 G, and demonstrated that $B_{\rm l}$ in quieter Sun-like stars is measurable with reasonable uncertainties. However, this survey and the majority of previous polarimetric surveys focused on obtaining mostly single snapshot observations. The great majority of FGK stars lack the time series of polarimetric data necessary to do period detection analysis. As previous studies have done to better understand stellar variability and its dependence to other measurable quantities (e.g., Haywood et al. 2016; Collier Cameron et al. 2019; Haywood et al. 2022), I turned to the best observed star that falls in that category: the Sun.

The origin of $B_{\rm l}$, in solar science often also called Solar Mean Magnetic Field (SMMF) or General Magnetic Field (GMF), is still strongly debated. Some attribute the largest contribution to the signal of $B_{\rm l}$ to the weak, large-scale magnetic flux over the entire visible disc (i.e., the "quiet" Sun flux) (e.g., Severny 1971; Xiang and Qu 2016). In fact, Bose and

Nagaraju (2018) claimed that 80% of the signal of B_l is generated by the background magnetic flux. Using resolved full-disc solar data, they partitioned the solar surface between sunspots, faculae and background. They then calculated the percentage variation of B_l due to each region independently using the coefficient of determination method based on linear regression analysis. They found that there exists a clear correlation between B_l and its component from only the background. They also found no correlation between the mean longitudinal magnetic field and the active regions filling factors. They concluded that the presence of active regions does not directly influence the structure of the signal in B_l, but their location on the disc can influence the amplitude of the signal (as further explained in Section 5.2.3).

These results are in opposition to the ones of others. For example, Scherrer et al. (1972) showed that the largest correlation between B_l and the interplanetary magnetic fields is reached when considering only the innermost fourth of the solar disc, which is more sensitive to active latitudes and therefore to active regions. Furthermore, Kutsenko et al. (2017) used a similar technique to Bose and Nagaraju (2018) on similar data, but recovered different results. They considered a magnitude threshold of 30 Mx cm^{-2} and found that the B_l component derived from active regions contributed from 65 to 95% of the total field. They therefore claimed that B_l is directly generated by magnetic flux concentrations, meaning spots, faculae and network. They asserted that the strong rotational modulation measured in B_l is a clear, if indirect, proof of its relationship with the active region flux. Overall, the source of the opposing results seems to be the different methods for the definition for active regions: Kutsenko et al. (2017) sectioned the solar surface with a magnetic flux concentration mask on the magnetogram, while Bose and Nagaraju (2018) separated active regions from background with a combination of intensity thresholds on the AIA 1600 Å and the 4500 Å images for plage and sunspots respectively. Nevertheless, assessing the true source of the variations of B_l is beyond the scope of this work. I will instead focus on addressing how its behaviours can help us understand stellar activity in the RVs and pinpoint the stellar characteristics needed for activity modelling.

In this Chapter, I use resolved solar observations to extract the mean longitudinal magnetic field of the Sun, and compare it to the radial velocities and the common activity proxies of Sun-as-a-star observations. The aim of this analysis is to determine whether B_l

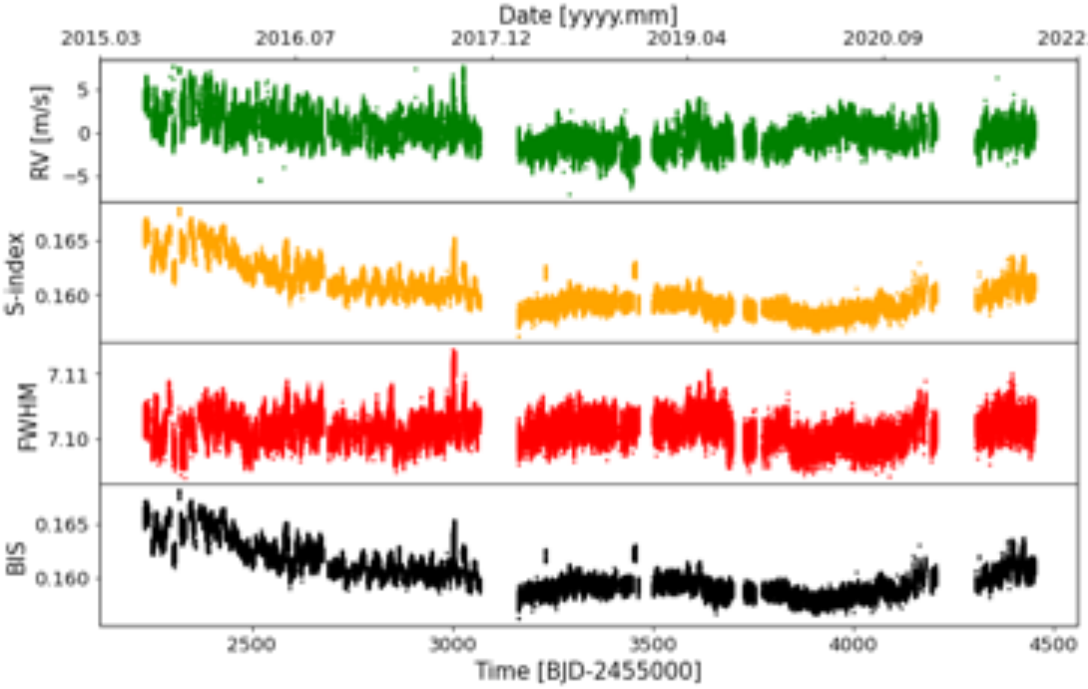

Figure 5.3: HARPS-N Solar telescope data. From the top, the corrected radial velocities in green, the S-index in orange, the full-width at half-maximum in red, and the bisector span in black. Uncertainties are included but may be too small to be visible.

can be a useful tracer of stellar activity in Sun-like stars in the context of RV surveys. I describe the data in Section 5.2.2. I analyse the derived time series to better identify the properties and periodicities of B_l in Section 5.2.3. Section 5.2.4 covers how I undersampled the data in order to emulate stellar observations, and the tests to assess the ability of the mean longitudinal magnetic field to recover the stellar rotational period and to support RV analysis in a GP regime. Finally, I conclude in Section 5.2.5.

5.2.2 Observational Data

5.2.2.1 HARPS-N Sun-as-a-star Data

The HARPS-N solar telescope (Dumusque et al. 2015; Phillips et al. 2016; Collier Cameron et al. 2019) is a 7.6-cm achromatic lens which feeds the sunlight to an integrating sphere and through an optical fiber into the High Accuracy Radial-velocity Planet Searcher for the Northern hemisphere spectrograph (HARPS-N: Cosentino et al. 2012; Cosentino et al. 2014). It is mounted on the Telescopio Nazionale Galileo (TNG) at the Observatorio del Roque de Los Muchachos in La Palma, Spain. Sun-as-a-star spectra are taken continuously throughout the day, with exposure times of 5 minutes in order to average over the solar oscillations. RVs are then extracted using the 2.3.5 version of the ESPRESSO pipeline applied to HARPS-N, the Data Reduction Software (DRS: Dumusque et al. 2021), which computes the cross-correlation function using a G2 stellar mask. From the CCFs, it also calculates the standard activity indicators: the full-width at half-maximum and the bisector span. Using the Ca H&K lines, the pipeline also measures the S-index. The data is corrected from solar system peculiarities, so they can treated as if they were truly coming from a star. The details of these corrections, which are summarised here, are explained in Collier Cameron et al. (2019) and Dumusque et al. (2021). First in order to strip the signal of the Solar System planets, the extracted spectral data are interpolated on the wavelength scale of the heliocentric frame of reference. The effects of differential extinction (noticeable due to the Sun being resolved in the sky) are then removed. The FWHM is corrected for the effects of the Earth's orbital eccentricity and obliquity. Finally, the S-index is corrected for ghosts on the CCD.

Some of the observations will be affected by clouds or other bad weather. To select the best data, strict cuts based on a data quality factor, Q_f, and on the metrics of the exposure meter are applied. For a detailed description of the applied data cuts, see Rescigno et al. (2024). In total, I considered 64,332 data points from BJD 2,457,232.873 (2015-Jul-29) to BJD 2,459,449.104 (2021-Aug-22). All the mentioned HARPS-N time series are shown in Fig. 5.3.

5.2.2.2 SDO/HMI Resolved-Sun Images

In this work, I use the 720-second integrated SDO/HMI exposures of the continuum photometric intensity, the Dopplergrams, and the magnetograms reduced with the Vector Pipeline. An example set of images is shown in Fig. 5.4. While the telescope produces near continuous observations, I choose a cadence of four hours, yielding six images per 24-hour period and 31,755 images spanning nearly 13 years from BJD 2,455,318 (2010-May-1) to BJD 2,459,945 (2022-Dec-31).

5.2.2.2.1 Estimating the full-disc solar longitudinal magnetic field and radial velocities

While the data is corrected to account for most instrumental effects, long baseline analysis of the solar Doppler velocities was not the original aim of the SDO mission, as addressed

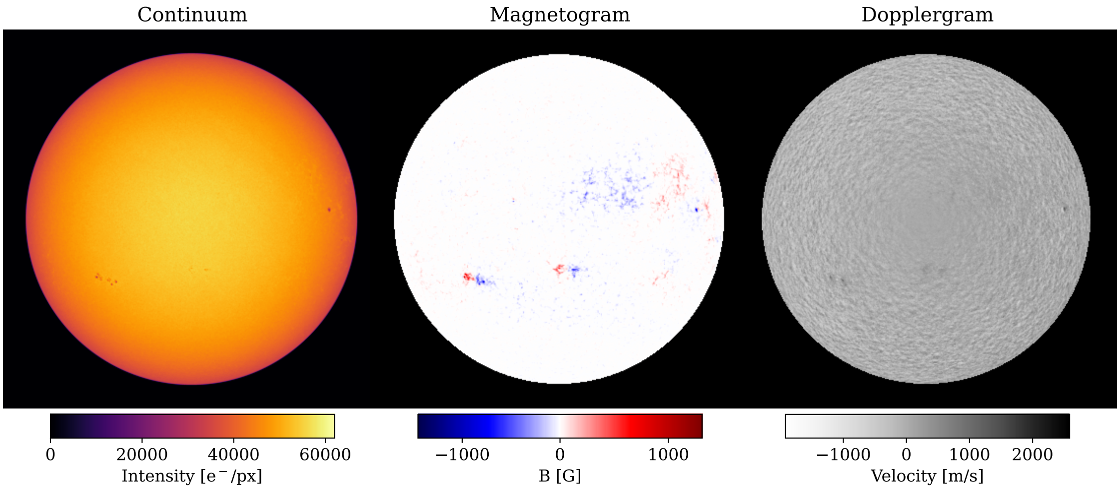

Figure 5.4: Example SDO/HMI images from 2015-Jul-29. From left to right: the continuum intensity (uncorrected for limb darkening), the line-of-sight magnetogram, and the Dopplergram (corrected for the solar rotation and spacecraft motion).

in Section 5.1, and long-term stability of the instrument was not prioritised. I thus use the SolAster pipeline to extract disc-integrated time series of all considered filtergram sets. The mean longitudinal magnetic field and the model-derived ΔRVs are plotted in Fig. 5.5. In this Section, I also describe in more detail the derivation of the B_l data.

In SDO/HMI data, the line-of-sight magnetic field, B_los, is computed for each pixel as the difference of the Doppler velocities observed in two circular polarisations, V'_LCP and V'_RCP:

$$B_\mathrm{los} = (V'_\mathrm{LCP} - V'_\mathrm{RCP})K_\mathrm{m}, \tag{5.4}$$

in which $K_\mathrm{m} = 0.231405$ for a Landé g factor of 2.5. HMI actually directly measures flux density in each pixel, but because a filling factor of one is assumed, a flux density of 1 Mx cm^{-2} is equivalent to a field strength of 1 G (Couvidat et al. 2016). This method is analogous to how the magnetic field is extracted for Magnetic Detection and Imaging (MDI). The 720s version of this variable is computed using selected filtergrams for ten 135s vector fields sequences from Camera 2. After full-disc foreshortening corrections, the disc-averaged, longitudinal magnetic field of the Sun is computed in each observation by summing the continuum intensity-weighed, line-of-sight magnetic field in each pixel of coordinates i and j on the resolved disc as

$$B_\mathrm{l} = \frac{\sum_{ij} B_\mathrm{los,ij} I_{ij}}{\sum_{ij} I_{ij}}, \tag{5.5}$$

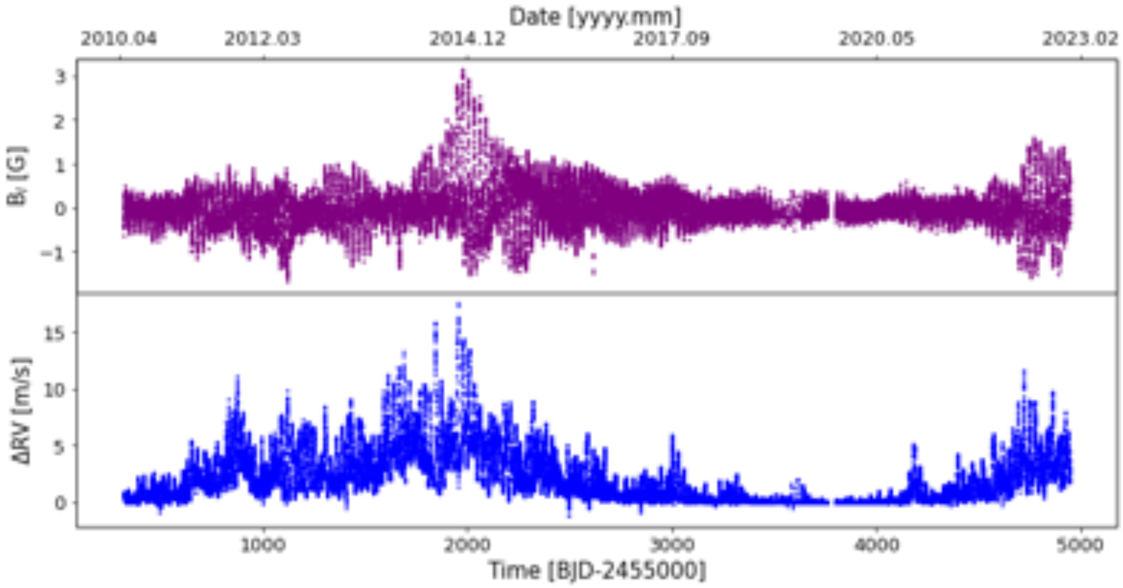

Figure 5.5: SDO/HMI-derived mean longitudinal magnetic field on the top, and the model radial velocities on the bottom. Uncertainties are not included as they would be too small to be visible.

in which I_{ij} is the observed, non-flattened continuum intensity in the same pixel. The derived time series is plotted on the top of Fig. 5.5 in purple. The uncertainties of the longitudinal magnetic field at each pixel increase as a function of their position on the disc and distance from the centre, expressed as μ angle, with them being ~5 G at disc centre and ~8 G at the limbs (Yeo et al. 2013). Even assuming a consistent 8 G noise level, the Poisson-derived uncertainties on the disc-averaged values are incredibly small. Therefore a larger uncertainty will be assumed for the majority of the analysis, as fully addressed in Section 5.2.4.1. The maximum field strength derived for the solar magnetic cycle (Cycle 24 and the beginning of Cycle 25) included in the data is of 3.05 G, which is comparable to the average maximum field derived by the BCool collaboration for G stars of 3.2 G, once again underlining the validity of the comparison.

5.2.3 Full Time Series Analysis: how does B_l relate to the RV variations?

I begin the analysis by assessing the basic properties of the mean longitudinal magnetic field compared to the other derived time series. With this study I aim to answer the following questions: i) Is B_l a direct proxy of activity-induced RV variations? ii) How does B_l change with activity and does it behave like the RVs? iii) What information can be extracted from analysing B_l that cannot be derived from other common indicators? iv) Can

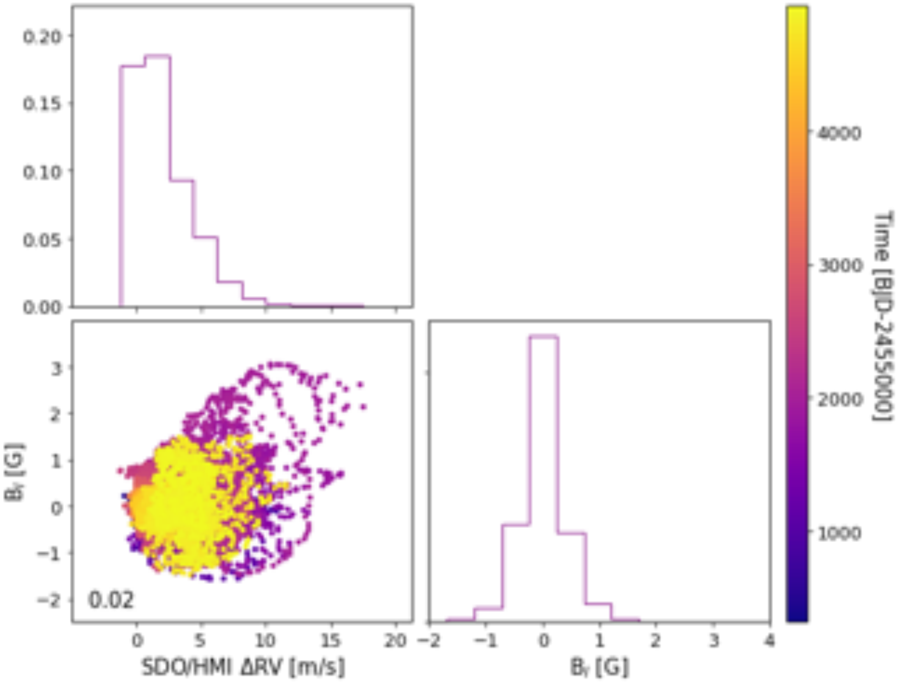

Figure 5.6: Correlation plot between the SDO/HMI-derived radial velocities and the mean longitudinal magnetic field. The colour indicates the Julian date of each datapoint. The computed Spearman rank correlation factor is also included.

B_l be used to trace and model the solar magnetic cycle?

5.2.3.1 Full Time Series Correlation Analysis

To assess whether the mean longitudinal magnetic field can be used to directly map the SDO/HMI rotationally-modulated stellar activity-induced RVs, I compute the Spearman rank-order correlation coefficient of the two time series. When considering all 13 years of data, I calculate a correlation coefficient of 0.02, as shown in Fig. 5.6, meaning that B_l does not correlate with its contemporaneous SDO/HMI-derived ΔRVs. I also compute the correlation between the absolute values of B_l and the ΔRVs. Their Spearman rank correlation coefficient is equal to 0.42, a low/moderate correlation. I compare these results to the correlations calculated between the entirety of the HARPS-N RVs and its activity indicators: 0.54 with the S-index, 0.06 with the FWHM, and 0.52 with the bisector span, as plotted in Fig. 5.7. With the exception of the FWHM, the HARPS-N radial velocities correlate well with the indicators most commonly employed in stellar activity analyses. In particular, a visual inspection of the time series also shows that the BIS and the S-index are sensitive to the long-term trend of the magnetic cycle. These similarities between the RVs and

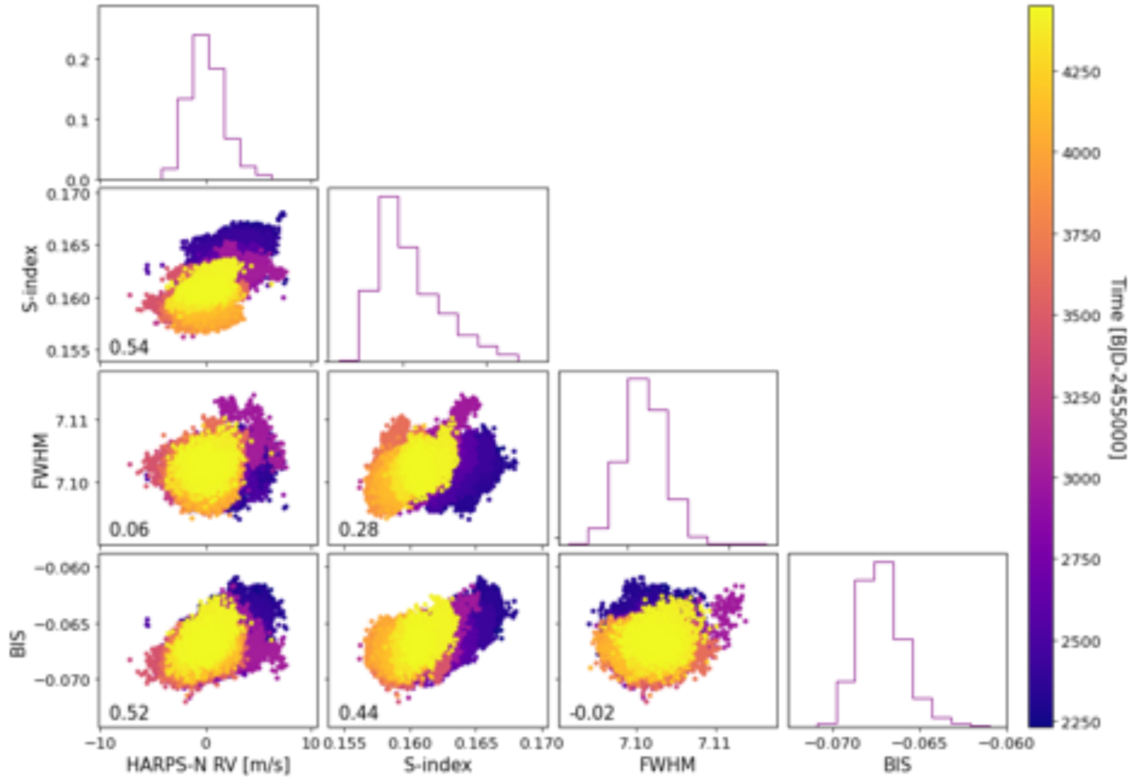

Figure 5.7: Correlation plots of the HARPS-N radial velocities and their activity proxies, S-index, full-width at half-maximum, and bisector span. The colour bar indicates the BJD of each datapoint. The Spearman Rank correlation factor for each set is also included.

the activity proxies are at the basis of many mitigation techniques. This good correlation however is not stable in time nor in activity level. In fact, during periods of minima the correlation becomes completely negligible, as addressed in Section 5.2.4.2. This behaviour can be attributed to the fact that these proxies are sensitive to a mixture of different active regions (spots, faculae, and network) (Cretignier et al. 2024). At low activity other effects not tied to active regions (and therefore not probed with traditional indicators) dominate the stellar variability (Lakeland et al. 2024). Nevertheless, I can now answer the first of our questions: B_l cannot be used as a direct one-to-one proxy to correct for stellar activity in the RVs over all timescales. The significantly worse correlation (especially when considering all levels of activity) between B_l and the ΔRV versus the one between the HARPS-N RVs and their proxies is to be expected after a simple visual inspection. As an example, B_l oscillates around 0 G, with an overall mean value of 0.02 G stable in time. The mean value of the ΔRVs changes with changing magnetic cycle phase, going roughly from 2.4 m s^{-1} at high activity, to 0.3 m s^{-1} during minimum.

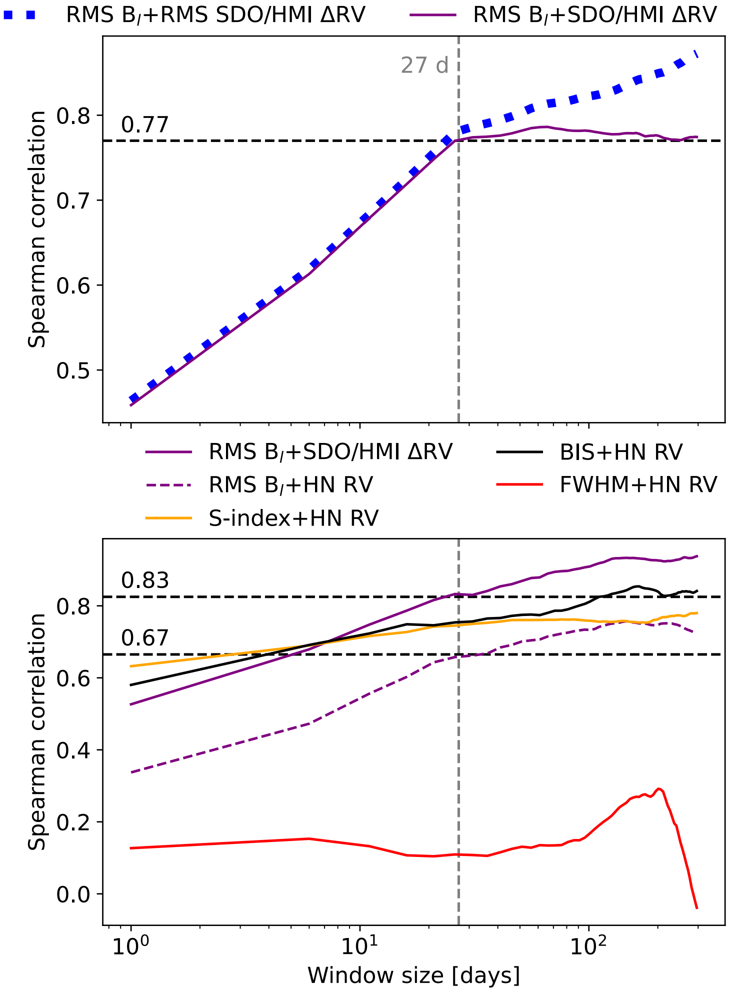

Figure 5.8: Spearman rank correlation coefficient between two time series against the size of the window (in days) used to smooth the signal (in logarithmic scale). *Top panel:* the correlation between the RMS of B_l and the RMS of the ΔRVs is plotted in blue dotted line, and between the RMS of B_l and the time-aware mean of ΔRVs is shown as a purple solid line. All considered time series are derived from SDO/HMI data and include all available observations. The knee of both plots is identified and its averaging window and correlation value are labelled. *Bottom panel:* Spearman correlation coefficient values plotted against varying window size. The time series considered have been matched following the method in Section 5.2.3.3. Colours represent, in order, the correlations between the RMS of B_l and the time-aware mean of SDO/HMI ΔRVs (solid purple), between the RMS of B_l and time-aware mean of the HARPS-N RVs (dashed purple), between the time-aware mean of the bisector span (black), the S-index (yellow), and the FWHM (red) with the HARPS-N RVs. The smoothing window equal to a solar rotation period is highlighted with a vertical gray dashed line. Horizontal black dashed lines indicate the correlation coefficient achieved when smoothing over a window of 27 days.

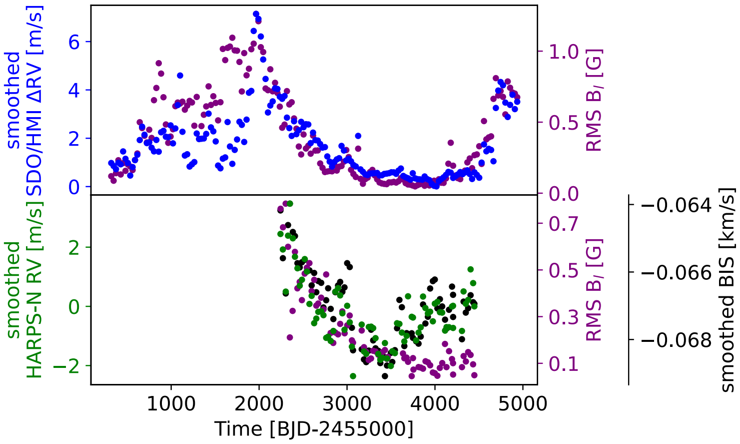

Figure 5.9: *Top panel:* the time series of the time-aware mean over an averaging window of 27 days of the SDO/HMI ΔRVs is plotted in blue. The RMS of B_l over the same window is also included in purple. *Bottom panel:* the time series of the time-aware mean of the matched HARPS-N RVs is plotted in green, of the matched bisector span in black, and the RMS of the matched B_l is shown in purple over the a window of 27 days.

I however notice a general trend shared between the B_l and the ΔRV time series through the solar cycle. I therefore investigate further. I postulate that, while direct measurements do not correlate, the root-mean squared scatter (RMS) of B_l may correlate to the general envelope shape of the RVs, and could therefore be useful information to model the long-term variations due to the magnetic cycle. To test this theory, I extract two new time series: I compute the rolling RMS of B_l over an "averaging window" of a day and the rolling time-aware mean of the SDO/HMI ΔRVs over the same window. The correlation between these new time series improves to 0.48. In order to find the best averaging window size, I repeat the same steps with window lengths between one day and one year. The results of this analysis are plotted in the top panel of Fig. 5.8 as a purple solid line. I also include the correlations between the RMS of B_l and the RMS of the ΔRVs for all the window sizes as a blue dotted line. Both correlations increase steadily until a window size of 27±1 days reaching a coefficient value of 0.77. At this point, the time series are not mapping the rationally-modulated variations, and are only sensitive to the overarching magnetic activity over the cycle, as shown in the top panel of Fig. 5.9. Therefore, the RMS of B_l over windows larger than the solar rotation period are able to successfully map the long-term variations in the ΔRVs. This time series can thus be used to correct for the long-term magnetic activity signal via techniques such as contemporaneous fit, or can be employed as a training set for a squared exponential kernel in a GP regression framework. As a simple test, I find the best-fit sine function to the RMS of B_l over a window of 27 days.

I then use the derived parameters to subtract the magnetic cycle long-term trend in the ΔRVs. This very rudimentary method is able to flatten the ΔRVs and reduce their RMS by more than 60%.

5.2.3.2 Structure Functions

The structure functions of all time series are also computed. The structure function measures the variability in a time series at each timescale (for a detailed explanation, see Chapter 2 Section 2.3.2). $\sqrt{\frac{1}{2}\mathrm{SF}}$ is used here to quantify the variability at each timescale, to better draw analogy with the RMS. The structure functions of the time series shown in Figs. 5.3 and 5.5 are plotted in Fig. 5.10. To ensure each time series is well sampled (i.e., with many pairs of observations contributing to each structure function calculation), at least 50 pairs of observations in each τ bin are required. To allow for direct comparison, I only consider the SDO/HMI data over the overlapping years with HARPS-N (2015 to 2021) instead of the whole time series. In Fig. 5.10, the higher cadence of HARPS-N is highlighted by the presence of data at shorter timescales, while the SDO/HMI-derived time series have a minimum τ of 4 hours.

Both RV time series (HARPS-N in green, and SDO/HMI in blue) have similar structure function behaviours. They both increase until a timescale of ~10 days, they then grow at a significantly slower rate (forming plateaus of sorts), to finally start increasing more strongly after 100 days, as the structure functions probe the solar rotation and activity cycle respectively. The RV RMS due to solar activity is of the order of 1 m s^{-1}. A somewhat similar behaviour is shown by B_l, with an initial increase until ~10 days, and a plateau at a $\sqrt{\frac{1}{2}\mathrm{SF}}$ of ~0.3 G (a low value expected for the extended minimum covered by the considered time series). Note that the structure function shows a slight decrease in this plateau region. I explain this behaviour by considering the magnetic cycle. Overall, the signal of B_l at comparable levels of stellar activity over different cycles is similar. That is to say that there are similarities between variations of B_l at the rise and at the decline of Cycle 24 versus at the rise of Cycle 25. The most interesting feature to notice in the structure function of the magnetic field is the significant dip at ~27 days (and a second smaller one at ~55 days). These dips highlight the strong modulation of the time series

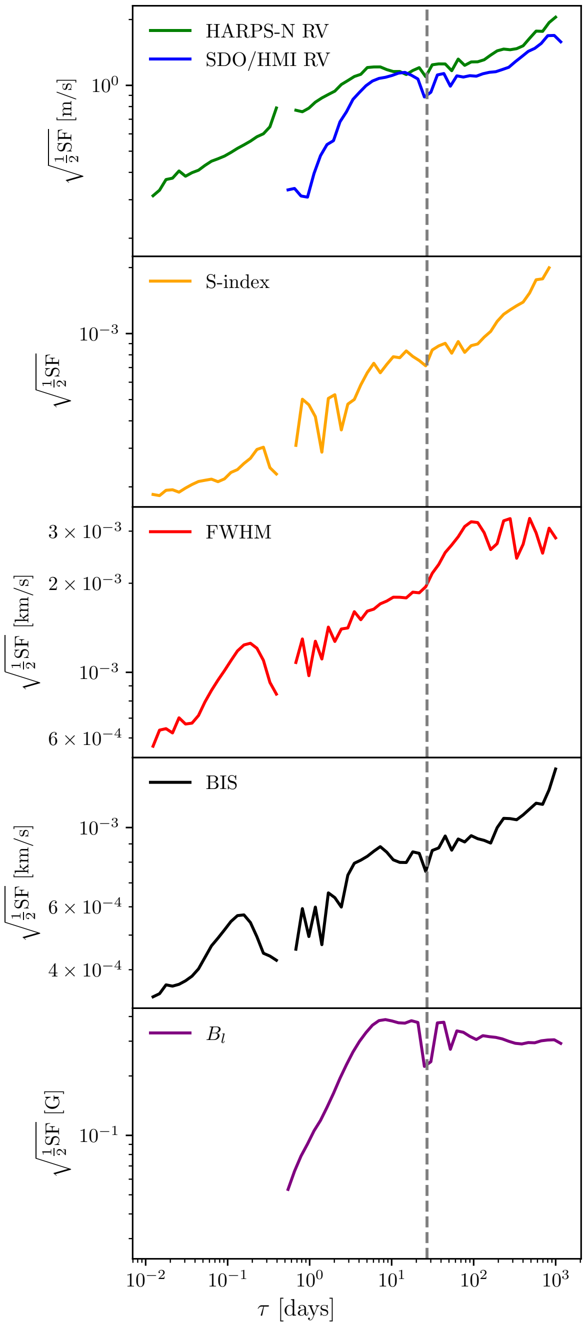

Figure 5.10: Structure functions of the time series shown in Figs. 5.3 and 5.5. See the main text for more details. From top to bottom: the structure functions for the HARPS-N and SDO/HMI ΔRVs (in green and blue respectively), the S-index, the CCF FWHM, CCF bisector span, and the mean longitudinal magnetic field. The higher cadence of the HARPS-N data is visible in the structure functions as the smaller minimum timescale. Likewise, the diurnal cycle of the ground-based observations gives rise to a gap in the structure function at ~ 0.5 days, since there are no pairs of observations separated by this timescale. The solar rotation period at 27 d is indicated by a grey dashed line. Figure taken from Fig. 7 of Rescigno et al. (2024) and produced by B. Lakeland.

at the solar rotation period. The SDO/HMI-derived radial velocities also show similar dips at one- and two-times the solar rotational period. While the HARPS-N RVs do share this feature, it is much less prominent. This is because the HARPS-N RVs are sensitive to variability from all processes on the Sun, whereas the ΔRVs calculated from SDO/HMI only consider the effect of active regions, which show a stronger rotational modulation. All three HARPS-N activity indicators have very similar structure functions, with a gradual but consistent increase at all timescales. They all present dips at ~1.5 and 2.5 days, which are not replicated in their radial velocities. The lack of a plateau region in any of the HARPS-N activity proxies means that no characteristic timescale of the variability can be retrieved: the activity indicators are affected by multiple physical processes all with different timescales. On the other hand, B_l shows a characteristic timescale of the order of half the solar rotation period, meaning that its behaviour can be sampled with two observations per period. This analysis therefore highlights the elevated complexity of the signal of the common activity proxies versus the mean longitudinal magnetic field, and is a first proof of the strong rotational modulation of B_l.

5.2.3.3 Matching the Data between HARPS-N and SDO/HMI

As the timestamps for the HARPS-N and SDO/HMI data are different, it is necessary to match the observations of the two time series in order to investigate the relationship between B_l and the RVs from HARPS-N. To do this, I interpolate the SDO/HMI data onto the timestamp of the nearest HARPS-N observation if the time between the two is less than one hour. If the time between an SDO/HMI observation and its closest match in the HARPS-N dataset is more than one hour, the data point is omitted. A justification for this approach is provided in the Appendix A1 of Rescigno et al. (2024). The resulting time series has 2,891 datapoints and includes the diurnal cycle and realistic poor-weather breaks from the HARPS-N data, and the maximum of six observations per 24-hour period of the selected dataset for SDO/HMI.

5.2.3.4 Correlation Analysis

I then reassess the correlation and recompute the Spearman coefficients between both the HARPS-N RVs and SDO/HMI ΔRVs, with all the considered activity tracers, including B_l,

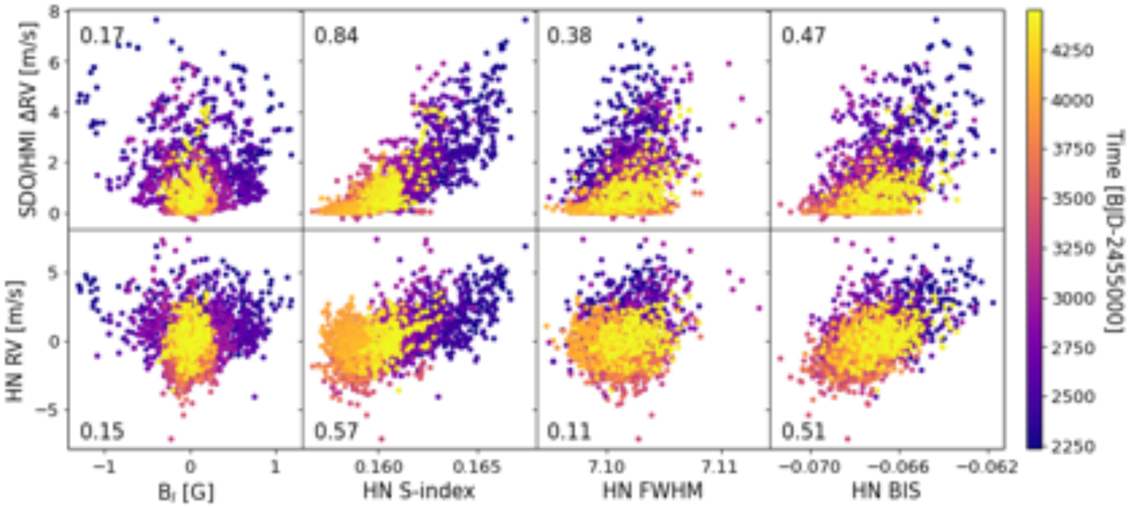

Figure 5.11: Correlation plots between the matched HARPS-N (here HN) and SDO/HMI time series. The SDO/HMI ΔRVs are in the first row, while the HARPS-N RVs are in the second. From the leftmost to the rightmost column, we plot the mean longitudinal magnetic field, the S-index, the full-width at half-maximum, and the bisector span. The data is colour-coded based on observation time and the Spearman rank correlation coefficients for each set are also included.

as shown in Fig. 5.11. As derived previously in Section 5.2.3.1, B_l does not correlate with either of the radial velocities. It is interesting to note, however, that correlations between the HARPS-N activity indicators and the RVs derived from the same instrument are lower than (or in the case of the BIS, comparable to) their correlation to the SDO/HMI radial velocities. As aforementioned, the SDO/HMI ΔRVs only include rotationally-modulated active region-induced RV variations, while the HARPS-N RVs include all other physical processes on the solar surface as well as instrumental systematics. These results highlight how these indicators are not able to successfully map the RV signals imprinted by processes such as granulation or supergranulation. I also replicate the same averaging window study undertaken in Section 5.2.3.1, as shown in the bottom panel of Fig. 5.8. I compute the correlations between the RMS of the matched B_l and the time-aware mean of the SDO/HMI ΔRVs with a rolling window size between one day and one year. The results are plotted as a purple solid line. With a window size of roughly the solar rotation period, the two time series reach a strong correlation of 0.83, reconfirming the earlier conclusion. For comparison, I also plot the correlations between the time-aware mean of the matched time series of the HARPS-N activity indices and their RVs. The FWHM (in red) does not map the long-term trend and therefore does not correlate well over all considered windows. On the other hand, the S-index and the BIS (respectively in yellow and black) reach similar correlations of 0.75 at a window size of ~27 days. Thus, the RMS of B_l over a solar rotation period correlate better to their smoothed ΔRVs than the HARPS-N activity proxies do to the smoothed RVs derived from the same instrument. As a further test, I also include the correlation with increasing rolling window size between the RMS of the matched B_l and the matched HARPS-N RVs, plotted as a dashed purple line. Unlike before, the correlation is lower, with it reaching only 0.67 at the solar rotation. To investigate this behaviour, in the bottom panel of Fig. 5.9 I plot the matched HARPS-N RVs and BIS time series smoothed over a solar rotation period (in green and black respectively) alongside the RMS of the B_l over the same window (in purple). It is clear that, while the RMS of B_l match the slow general decrease at the end of Cycle 24, the time series extracted from SDO/HMI diverge from the ones derived from the HARPS-N spectrograph around BJD 2,458,500, roughly the start of the extended minimum. The SDO/HMI ΔRVs follow the same shape as the RMS of B_l, instead of bending back up, as the HARPS-N data do. This different trend between HARPS-N and SDO/HMI can be caused by a variety of sources, the study of which is above the scope of this work. It is however important to note that this differing trend is partially enhanced in Fig. 5.9 by the axes chosen for the comparison.

5.2.3.5 Periodogram Analysis

I compute the Generalised Lomb-Scargle periodograms (GLS: Zechmeister and Kürster 2009) of all the time series for the complete and the matched datasets. Both produce similar results. For this analysis, I focus on signals with periods smaller than 100 days, as longer magnetic cycle periodicity would not be reliably picked up with the available baseline. In particular, I am interested in assessing the ability of B_l to systematically recover the solar rotation period. For these reasons, I remove all long-term signals with a low-pass filter. Via this comparison I am also able to confirm that no significant periodic signal is introduced in the data-matching step by interpolating the SDO/HMI data on the HARPS-N timestamps. In this work, I only include the periodograms for the matched datasets and plot them in Fig. 5.12. In the Figure, the Carrington solar rotational period of P_rot =27.2753 days as seen from the Earth is highlighted as a reference with a gray dash-dotted line. I also include the first and second harmonics of the Carrington period as dotted lines. By quick visual

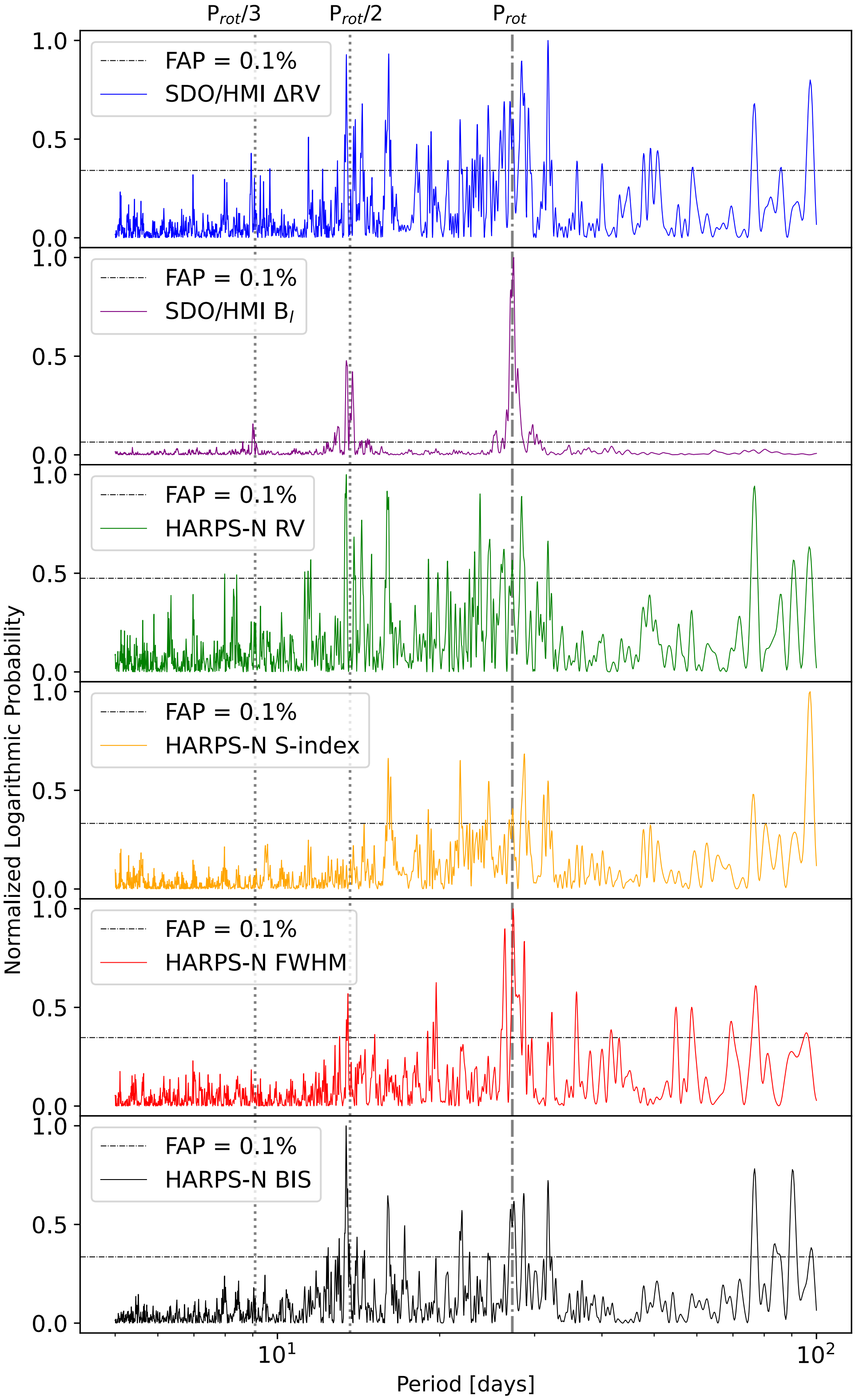

Figure 5.12: Generalised Lomb-Scargle Periodogram of the matched time series. On the x-axis the period in days, on the y-axis the normalised logarithmic Zechmeister-Kürster power (or probability). From top to bottom, the matched time series of SDO/HMI ΔRVs, and mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and bisector span. The Carrington Solar rotation period is indicated by a gray dash-dotted line. The first and second harmonics of the rotation are also highlighted by dotted lines. The False Alarm Probability (FAP) equal to 0.1% are included as dashed gray horizontal lines.

inspection, it is clear that the mean longitudinal magnetic field strongly outperforms all activity indicators in finding the expected rotational period. The only other relevant peaks in the periodogram of B_l are generated by the first and second harmonics of P_rot. This behaviour has been noted before for other B_l measurements (e.g., Kotov and Levitskii 1983; Grigor'ev and Demidov 1987; Obridko and Shelting 1992) and is similar to the results obtained by Xie et al. (2017) via wavelet transformation.

Both RVs are slightly more sensitive to the half rotational period than the full one, although they present wide forests of peaks at P_rot and $P_\mathrm{rot}/2$ both. The S-index has its most significant peak around 100 days, followed by one at ~29 days. Even considering all peaks above the 0.1% False Alarm Probability (FAP) level, the S-index does not reliably recover the solar rotation period. The FWHM is the most sensitive out of the HARPS-N proxies to P_rot, with a forest of peaks centred in ~29 days. It also shows peaks at $\sim P_\mathrm{rot}/2$, as well as ~19 days. The periodogram of the BIS is nicely peaked around $P_\mathrm{rot}/2$, with some signal around P_rot, as well as ~32 and 22 days. The periodograms of all time series excluding B_l are complex at high frequencies, have power at longer periods, and the true solar P_rot cannot be recovered to a reasonable level of accuracy or precision. In contrast, the periodogram of the mean longitudinal magnetic field is much simpler and does not present any significant power at long periods. As mentioned previously, all the power is concentrated at P_rot and its harmonics. In fact, given the formulation of periodograms, no power at low frequencies should be expected. Periodograms fit sinusoidal curves to the data for all periods and assess the goodness of the fit. Long-term effects can be fit by a sine curve in the radial velocities and its common proxies, but they behave differently in B_l. As an example, the magnetic cycle imprints on the RVs a general increase in their mean value over time of maximum and a decrease over times of minima. On the contrary, the average value of B_l stays constant in time. The magnetic cycle only affects the amplitude of the oscillations, not their mid-point, meaning that it cannot be fit by a long-period sine curve. This effect yields a much simpler periodogram. It can overall be concluded that a Fourier analysis of B_l is significantly more sensitive to P_rot and allows for a more precise and accurate identification of the solar rotation period than all other analysed time series.

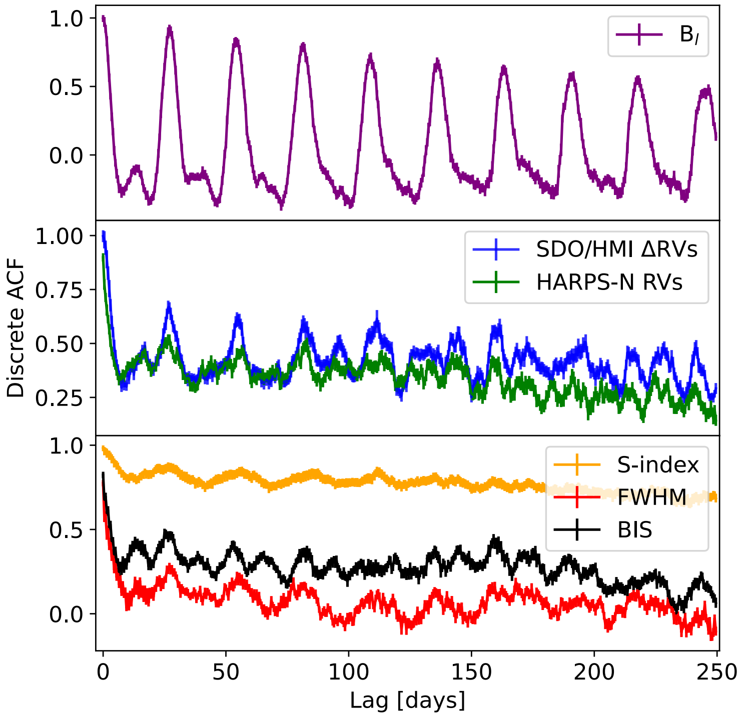

Figure 5.13: Autocorrelation function over a lag window of 250 days of the mean longitudinal magnetic field (top row in purple), the SDO/HMI and HARPS-N RVs (middle row in blue and green respectively), and the HARPS-N activity proxies S-index, full-width at half-maximum and bisector span (bottom row in orange, red and black). Uncertainties are included as errorbars.

5.2.3.6 Autocorrelation Function Analysis

Another way of isolating the rotational period is to compute the autocorrelation function of the time series (Giles et al. 2017; Collier Cameron et al. 2019). An autocorrelation analysis measures the relationship between observations at different points in time, and can therefore isolate patterns over the time series. I use the method developed by Edelson and Krolik (1988) and updated in Robertson et al. (2015) to compute the autocorrelation function (ACF) for unevenly sampled datasets. In very simple terms, we "slide" in time the data and compute how well it correlates to its original version via Pearson rank-order correlation coefficient (as described in Section 2.3.3). I use a code adapted from PYDCG[4]. Assuming significant rotational modulation, the solar rotation period can be extracted as the lag between each major peak in the ACF. I obtain the ACF for all matched time series, as shown in Fig. 5.13. As in Section 5.2.3.5, B_l is especially good at recovering the rotational period of the Sun, and its periodic signal stays strong and clear over multiple rotations. I compute the half-life of the autocorrelation to be 2.74±0.02 days. While not

---

[4] Available at: https://github.com/astronomercdamo/pydcf

wholly insensitive to the rotation period in this analysis, the HARPS-N RVs and the other proxies do not show as clear or well-peaked signals. As expected, the SDO/HMI ΔRVs present a smoother ACF signal than the HARPS-N ones, as they are derived with a model that considers only rotationally-modulated components.

5.2.3.7 Lag Analysis

Previous works have proposed or attempted to constrain the presence of time lags between the stellar activity proxies and the RVs (Boisse et al. 2011; Santos et al. 2014; Collier Cameron et al. 2019; Costes et al. 2021). I therefore also investigate for the presence of any significant lag between the mean longitudinal magnetic field and both radial-velocity time series. I use the same method as described in the previous section, this time computing the correlation between two different time series and shifting in time one with respect to the other. I limit the investigation to lag values between −30 and 30 days. The results of this analysis are plotted in the top panel of Fig. 5.14. I first examine the cross-correlation function between B_l and the total SDO/HMI ΔRVs, plotted in dark blue. When considering the entirety of the available 13 years of data, no significant lag can be found. A best correlation of 0.17 is recorded at ~ -7 days. As a next step I compute the cross-correlation functions between time-matched datasets. Starting from the matched SDO/HMI ΔRVs, I find a lag with B_l of -7.5 ± 0.5 days with a correlation coefficient of 0.47 ± 0.05. Similarly, the most probable lag between the matched B_l and HARPS-N RVs is found at -7.4 ± 0.5 days with a correlation coefficient of 0.32 ± 0.05. I also note second possible peaks for all RV time series at ~20 days. I interpret this as the repetition of the same lag in the "next" rotational period (assuming a $P_{\mathrm{rot}} \sim 27$ days). The derived best lag is comparable to a forth of the solar rotation, or roughly the difference between disk centre and limb. I note that, with the same time sampling and baseline, the SDO/HMI ΔRVs reach a higher correlation than the HARPS-N RVs for the same lag. In order to test whether this possible lag is driven by the presence of active regions, I also compute the cross-correlation between the matched B_l and the "quiet-Sun" RVs. This last time series is computed as the subtraction between the HARPS-N RVs (expected to include all processes) and the SDO/HMI ΔRVs (which only include active region-induced effects). It represents the RV variations caused by all physical processes on the Sun that are not directly tied to either the flux imbalance or the suppression of convective blueshift generated by the presence of large active regions. This method is justified in Lakeland et al. (2024). No significant lag can be extracted between B_l and the quiet RVs. These results point to the conclusion that active regions, such as spots and faculae, are the driving force behind the possible lag between B_l and the RVs.

To further investigate this, I also plot in red and green respectively in the bottom panel of Fig. 5.14 the cross-correlation function between B_l and the SDO/HMI ΔRVs during high activity (when the active region filling factors are maximised) and low activity (when active regions are few and far in between). Only the low activity ΔRVs show a clear lag with B_l at -8.5 ± 0.5 days with a 0.40 ± 0.05 peak correlation coefficient. The cross-correlation also peaks at $\sim$20 days, but differently from before, it also presents somewhat significant peaks at $\sim$9 days (and the related rotation peak at ~ -19 days). The high activity cross-correlation is as flat as the one between B_l and the full ΔRVs (also re-plotted in the bottom panel for comparison). These results seem to oppose the earlier conclusion. However, it is important to note that while at low activity the Sun does develop substantially less spots and faculae than during maximum, the surface is never fully bereft of them. In fact, even though the model to compute the SDO/HMI ΔRVs only considers the direct effects of large active regions, there is still some variability during minimum. I can therefore explain these results as follows: at high activity the larger number of active regions allows their longitudinal distribution to be significantly more even over the solar disc. Their contributions to a lag may therefore be "smoothed" away. On the other hand, during minimum active regions can more easily be approximated to a single cluster. This yields a "simpler" signal and the lag can be more successfully recovered. Similar results are derived when fitting quasi-periodic kernels to different sections of solar data, with the signal being smoother at solar maximum and increasing in complexity towards solar minimum (Klein et al. 2024). Further analysis is required to truly understand the source and the reason behind this lag. Nevertheless, this best-derived lag between B_l and the RVs is only minorly significant. In fact, the best-fit lag is not constant in time. To confirm this, I divide the SDO/HMI data in rolling 200-day chunks and find the cross-correlation coefficient at the best lag for each section. The computed lags for the 31,555 data chunks range from -8 days to 5 days. Roughly 80% of best-fit lags have correlation coefficients under 0.4. I retrieve no significant trends with time. The -7.5 days result is only recovered from

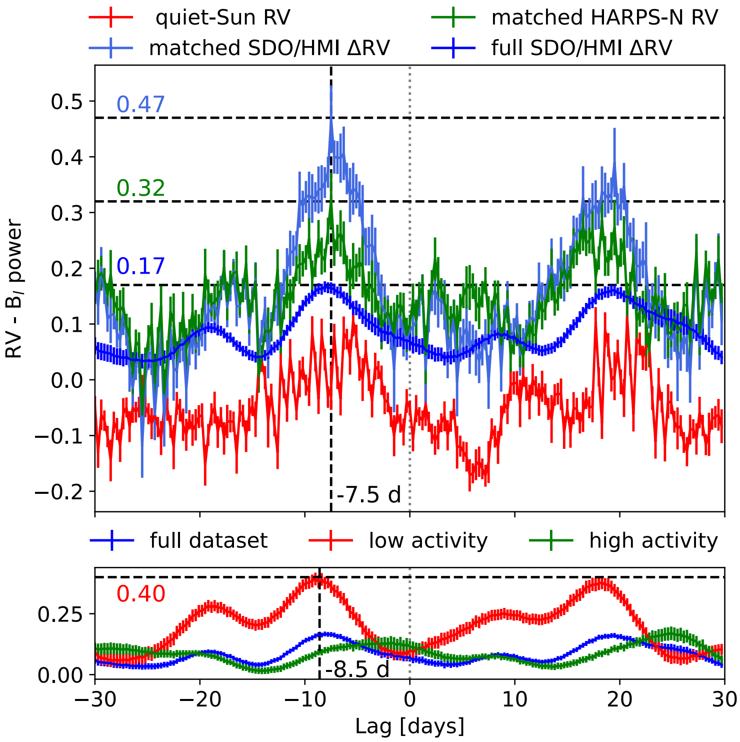

Figure 5.14: *Top panel:* lag plot between B_l and the RV time series. The lag against the matched SDO/HMI ΔRVs are plotted in pale blue, while the lag against the HARPS-N matched RVs are plotted in green. The lag between B_l and the full SDO/HMI ΔRVs time series is included in dark blue. The lag between the matched B_l and the "quiet-Sun" RVs is plotted in red, and is computed as the subtraction between the matched HARPS-N RVs and the active regions-derived ΔRVs from SDO/HMI. On the y-axis is the Pearson rank correlation coefficient computed between B_l and the time shifted RVs. Uncertainties on the power are included as errorbars. The best correlation achieved and the best-fit lag are highlighted by black dashed lines. The 0 lag is highlighted with a gray dotted vertical line. *Bottom panel:* lag plot between B_l and SDO/HMI ΔRVs. In blue, as in the top panel, the full SDO/HMI dataset, in red the low activity section of the same RVs (2015-Dec to 2021-Jan), in green the high activity section of the RVs (2010-May to 2015-Nov). The best fit lag of the low activity RVs and its respective correlation are highlighted with a black dashed lines. The 0 days lag is identified by a vertical dotted gray line.

the distributions of all the best-fit lags when considering solely correlation coefficients above 0.4.

Looking back to Fig. 5.5, a general visual inspection does point to an overarching possible longer time lag between B_l and the SDO/HMI ΔRVs, especially for the years 2013 to 2015, when the activity is at the highest. This behaviour has been noticed and investigated before: Sheeley and Wang (2015) showed that, for most observed magnetic cycles, the solar large-scale field undergoes sudden rejuvenation only after the Sun has reached its maximum. They, in fact, stated that a significant increase in the mean of the absolute B_l marks the start of the declining phase of the cycle. Therefore, this increase in amplitude is not directly tied to the increase of solar photospheric activity (measured by the number of sunspots and other active regions, as the maxima of the solar cycle is normally defined). It is instead caused by the longitudinal distribution of sources of flux, in particular a specific arrangement that (together with contribution from the axisymmetric element) reinforces the equatorial dipole component of the magnetic field. Since the migration and emerging patterns of active regions are expressions of the magnetic cycle of the Sun (Hathaway 2015), the rejuvenation of B_l is inherently tied with the stage of the cycle. Overall, this effect means that the increase in B_l, while correlated to the magnetic cycle, is not symptomatic of the same process as the increase of the variability in the radial velocities (which is primarily dependent on the filling factor of the active regions). The time delay of the maximum amplitudes of the mean magnetic field is therefore explained by the time required by the larger amount of active regions to migrate inward (Huang et al. 2017) and emerge in the "correct" arrangement. The length of this "lag" is not fully constrained yet, given the long baselines required to have enough data for a proper statistical approach, but it can be approximated to be of the order of months to a year. This significant correlation between the pattern of emergence of flux and the value of B_l can also inform us about the physical locations of the largest active regions on the surface of the star. In fact, Wang and Robbrecht (2011) found that the increased bias towards larger positive B_l values during 2014 was generated by a north-south asymmetry in the distribution of flux emergence. In the Sun, poleward surface flows maintain a north-south asymmetry in the photospheric field, which in turn generates an asymmetric quadrupole component. This means that at times, one sector of polarity can dominate over the other at the solar equator. The overall sign bias of the oscillations of B_l is therefore dictated by the leading polarity in the more active hemisphere. In this case, the wide positive amplitudes were induced by greater sunspot activity in the southern hemisphere of the Sun, as the southern wing polarity for Cycle 24 was positive (e.g., Norton et al. 2023). Differently from the RVs, the mean value of B_l and its evolution with time informs us about the leading polarity of the active regions, and in cases in which the magnetic field is better understood, they inform us about the hemispheric positions of the active regions. In time series of stellar observations, this information could also be employed as further constraints in Zeeman Doppler Imaging (Semel 1989; Brown et al. 1991; Piskunov and Kochukhov 2002).

5.2.4 Stellar-like Observations: can we use B_l to measure P_{rot}?

In Section 5.2.3, I have proven that with high cadence and a long baseline, the mean longitudinal magnetic field is the ideal period detector due to the strength and the simplicity of its signal. However, the value of a good activity tracer is its ability to inform us about the stellar variability successfully over much shorter timescales. Is B_{l} as good as a rotational period detector with larger uncertainties and with significantly less data, as is the case with most stellar datasets? Can B_{l} be relied on over all levels of magnetic activity, or will it fail at low activity, as do most of the other common activity proxies? I therefore test the mean longitudinal magnetic field as a "stellar" activity tracer. I limit the data to the average stellar season length, roughly 100 days. I also select three chunks of data over the available years in order to test the effectiveness of B_{l} over multiple phases of the solar magnetic cycle, at highest, medium and lowest activity covered by the HARPS-N dataset.

5.2.4.1 Choosing a Realistic Stellar-like Cadence and Precision

In order to represent a typical observational schedule for a star in the context of exoplanet detection, I pass through a second data selection process. A typical cadence for stars is maximum of an observation a night. I therefore select only one observation taken each day of data. I do not average all datapoints to daily bins, as that would get rid of effects such as granulation and it would not be representative of the type of observations undertaken for stars. I instead randomly select one observation over each 24 hour window. In this analysis I do not account for the difference of integration time per exposure. The 5-minute exposure length of HARPS-N solar data is long enough to average out p-modes, and all other physical process that can significantly influence the RV variations (e.g., supergranulation) have baselines longer than the average exposure time of stellar observations. I select three 100-day chunks over three stages of stellar activity. A high stellar activity case is selected for BJD 2,457,235 to 2,457,335 (2015-Ju-31 to 2015-Nov-8), close to the start of the HARPS-N solar observations, at the highest currently observed activity level. A medium activity case is selected for BJD 2,457,716 to 2,457,816 (2016-Nov-23 to 2017-Mar-3). A low stellar activity case is selected during the extended minimum of Cycle 24 for BJD 2,458,950 to 2,459,050 (2020-Apr-10 to 2020-Jul-19). The selected chunks are shown in Fig. 5.15 with vertical dashed lines. From here onward, only the observed HARPS-N

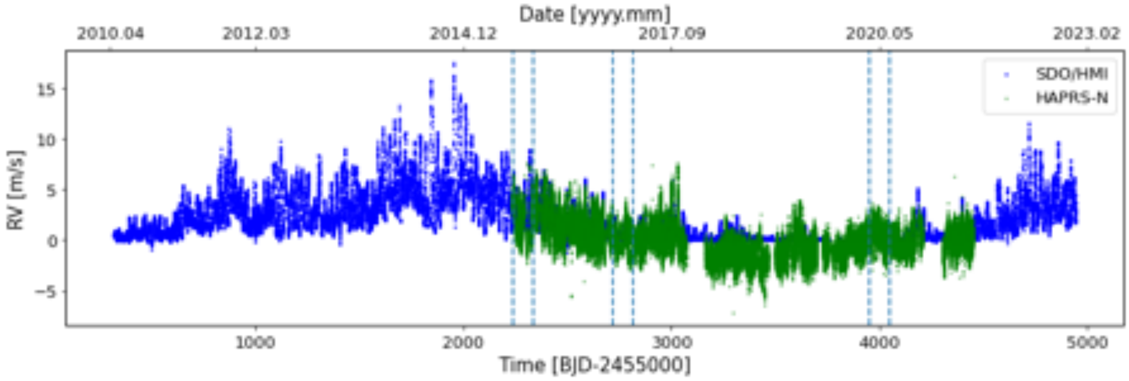

Figure 5.15: The selected stellar season-like 100 days chunks for the analysis in Section 5.2.4 are shown by vertical dashed lines. The SDO/HMI radial velocities are plotted in the darker blue, and the HARPS-N RVs in the lighter green. Uncertainties are included but may be too small to be visible.

radial velocities will be considered in the analysis. As a reminder, given the matching method summarised in Section 5.2.3.3, bad-weather breaks are already included. I note that this already significantly reduced dataset will very likely still represent an ideal stellar cadence. The Sun is still observed even with predicted SNR values down to 200. This is not the case with EPRV targets. Although the two cuts mentioned in Section 5.2.2.1 will eliminate data taken under not ideal conditions, these requirements are still more relaxed than what would be expected of an EPRV target. Moreover, I have not considered the possibility of telescope time competition. At the TNG, every hour of light is dedicated uniquely to the solar telescope. Conversely at night multiple programs are competing for time. It is therefore unlikely for a telescope with multiple programs to be able to achieve the "once-a-day" cadence here selected over the entire season. Nevertheless, I have reduced the dataset considerably to a cadence similar to what new missions such as the Terra Hunting Experiment (THE: Thompson et al. 2016) are aiming to achieve.

As mentioned in Section 5.2.2.2.1, the uncertainties on the mean longitudinal magnetic field derived directly from SDO/HMI errors are very small. No existing or planned polarimetric survey of far-away stars could reach those levels. To better represent the stellar case, I instead inflate the uncertainties of B_l to the best achieved uncertainty level of 0.2 G on fully detected mean longitudinal fields for Sun-like star by the BCool collaboration (Marsden et al. 2014; Mengel et al. 2017). This is an optimistic floor that has been proven to be achievable by polarimetric observations of Sun-type stars before, and it is the precision level new spectropolarimetric instruments, such as the one currently in construction for the upcoming HARPS3, aim to achieve. I therefore use a constant error on B_l measurements of 0.2 G. To match this uncertainty and to truly represent the inflated error, I also inject into the dataset white noise randomly extracted from a Gaussian distribution with a FWHM of 0.2 G.

5.2.4.2 Preliminary Analysis

High Stellar Activity Case: I start with the chronologically-first 100-day chunk: the high stellar activity case. Over this window, I match 69 datapoints. The selected data is plotted in Fig. 5.16. Even with a quick visual inspection, a clear periodic signal can be identified in the mean longitudinal magnetic field. Similarly to Section 5.2.3, I compute the correlation between each considered activity indicator and the HARPS-N RVs, shown in the first row of Fig. 5.18. At this stage the solar activity is strong and dominated by rotationally-modulated effects, as highlighted by the high correlation between the RVs and the indicators S-index, FWHM and bisector span. I also plot the Generalised Lomb-Scargle periodograms of the all the time series for this chunk in Fig. 5.17. In this case, as expected, all HARPS-N proxies and to a lower degree the radial velocities themselves have power at the solar rotation period. Once again B_l is sensitive to the P_{rot} and $P_{\mathrm{rot}}/2$ signals.

Medium Stellar Activity Case: A similar analysis is then undertaken for the medium activity case. Over this window I match 47 datapoints. I plot the derived time series in Fig. 5.19. The correlation relationships between the RVs and the activity indicators are plotted in the second row of Fig. 5.18. The computed Spearman rank correlation coefficient are now significantly lower for all proxies and no correlation above 0.5 can now be found. I can postulate that most of the rotationally modulated effects are now reduced in significance with respect to other photospheric and chromospheric variability. As before, I also perform a GLS periodogram analysis, as shown in Fig. 5.20. The activity signal is now not strong enough to be picked out from a periodogram analysis of the RVs only, but it is present in most of the investigated indicators. The FWHM retrieve the rotation period to a False Alarm Probability of 1%. The S-index and B_l are the only ones that present power at the rotational period over the 0.1% FAP level. It is of note that all HARPS-N activity

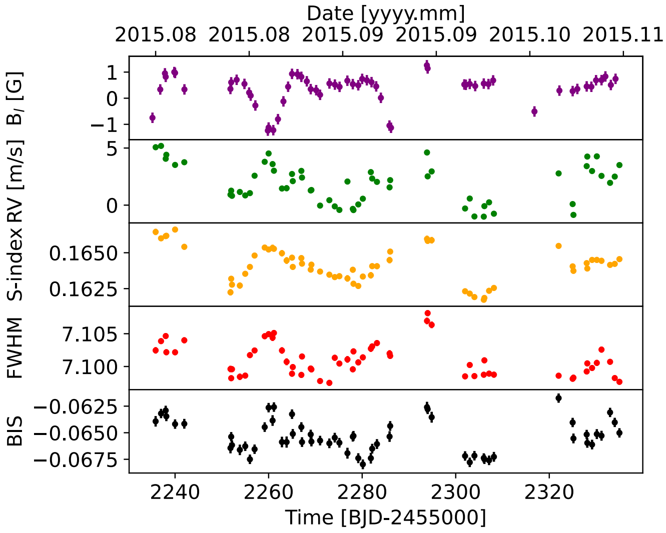

Figure 5.16: High activity dataset selection. From top to bottom: mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and bisector span. Some uncertainties may be too small to be clearly visible.

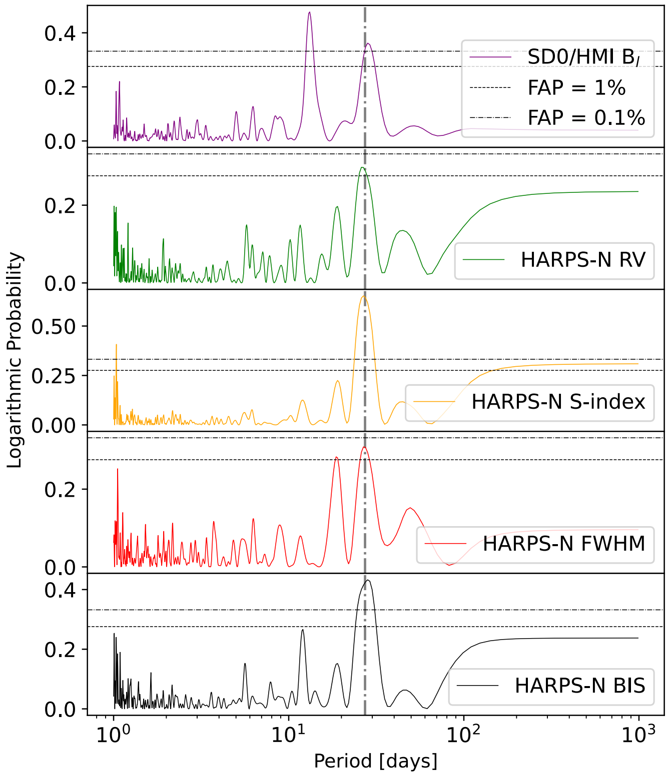

Figure 5.17: GLS periodograms of the high activity data. From top to bottom: mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS. 1% and 0.1% False Alarm Probabilities are shown as dotted and dashed gray lines. The vertical dash-dotted black line highlights the Carrington solar rotational period.

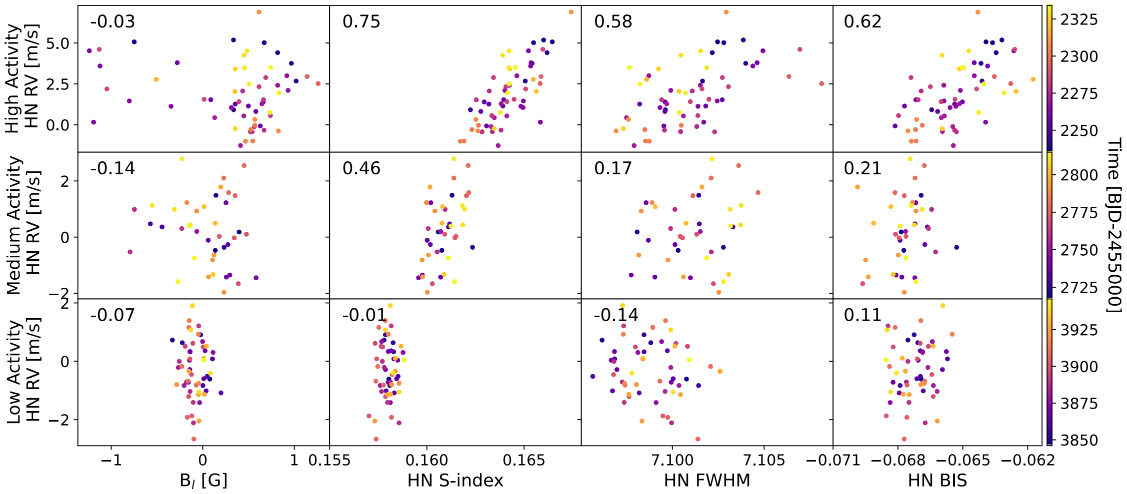

Figure 5.18: Correlation plots of the (from top to bottom) high, medium and low activity data selection of the HARPS-N (here HN) radial velocities against the considered activity proxies. The data is colour-coded based on date. The Spearman rank-order correlation coefficient of each pair is also included.

indicators are now also presenting a peak at ~40 days (not an harmonic of the rotational period or one of its aliases). This peak exceeds the 0.1% FAP level in the S-index, the most reliable of the common proxies in the previous analysis, making period determination only based on its information trickier. Overall at medium activity, B_l already starts to outperform other proxies in this preliminary analysis.

Low Stellar Activity Case: I repeat the same analysis once more with the last selected dataset over the extended solar minimum. I match 79 datapoints. The extracted time series are plotted in Fig. 5.21. At this level of activity the considered uncertainty of 0.2 G is comparable to the B_l signal itself. There is now no correlation between any of the activity proxies and the radial velocities, as shown in the bottom row of Fig. 5.18. Most of the rotationally-modulated effects are now overshadowed by other sources of activity. Even a periodogram analysis yields no information regarding the periodicity of the Sun. Most of the signal seems to in fact be aperiodic, as illustrated in Fig. 5.22.

5.2.4.3 Gaussian Process Regression Analysis

The usual next step in the analysis of radial-velocity data, especially in cases with high correlation between indicators and RVs is to employ Gaussian process regression to model the activity in the stellar proxies, as I have done before in Chapter 4 Section 4.2.6. This is done in order to identify the hyperparameters that better fit the stellar signal, which

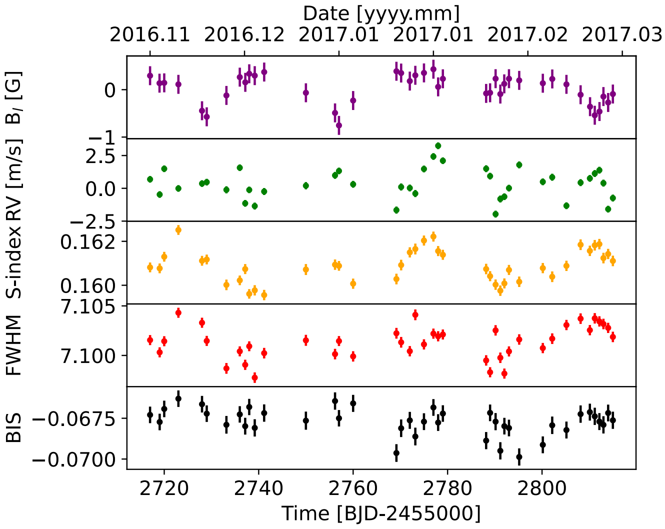

Figure 5.19: Medium activity dataset selection. From top to bottom: mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS. Some uncertainties may be too small to be clearly visible.

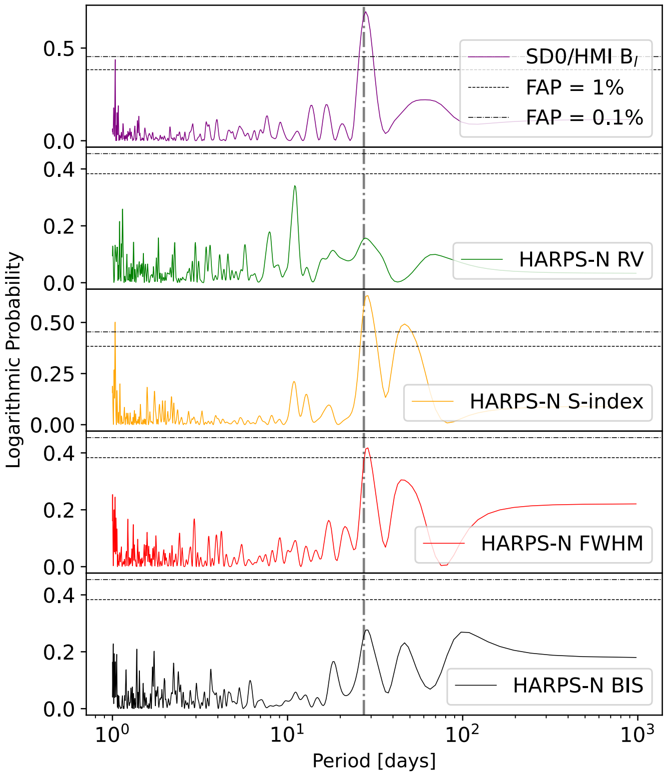

Figure 5.20: GLS periodograms of the medium activity data. From top to bottom: mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS. 1% and 0.1% False Alarm Probabilities are shown as dotted and dashed gray lines. The vertical dash-dotted black line highlights the Carrington solar rotational period.

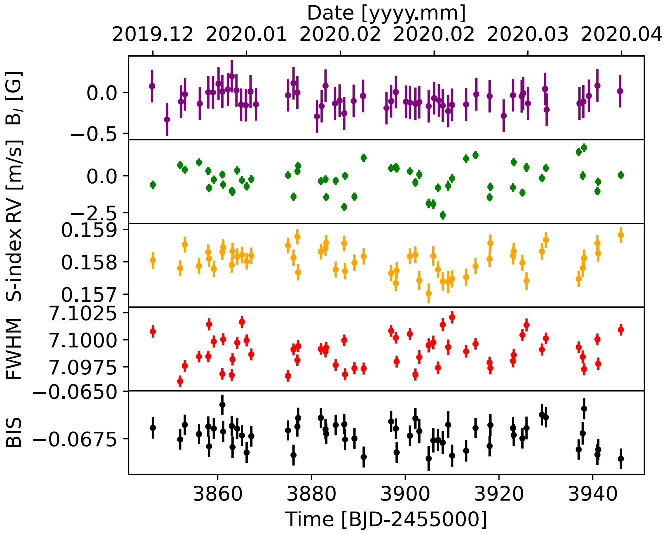

Figure 5.21: Low activity dataset selection. From top to bottom: mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS. Some uncertainties may be too small to be clearly visible.

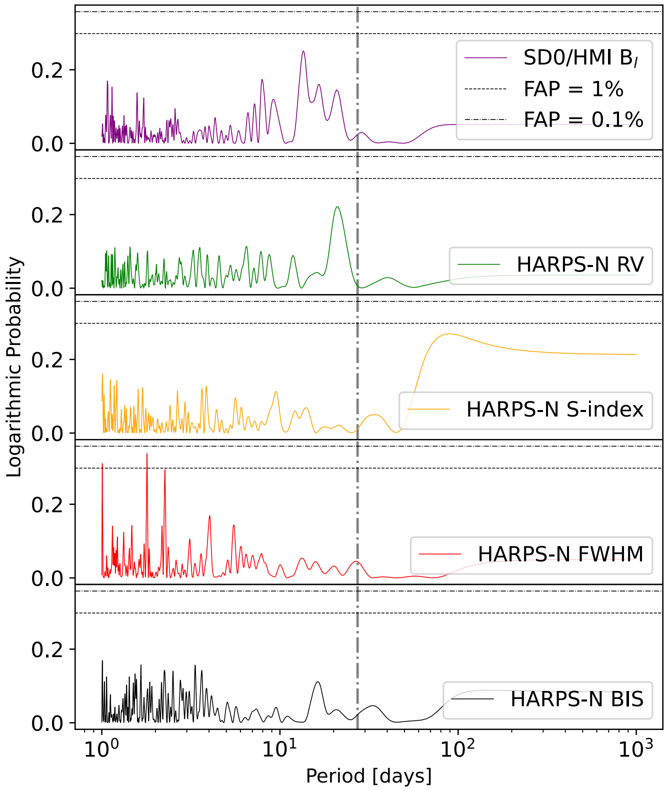

Figure 5.22: GLS periodograms of the low activity data. From top to bottom: mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS. 1% and 0.1% False Alarm Probabilities are shown as dotted and dashed gray lines. The vertical dash-dotted black line highlights the Carrington solar rotational period.

can then be used to inform priors in a second GP analysis of the RVs themselves. In this work, I undertake the most uninformative Gaussian process regression analysis in order to simulate a preliminary stellar rotation period search, or a "worse case scenario", in a typical exoplanet detection. To model the stellar activity, I use a Quasi-periodic kernel with an added white noise "jitter" term, as described by Eq. 4.2 in Chapter 4 Section 4.1.6.

The QP kernel has been successfully employed to model stellar activity in both radial-velocity (e.g., Rajpaul et al. 2015; Barros et al. 2020) and stellar activity proxy analyses (e.g., Haywood et al. 2014; Grunblatt et al. 2015). In this work, I test whether a similar analysis could be undertaken with B_l and whether it could be more successful than the same study done on other activity proxies. To do so I once again use the pipeline MAGPy-RV. I run the same analysis for all time series: B_l, HARPS-N radial velocities, S-index, FWHM and bisector span. All hyperparameters are bound by forced positive (larger than 0) uniform priors. The harmonic complexity μ is bound by a uniform prior between [0,1]. I also bind both the period $P_{\rm rot}$ and the evolution timescale l with uniform priors between [0,100] given the length of the selected window. I bind the white noise β with a Gaussian prior centred in the mean value of the uncertainties of the considered dataset and of width equal to 25% of said value, in order to avoid the GP explaining all the variability in the form of white noise. These priors represent the amount of information I am able to derive from initial analysis in the low activity case. For ease of comparison, I use the same priors in all runs. For all analyses, I simultaneously evolve 200 chains over 50,000 iterations each, with a discarded burn-in phase of 10,000 steps. I assess the convergence of the chains by computing the Gelman-Rubin statistics and define a chain as converged only under a 1.1 convergence cut. Not all chains are able to converge with the described priors. Instead of aiming for full convergence, I select a number of iterations with which most of the parameter space for all datasets is investigated, and with which all B_l chains are fully converged.

In this work, I focus on the hyperparameters useful for subsequent radial-velocity fitting: the period of the solar activity signal, its evolution timescale and its harmonic complexity. The amplitudes and jitters of each time series are not comparable. I plot the posterior distributions after MCMC analysis of each considered hyperparameter in Figs.

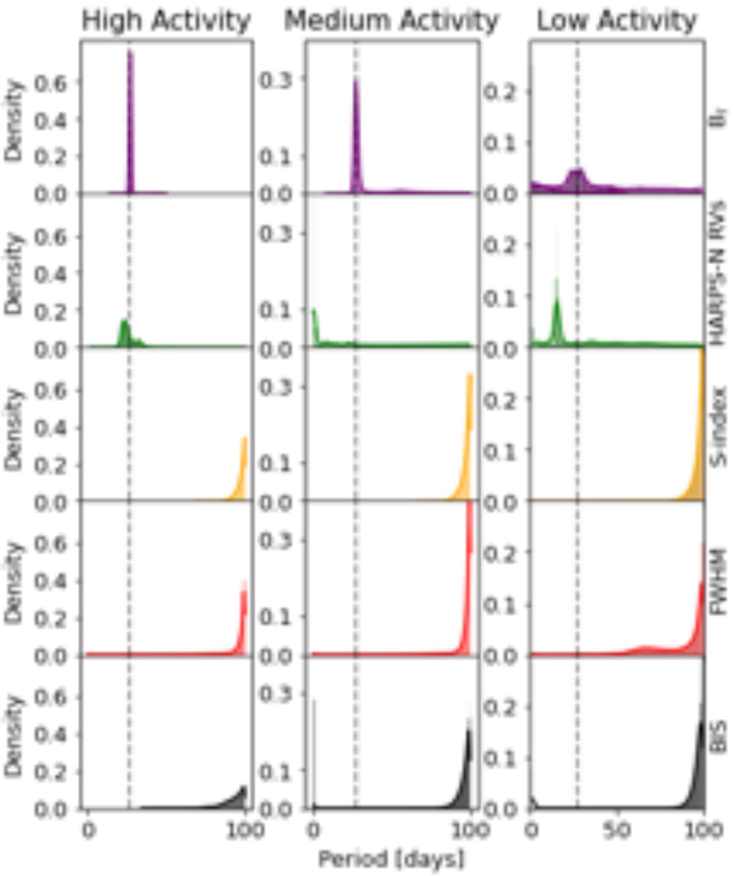

Figure 5.23: Collection of posteriors for the period P_{rot} of the Quasi-periodic kernel after GP regression. From left to right we consider the high, medium and low activity cases. From top to bottom we see the posteriors of mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS in their respective colours. The Carrington solar rotation is here highlighted with a black dashed line. Note the shared y-axis for each column.

5.23, 5.24 and 5.25. The high solar activity case is shown in the first column, the medium is in the middle column, and the low activity case is in the third column. Each time series is plotted on a different row.

The Rotational Period: I start the assessment from the period of the solar activity, in Fig. 5.23. For the highest activity case, the mean longitudinal magnetic field far outperforms all other proxies and the RVs themselves in identifying the "correct" solar activity period (here defined again by the Carrington solar rotational period and shown by a dashed vertical line in the figure). Therefore, even in the case in which B_l is comparable to other common proxies in a simple Fourier analysis, the mean longitudinal magnetic field gains an edge in a GP regression framework. At medium activity, B_l is still able to cleanly converge for the expected value. Even during the prolonged minimum, although not to high precision, B_l is the only time series able to identify the solar rotational period, and the radial velocities are only sensitive to its first harmonic. When looking at all the posterior results together, it is clear that only B_l is consistently successful at recovering the solar rotation period. To do so, it requires little to no prior information, making it more versatile, and it can converge much quicker than any other proxy, lowering the computational expense. In all cases, the HARPS-N activity proxies are unable to converge for any periodicity and instead their posteriors peak at the top of the time window available for exploration: 100 days. They model the activity in the data as a long period (longer than the dataset) with shorter evolution timescale l and higher h. They therefore "assign" more of the signal to other time-dependent hyperparameters. This is another confirmation of the sensitivity of B_l to the solar rotation period. Radial velocities and their spectra- or CCF-derived proxies rely on surface features and limb darkening modulation to pick up the rotational period. On the other hand, the mean longitudinal magnetic field extracted with spectropolarimetric observations is not only affected by limb-darkening and foreshortening. Its change in intensity with rotation is also exacerbated by the fact that we are observing the line-of-sight component of the radial field, which will be at the largest when the field is pointing directly at the observer and will approach zero when rotating perpendicular to the line of sight, all together yielding a larger and clearer modulation in the signal.

The Evolution Timescale: Similarly to the period, I plot the posterior distributions of the evolution timescale l in Fig. 5.24. All common activity proxies as well as the mean longitudinal magnetic field prefer longer evolution timescales than the radial velocity. The l posterior distributions of the RVs peak at values comparable to the rotation period. This result is also supported by previous analyses (e.g., Camacho et al. 2023). The longer timescale recovered by B_l does not oppose the result of the RV analysis. In fact, even assuming that the main source of the variations in B_l is the magnetic flux in the active regions, the magnetic fields concentrated in said active regions have been shown to have longer lifetimes than their photometric expression in the form of sunspots or faculae (which are the source of the RV variations). A nascent active region (before any photometric brightening or dimming of the solar photosphere) is an ensemble of small-scale emergence events with a preferred magnetic orientation (Strous and Zwaan 1999). Af-

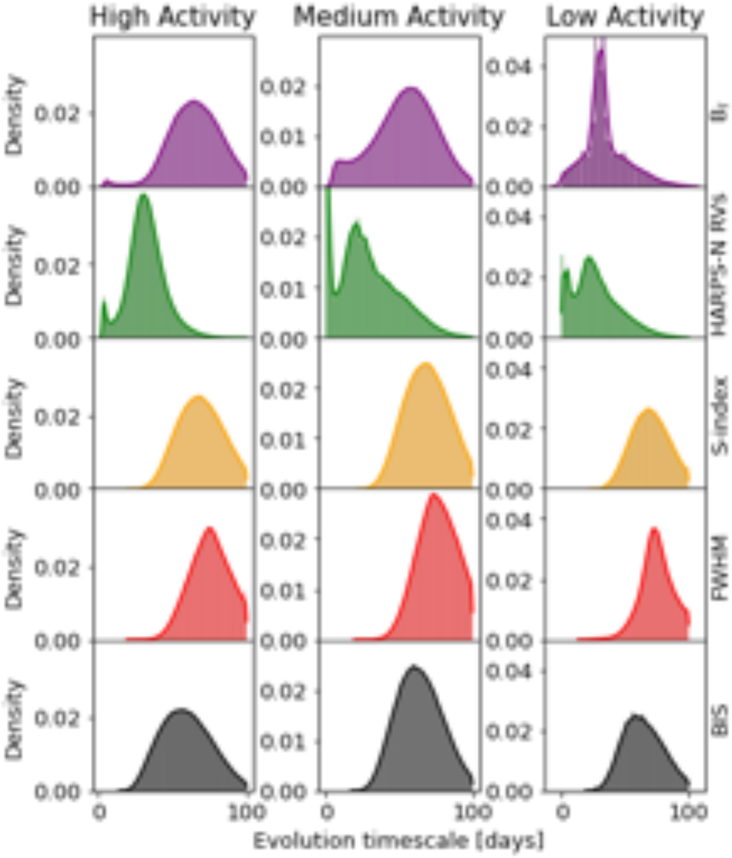

Figure 5.24: Collection of posteriors for the evolution timescale after τ of the Quasi-periodic kernel GP regression. From left to right we consider the high, medium and low activity cases. From top to bottom we see the posteriors of mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS in their respective colours. Note the shared y-axis for each column.

ter the emerging of magnetic field concentrations, the Ca II intensity begins increasing, usually with a time lag in the order of tens of minutes (Bumba and Howard 1965), and convective collapse begins. It is then reasonable for the S-index also to converge to comparable evolution timescales to B_l. Opposite magnetic polarities separate after 24 hours and areas of the same polarity migrate towards each other to coalesce into larger features such as pores (Van Driel-Gesztelyi and Green 2015). With increasing total field and as further areas migrate and conglomerate, pores evolve in photometric active regions in the form of spots and faculae (Cheung et al. 2017). With time (and with a lifetimes roughly of 15 to 60 days) photometric active regions gradually disappear. In this process opposite polarity fragments magnetically reconnect and the flux slowly cancels itself. As coronal heating decreases, the plages start dimming. Finally, the magnetic active region dissipates into the magnetic background. Just as the lifetime of spots and faculae depends on their size, the overall lifetime of magnetic active regions is proportional to the magnetic flux they reach at maximum development (Van Driel-Gesztelyi and Green 2015). The results of l in B_l can therefore be reliably employed as upper bounds in following RV GP regression analyses.

Harmonic Complexity: When looking at the posterior distributions for the harmonic complexity, shown in Fig. 5.25, B_l is consistently in better agreement with the radial velocity than other proxies. The dashed vertical line indicates an harmonic complexity equal to 0.5 (meaning that all complex features on the surface can be reduced to two active regions on opposite sides of the solar sphere rotating in and out of view). This result is in line with the conclusions in Jeffers and Keller (2009). From the formulation of the QP kernel in Eq. 4.2, a higher value of μ means a smoother curve in-between periods, or a lower inner-period complexity. These results seem to contradict the conclusions of the analysis in Section 5.2.3, that the mean longitudinal magnetic field exhibits less complex signals than the other proxies. However first, the posterior distributions in Fig. 5.25 need to be considered within the larger context of the Gaussian process analysis. I cannot do a direct comparison of the extracted best-fit parameters between B_l and the HARPS-N activity proxies, as the latters were not able to recover the "correct" period and are therefore modelling the activity in a completely different manner. For example, I note that for all levels of activity, the most probable jitter term describing the best-fit GP function to the

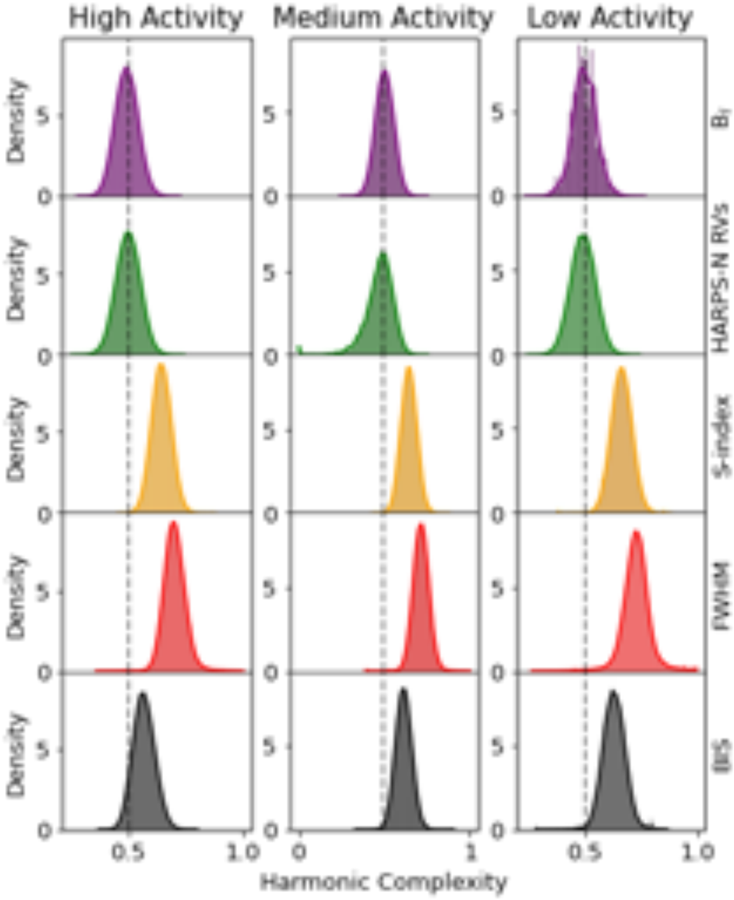

Figure 5.25: Collection of posteriors for the harmonic complexity μ of the Quasi-periodic kernel after GP regression. From left to right we consider the high, medium and low activity cases. From top to bottom we see the posteriors of mean longitudinal magnetic field, HARPS-N radial velocities, S-index, FWHM, and BIS in their respective colours. The black dashed line highlights the 0.5 harmonic complexity. Note the shared y-axis for each column.

activity indicators after MCMC optimisation is consistently larger (ranging between twice to 20 times as large) than the average uncertainty in the corresponding time series. This is not the case for B_l. I therefore postulate that a significant part of the signal in the proxies is not being modelled by the GP at all and it is instead accounted for by the large jitter.

Overall, with this analysis I show that the mean longitudinal magnetic field is a great rotational period detector. It is more effective than the RVs themselves or all other considered activity proxies, as it consistently outperforms them over all solar activity levels. It is more efficient than the other time series, as it requires the least amount of prior information to converge the the "correct" value and needs the shortest computational time. A GP regression analysis of B_l is not only useful to find or confirm the period of the quasi-periodic variations, but the results of other hyperparameters can also inform a second GP analysis of the RVs themselves. The harmonic complexity posterior of the mean longitudinal magnetic field can be used as a prior for the RVs, as I have proven that they are in agreement over all activity levels. Moreover, the evolution timescale derived for B_l can inform the upper bound of the same hyperparamter for the RVs.

5.2.5 Conclusions

In this work I analysed the solar mean longitudinal magnetic field as a rotational period detector and as a tracer for the mitigation of activity-induced variations in RV surveys in the context of exoplanet detection. I considered the longitudinal magnetic field extracted from SDO/HMI observations alongside the ΔRVs derived with a model from the same data. I performed correlation analysis, I computed their structure functions, Generalised Lomb-Scargle periodograms and autocorrelation functions, and I tested for the presence of any time lag between the two time series. In parallel I duplicated all the analyses with Sun-as-a-star observations taken by the HARPS-N spectrograph (I considered the HARPS-N derived radial velocities, the S-index, and the full-width at half-maximum and the bisector span of the CCF). I find the following:

- B_l does not directly correlate to the RVs. This lack of correlation is not activity-level dependent. B_l cannot therefore be employed as a direct proxy of the solar activity in the radial velocities. I however find that the time series of the RMS of B_l computed over a window size comparable to the solar rotation period does correlate well with the RVs smoothed over the same amount of time. With a rudimentary sine function fitted to the RMS of B_l and subtracted from the SDO/HMI ΔRVs, I am able to reduce the radial velocity scatter by more than 60%. B_l can therefore be used to successfully model out the long-term RV signal due to the magnetic cycle of the Sun.

- B_l has a significantly simpler structure function than all other considered time series, with a characteristic timescale of ~10 days.

- B_l is an effective solar rotation detector. Even when the same cadence and baseline are considered between the SDO/HMI and the HARPS-N data, the periodogram of B_l only presents peaks at the Solar Carrington rotational period and to a lesser extent to at its first and second harmonics. None of the other considered proxies or either of the radial velocities are as simple, and in most cases no clear rotation period can be isolated with a Fourier analysis. This point is further confirmed by an autocorrelation analysis, in which the rotation period signal of B_l stays clear and strong over multiple rotations.

- A lag analysis is performed and a minorly relevant lag between B_l and the RVs is found at roughly −7.5 days. This lag appears to be driven by the signal generated by active regions. These results, however, change significantly based on which section of the solar RVs are considered and based on the level of the magnetic cycle.

Overall, I therefore have proven that with high cadence and a long baseline, the mean longitudinal magnetic field is a very effective solar rotational period detector, and it can be used to inform the understanding of the physical processes happening on the surface of the Sun. This is, however, not representative of the type of observational time series taken for exoplanet detection. Therefore, I also test the B_l as a "stellar activity tracer". I degrade the time series to 100-day windows, with a single observation per night, and inflate the uncertainties in B_l to those achieved by previous polarimetric surveys. I then perform a typical preliminary analysis followed by Gaussian process regression with a Quasi-periodic kernel. I perform the same analysis for three chunks of data over high, medium and low activity levels. I find the following:

- The mean longitudinal magnetic field starts outperforming the other activity indicators in a preliminary periodogram analysis from the medium activity level.
- After one-dimensional GP regression, B_l is the only time series that is able to successfully recover the "correct" rotational period over all levels of activity. It does so with the shortest convergence time and with little to no prior information.
- B_l recovers a longer evolution timescale than the RVs.
- The best-fit harmonic complexities of B_l and the RVs strongly agree within uncertainties.

With this analysis, I have re-confirmed the mean longitudinal magnetic field as an effective and efficient rotational period detector, with exoplanet-survey-like time series and over all levels of solar activity. The best-fit values extracted from the posteriors of the other hyperparameters can be used as prior information for a follow-up RV GP regression. This work also highlights the need of time series of polarimetric data for less magnetically active stars, for more fields than simply exoplanetology.

Chapter 6

The Conclusion

> *"If you torture the data long enough, it will confess"*
>
> — *Ronald Coase*

6.1 Concluding Remarks

The radial-velocity method is one of the most promising avenues for the detection of exoplanets. With the recent sub-m s^{-1} precision reached by ultra-stable spectrographs such as HARPS-N, a new horizon of rocky planets in long-orbits has become within reach. However, the greatest challenge for the detection and characterisation of exoplanets is now stellar variability. Activity on the surface of stars often strongly dominates the RV budget, and it can easily obscure or mimic a Keplerian signal. Chapter 1 addresses and describes the main sources of stellar variability over all timescales relevant to radial-velocity surveys. Dramatic events such as flares and coronal mass ejections can easily be accounted in an RV analysis, and some oscillation modes can be averaged-out with long enough exposures, but the quasi-periodic variability imprinted by active regions on the stellar surface is still tricky to model. Moreover, we are now probing for the first time stellar effects with induced scatter under the m s^{-1}, such as supergranulation and meridional flows. These processes are still not fully understood and cannot yet be corrected for.

To address these signals, the exoplanet community has developed and adapted a large variety of modelling techniques aimed at the mitigation of stellar variability and the precise retrieval of planetary oscillations. In Chapter 2 I describe the most common methods to model stellar activity. All techniques rely on the fact that a pure Doppler-shift caused by the gravitational pull of an exoplanet orbiting the star affects all absorption lines of the observed spectrum in the same way and with the same velocity. On the opposite, stellar activity tends to distort the profiles of spectral lines, depending on their physical characteristics. However, the limited resolution and precision of observations is such that a precise RV cannot be derived simply by computing the shift of a single spectral line, and stellar processes affect different lines in unique ways. Basic methods for RV extraction generally combine all the information of all absorption lines in the spectrum into a single line profile, and compute RVs based on its peak. It is possible to account for some of the effects of stellar activity in the RV extraction, by either modelling the spectra or by using their cross-correlation functions (CCFs), as listed in Sections 2.1.2 and 2.1.3. Activity indicators can also help in identifying stellar signals. Whether they are derived from lines in the spectrum or from the CCF, they are designed to not be sensitive to Keplerian signals and to only map the variability caused by magnetic activity. Finally, there are also many techniques to study the RV time series themselves that either focus on the recognition of periodic signals, such as periodograms, or aim to directly model the stellar variability, as Gaussian processes.

Gaussian Process (GP) regression has in fact been proven to be a very successful technique for the mitigation of stellar activity, as it is able to model the variability without making any assumption about its functional form. In Chapter 3, I describe GPs in the context of probabilistic theory and Bayesian inference, and I introduce MAGPʏ-RV. MAGPʏ-RV is a Gaussian process regression pipeline with Markov Chain Monte Carlo parameter space searching algorithms I developed in the context of exoplanet detection and characterisation. It allows to simultaneously model stellar activity, described by a GP with the chosen covariance function, and Keplerian signals in the RVs as well as transits in photometric data.

I then use this pipeline for the analysis of two planetary systems, as reported in Chapter 4. In Section 4.1, I present the detection and characterisation of an inner mini-Neptune in a 9.2292005±0.0000063 day orbit and an outer mono-transiting sub-Saturn

planet in a $95.50^{+0.36}_{-0.25}$ day orbit around the moderately active, bright K5V star TOI-2134. Based on the analysis of five sectors of TESS data, I determine the radii of TOI-2134b and c to be 2.69±0.16 $R_\oplus$ for the inner planet and 7.27±0.42 $R_\oplus$ for the outer one. After careful periodogram analysis, the masses of both planets are derived based on RVs observed with the HARPS-N and the SOPHIE spectrographs via Gaussian Process regression: $9.13^{+0.78}_{-0.76}$ $M_\oplus$ for TOI-2134b and $41.89^{+7.69}_{-7.83}$ $M_\oplus$ for TOI-2134c. I analysed the photometric and radial-velocity data first separately, then jointly. The inner planet is a mini-Neptune with density consistent with either a water-world or a rocky core planet with a low-mass H/He envelope. The outer planet has a bulk density similar to Saturn's. The outer planet is derived to have a significant eccentricity of $0.67^{+0.05}_{-0.06}$ from a combination of photometry and RVs. I compute the irradiation of TOI-2134c as 1.45±0.10 times the bolometric flux received by Earth, positioning it for part of its orbit in the habitable zone of its system.
In Section 4.2, I include the work done in collaboration with Dr. Dalal and published in Dalal et al. (2024). As part of the HARPS-N Rocky Planet Search programme, we present the discovery of three super-Earth candidates orbiting HD 48948, a bright K-dwarf star with an apparent magnitude of $m_V = 8.58$ mag. Various methodologies were applied to extract the radial velocities from the spectra, and a comprehensive comparative analysis of possible system architectures obtained through these diverse extraction techniques was also conducted. In this work, I focus on the study of the activity-corrected RVs (after YARARA post-processing) and the one-dimensional GP analysis undertaken to identify the masses of the three planet candidates. This study reveals three planetary candidates with orbital periods of 7.3, 38, and 151 days, and minimum masses estimated at 4.96 ± 0.42 $M_\oplus$, 7.45 ± 0.75 $M_\oplus$, and 10.67 ± 0.90 $M_\oplus$, respectively. The outermost planet is also found to reside within the (temperate) habitable zone, positioned at a projected distance of 0.029″ from its star.

In parallel, I also analysed solar data in order to develop a better understanding of the processes driving stellar variability. The Sun is a fairly representative star of the sample of targets that are generally selected for RV surveys, and most crucially it is the only star we can spatially resolve. In Chapter 5 I list the most important solar disc-integrated RV dataset and I introduce the SolAster pipeline (Ervin et al. 2022a; Ervin et al. 2022b). SolAster is a tool that allows for the scrambling of resolved, full-disc solar observations taken by the

Helioseismic and Magnetic Imager on board of the Solar Dynamics Observatory, into Sun-as-a-star time series, comparable to the data measured by ground-based spectral surveys. Building on the work of Haywood et al. (2022), I use this pipeline to derive disc-integrated longitudinal magnetic field data in order to asses its uses in radial-velocity surveys. In Section 5.2, I show that the mean longitudinal magnetic field is an excellent rotational period detector and a useful tracer of the solar magnetic cycle. In order to put these results into context, I compare the mean longitudinal magnetic field to three common activity proxies derived from HARPS-N Sun-as-a-star data: the full-width at half-maximum, the bisector span and the S-index. I find that the mean longitudinal magnetic field does not correlate with the RVs and therefore cannot be used as a one-to-one proxy. However, with high cadence and a long baseline, the mean longitudinal magnetic field outperforms all other considered proxies as a solar rotational period detector, and can be used to inform our understanding of the physical processes happening on the surface of the Sun. I also test the mean longitudinal magnetic field as a "stellar activity tracer" on a reduced solar dataset to simulate stellar-like observational sampling, including realistic telescope time scheduling and seasonal visibility. With a Gaussian process regression analysis, I confirm that, as for the Sun, the mean longitudinal magnetic field is the most effective and efficient rotational period indicator over different levels of stellar activity. These results highlight the need for polarimetric time series observations of stars.

6.2 Next Steps

Through this thesis, I have highlighted the need of novel techniques to mitigate the stellar activity in order to reach the precision necessary to detect Earth-analogues. Gaussian processes represent an exciting tool to develop in order to break this stellar variability barrier. In my future work I intent to tackle this issue with two separate but complementary approaches.

6.2.1 Introducing Non-Stationarity to GPs for Exoplanet Detection

Recent works (e.g., Klein et al. 2024) have proven that new modelling techniques are required for the accurate and precise detection of small exoplanets in wide orbits. While

GP regression with a Quasi-periodic kernel has been very successful at mitigating stellar activity due to rotationally modulated effects, it does not perform as well on longer solar time series. Modelling solar data with baselines longer than roughly two years does not yield better results than modelling the same time series in separate chunks. As all stars are expected to undergo magnetic cycles, these results mean that missions such as the Terra Hunting Experiment will not be able to count on their long-term monitoring for increased planet detectability. The reason for this is simple: the covariance relationships between each set of datapoints change based on the stage of the magnetic cycle. We therefore cannot employ a single kernel with static hyperparameters to model the entire time series. I aim to develop a novel technique to model long-baseline time series and in particular their behaviours due to the magnetic cycle. As a first step, I plan to pinpoint the best way to model the long-term solar variability due to the cycle. The change in filling factors on the surface of the Sun imprints a typical skewed sinusoidal trend on the RVs. I plan to test whether this effect is best mitigated by fitting an additional squared exponential kernel (for example by simultaneous fit with activity indicators). The outputs of this preliminary analysis will not only guide my decisions for following developments, but will also represent the first systematic study on the mitigation of the magnetic cycle in the context exoplanet hunting. The real challenge is presented by the development of non-stationary physically motivated kernels. I plan to approach this problem with two separate but compatible approaches. On one hand, I plan on building on the previous work of Remes et al. (2017), and extend their proven formulation for non-stationarity to the Quasi-periodic kernel, following the programming formatting of open source Python packages such as GEORGE. On the other hand, I aim to also develop a technique to simultaneously model separate sections of the same time series with different stationary kernels, while requiring a continuous solution. This goal will most likely require the introduction of temporal-based damping and amplifying parameters to allow for one or another formulation of the same kernel to dominate. The results of this analysis will build towards a new modelling standard for Gaussian process regression in planet detection studies.

6.2.2 Modelling the Spectrum with GPs

I also plan to apply GP regression to spectral line characteristics and to eventually use GPs to model sections of spectrum. Building on the prior works of Miklos et al. (2020), Siegel et al. (2022), and Wise et al. (2022), I aim to first focus on absorption line-dependent information in order to disentangle how different lines react to stellar activity and how their distortions contribute to the RVs. Understanding how the physical properties of spectral lines influence the sensitivity of RV measurements to stellar variability is vital to isolate its effects and to develop new mitigation techniques. This method will allow me to robustly analyse the temporal changes in the shape of lines based on their physical properties (such as formation temperature or atomic source) or on their responses to physical processes (such as magnetic sensitivity measured via Landè g-factor). After computing line characteristics such as the average equivalent-widths and line-depths for each line group in each observation, I will model these time series using GP regression. This process will isolate which line-groups are most sensitive to specific sources of stellar activity at the various stages of the magnetic cycle, and are therefore the most useful to model them. As a next step, I aim to also look into the spectra directly. I will model distortions generated by stellar activity in each observation in the wavelength-domain using GPs. As described in Chapter 2 Section 2.1.1, Rajpaul et al. (2020) presented a proof-of-concept methodology in the realm of planet detection. They assume that the only variability in the spectra is generated by Doppler shifts and show that even a crude implementation yields comparable precision RVs with significantly lower dispersion than traditional RV extraction methods. Even in a basic form, implementing GPs to model the observed spectra can significantly facilitate their study. I thus plan to develop a robust technique to apply GP regression with MCMC likelihood optimisation to spectral data.

Bibliography

Aigrain, S., J. Llama, T. Ceillier, et al. 2015. 'Testing the recovery of stellar rotation signals from Kepler light curves using a blind hare-and-hounds exercise.' *MNRAS* 450, no. 3 (July): 3211–3226.

Aigrain, S., F. Pont, and S. Zucker. 2012. 'A simple method to estimate radial velocity variations due to stellar activity using photometry.' *MNRAS* 419, no. 4 (February): 3147–3158.

Akaike, H. 1983. 'Information measures and model selection.' *Internaltional Statistical Institute* 44:277–291.

Al Moulla, K., X. Dumusque, M. Cretignier, Y. Zhao, and J. A. Valenti. 2022. 'Measuring precise radial velocities on individual spectral lines. III. Dependence of stellar activity signal on line formation temperature.' *A&A* 664 (August): A34.

Al Moulla, K., X. Dumusque, P. Figueira, et al. 2023. 'Stellar signal components seen in HARPS and HARPS-N solar radial velocities.' *A&A* 669 (January): A39.

Albrecht, Simon, Joshua N. Winn, John A. Johnson, et al. 2012. 'Obliquities of Hot Jupiter Host Stars: Evidence for Tidal Interactions and Primordial Misalignments.' *ApJ* 757, no. 1 (September): 18.

Allen, C. W. 1973. *Astrophysical quantities.*

Amard, Louis, Julia Roquette, and Sean P. Matt. 2020. 'Evidence for metallicity-dependent spin evolution in the Kepler field.' *MNRAS* 499, no. 3 (December): 3481–3493.

Ambikasaran, Sivaram, Daniel Foreman-Mackey, Leslie Greengard, David W. Hogg, and Michael O'Neil. 2015. 'Fast Direct Methods for Gaussian Processes.' *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (June): 252.

Andretta, V., J. G. Doyle, and P. B. Byrne. 1997. 'The NaI $\lambda\lambda$5890,5896 resonance doublet as chromospheric diagnostics in M dwarfs.' *A&A* 322 (June): 266–279.

Anna John, A., A. Collier Cameron, J. P. Faria, et al. 2023. 'Sub-m s^{-1} upper limits from a deep HARPS-N radial-velocity search for planets orbiting HD 166620 and HD 144579.' *MNRAS* 525, no. 2 (October): 1687–1704.

Anna John, Ancy, Andrew Collier Cameron, and Thomas G. Wilson. 2022. 'The impact of two non-transiting planets and stellar activity on mass determinations for the super-Earth CoRoT-7b.' *MNRAS* 515, no. 3 (September): 3975–3995.

Artigau, Étienne, Charles Cadieux, Neil J. Cook, et al. 2022. 'Line-by-line Velocity Measurements: an Outlier-resistant Method for Precision Velocimetry.' *AJ* 164, no. 3 (September): 84.

Astudillo-Defru, N., R. F. Dıéaz, X. Bonfils, et al. 2017. 'The HARPS search for southern extra-solar planets. XLII. A system of Earth-mass planets around the nearby M dwarf YZ Ceti.' *A&A* 605 (September): L11.

Athay, R. G. 1976. *The solar chromosphere and corona: Quiet sun.* Vol. 53.

Babcock, Horace W. 1949. 'Magnetic Intensification of Stellar Absorption Lines.' *ApJ* 110 (September): 126.

Baglin, A., M. Auvergne, L. Boisnard, et al. 2006. 'CoRoT: a high precision photometer for stellar ecolution and exoplanet finding.' In *36th COSPAR Scientific Assembly*, 36:3749. January.

Bagnulo, Stefano, and John D. Landstreet. 2015. 'Stellar magnetic fields.' In *Polarimetry of Stars and Planetary Systems,* 224–243. Cambridge University Press.

Baranne, A., D. Queloz, M. Mayor, et al. 1996. 'ELODIE: A spectrograph for accurate radial velocity measurements.' *A&AS* 119 (October): 373–390.

Barkaoui, K., M. Gillon, Z. Benkhaldoun, et al. 2017. 'Assessment of the potential of the new Belgo-Moroccan telescope TRAPPIST-North for high-precision exoplanet transit photometry.' In *Journal of Physics Conference Series,* 869:012073. Journal of Physics Conference Series. June.

Barnes, J. R., J. S. Jenkins, H. R. A. Jones, et al. 2012. 'Red Optical Planet Survey: a new search for habitable earths in the southern sky.' *MNRAS* 424, no. 1 (July): 591–604.

Barragán, O., S. Aigrain, D. Kubyshkina, et al. 2019. 'Radial velocity confirmation of K2-100b: a young, highly irradiated, and low-density transiting hot Neptune.' *MNRAS* 490, no. 1 (November): 698–708.

Barragán, Oscar, Suzanne Aigrain, Vinesh M. Rajpaul, and Norbert Zicher. 2022. 'PYANETI - II. A multidimensional Gaussian process approach to analysing spectroscopic time-series.' *MNRAS* 509, no. 1 (January): 866–883.

Barros, S. C. C., J. M. Almenara, M. Deleuil, et al. 2014. 'Revisiting the transits of CoRoT-7b at a lower activity level.' *A&A* 569 (September): A74.

Barros, S. C. C., O. Demangeon, R. F. Dıéaz, et al. 2020. 'Improving transit characterisation with Gaussian process modelling of stellar variability.' *A&A* 634 (February): A75.

Bartlett, M. S. 1948. 'Smoothing Periodograms from Time-Series with Continuous Spectra.' *Nature* 161, no. 4096 (May): 686–687.

Batalha, Natalie M., Jason F. Rowe, Stephen T. Bryson, et al. 2013. 'Planetary Candidates Observed by Kepler. III. Analysis of the First 16 Months of Data.' *ApJS* 204, no. 2 (February): 24.

Batalha, Natasha E., Nikole K. Lewis, Michael R. Line, Jeff Valenti, and Kevin Stevenson. 2018. 'Strategies for Constraining the Atmospheres of Temperate Terrestrial Planets with JWST.' *ApJl* 856, no. 2 (April): L34.

Bedell, Megan, David W. Hogg, Daniel Foreman-Mackey, Benjamin T. Montet, and Rodrigo Luger. 2019. 'WOBBLE: A Data-driven Analysis Technique for Time-series Stellar Spectra.' *AJ* 158, no. 4 (October): 164.

Bellotti, S., P. Petit, J. Morin, et al. 2022. 'Mitigating stellar activity jitter with different line lists for least-squares deconvolution. Analysis of a parametric and a randomised line selection.' *A&A* 657 (January): A107.

Benz, W., C. Broeg, A. Fortier, et al. 2021. 'The CHEOPS mission.' *Experimental Astronomy* 51, no. 1 (February): 109–151.

Berger, T. E., and A. M. Title. 1996. 'On the Dynamics of Small-Scale Solar Magnetic Elements.' *ApJ* 463 (May): 365.

Bishop, Christopher M., and Nasser M. Nasrabadi. 2007. 'Pattern Recognition and Machine Learning.' *Journal of Electronic Imaging* 16, no. 4 (January): 049901.

Bloom, J. S., J. W. Richards, P. E. Nugent, et al. 2012. 'Automating Discovery and Classification of Transients and Variable Stars in the Synoptic Survey Era.' *PASP* 124, no. 921 (November): 1175.

Blunt, Sarah, Adolfo Carvalho, Trevor J. David, et al. 2023. 'Overfitting Affects the Reliability of Radial Velocity Mass Estimates of the V1298 Tau Planets.' *AJ* 166, no. 2 (August): 62.

Boisse, I., F. Bouchy, G. Hébrard, et al. 2011. 'Disentangling between stellar activity and planetary signals.' *A&A* 528 (April): A4.

Boisse, I., A. Eggenberger, N. C. Santos, et al. 2010. 'The SOPHIE search for northern extrasolar planets. III. A Jupiter-mass companion around HD 109246.' *A&A* 523 (November): A88.

Boisse, I., C. Moutou, A. Vidal-Madjar, et al. 2009. 'Stellar activity of planetary host star HD 189 733.' *A&A* 495, no. 3 (March): 959–966.

Bortle, Anna, Hallie Fausey, Jinbiao Ji, et al. 2021. 'A Gaussian Process Regression Reveals No Evidence for Planets Orbiting Kapteyn's Star.' *AJ* 161, no. 5 (May): 230.

Borucki, William J., David Koch, Gibor Basri, et al. 2010. 'Kepler Planet-Detection Mission: Introduction and First Results.' *Science* 327, no. 5968 (February): 977.

Borucki, William J., David G. Koch, Gibor Basri, et al. 2011. 'Characteristics of Planetary Candidates Observed by Kepler. II. Analysis of the First Four Months of Data.' *ApJ* 736, no. 1 (July): 19.

Bose, Souvik, and K. Nagaraju. 2018. 'On the Variability of the Solar Mean Magnetic Field: Contributions from Various Magnetic Features on the Surface of the Sun.' *ApJ* 862, no. 1 (July): 35.

Bouchy, F., R. F. Dıéaz, G. Hébrard, et al. 2013. 'SOPHIE+: First results of an octagonal-section fiber for high-precision radial velocity measurements.' *A&A* 549 (January): A49.

Bouchy, F., G. Hébrard, S. Udry, et al. 2009. 'The SOPHIE search for northern extrasolar planets . I. A companion around HD 16760 with mass close to the planet/brown-dwarf transition.' *A&A* 505, no. 2 (October): 853–858.

Brewer, Brendon J., Livia B. Pártay, and Gábor Csányi. 2009. 'Diffusive Nested Sampling.' *arXiv e-prints* (December): arXiv:0912.2380.

Brewer, John M., Debra A. Fischer, Ryan T. Blackman, et al. 2020. 'EXPRES. I. HD 3651 as an Ideal RV Benchmark.' *AJ* 160, no. 2 (August): 67.

Brown, S. F., J. -F. Donati, D. E. Rees, and M. Semel. 1991. 'Zeeman-Doppler imaging of solar-type and AP stars. IV. Maximum entropy reconstruction of 2D magnetic topologies.' *A&A* 250 (October): 463.

Brown, T. M., N. Baliber, F. B. Bianco, et al. 2013. 'Las Cumbres Observatory Global Telescope Network.' *PASP* 125, no. 931 (September): 1031.

Brun, Allan Sacha, and Matthew K. Browning. 2017. 'Magnetism, dynamo action and the solar-stellar connection.' *Living Reviews in Solar Physics* 14, no. 1 (September): 4.

Bryan, Marta L., Heather A. Knutson, Eve J. Lee, et al. 2019. 'An Excess of Jupiter Analogs in Super-Earth Systems.' *AJ* 157, no. 2 (February): 52.

Buchhave, Lars A., Martin Bizzarro, David W. Latham, et al. 2014. 'Three regimes of extrasolar planet radius inferred from host star metallicities.' *Nature* 509, no. 7502 (May 29, 2014): 593–595. Accessed June 1, 2014.

Buchhave, Lars A., David W. Latham, Anders Johansen, et al. 2012. 'An abundance of small exoplanets around stars with a wide range of metallicities.' *Nature* 486, no. 7403 (June): 375–377. Accessed July 17, 2018.

Bumba, V., and Robert Howard. 1965. 'A Study of the Development of Active Regions on the Sun.' *ApJ* 141 (May): 1492.

Burt, Jennifer A., Louise D. Nielsen, Samuel N. Quinn, et al. 2020. 'TOI-824 b: A New Planet on the Lower Edge of the Hot Neptune Desert.' *AJ* 160, no. 4 (October): 153.

Camacho, J. D., J. P. Faria, and P. T. P. Viana. 2023. 'Modelling stellar activity with Gaussian process regression networks.' *MNRAS* 519, no. 4 (March): 5439–5453.

Campbell, Bruce, G. A. H. Walker, and S. Yang. 1988. 'A Search for Substellar Companions to Solar-type Stars.' *ApJ* 331 (August): 902.

Cañas, Caleb I., Shubham Kanodia, Chad F. Bender, et al. 2022. 'TOI-3714 b and TOI-3629 b: Two Gas Giants Transiting M Dwarfs Confirmed with the Habitable-zone Planet Finder and NEID.' *AJ* 164, no. 2 (August): 50.

Cegla, H. M. 2019. 'The Impact of Stellar Surface Magnetoconvection and Oscillations on the Detection of Temperate, Earth-Mass Planets Around Sun-Like Stars.' *Geosciences* 9, no. 3 (March): 114.

Cegla, H. M., S. Shelyag, C. A. Watson, and M. Mathioudakis. 2013. 'Stellar Surface Magneto-convection as a Source of Astrophysical Noise. I. Multi-component Parameterization of Absorption Line Profiles.' *ApJ* 763, no. 2 (February): 95.

Cegla, H. M., C. A. Watson, T. R. Marsh, et al. 2012. 'Stellar jitter from variable gravitational redshift: implications for radial velocity confirmation of habitable exoplanets.' *MNRAS* 421, no. 1 (March): L54–L58.

Cegla, H. M., C. A. Watson, S. Shelyag, M. Mathioudakis, and S. Moutari. 2019. 'Stellar Surface Magnetoconvection as a Source of Astrophysical Noise. III. Sun-as-a-Star Simulations and Optimal Noise Diagnostics.' *ApJ* 879, no. 1 (July): 55.

Chabrier, G., I. Baraffe, F. Allard, and P. Hauschildt. 2000. 'Evolutionary Models for Very Low-Mass Stars and Brown Dwarfs with Dusty Atmospheres.' *ApJ* 542, no. 1 (October): 464–472.

Chaplin, W. J., H. M. Cegla, C. A. Watson, G. R. Davies, and W. H. Ball. 2019. 'Filtering Solar-Like Oscillations for Exoplanet Detection in Radial Velocity Observations.' *AJ* 157, no. 4 (April): 163.

Chaplin, William J., and Andrea Miglio. 2013. 'Asteroseismology of Solar-Type and Red-Giant Stars.' *ARA&A* 51, no. 1 (August): 353–392.

Charbonneau, Paul. 2020. 'Dynamo models of the solar cycle.' *Living Reviews in Solar Physics* 17, no. 1 (June): 4.

Chen, Jingjing, and David Kipping. 2017. 'Probabilistic Forecasting of the Masses and Radii of Other Worlds.' *ApJ* 834, no. 1 (January): 17.

Chen, Scott Shaobing, David L. Donoho, and Michael A. Saunders. 1998. 'Atomic Decomposition by Basis Pursuit.' *SIAM Journal on Scientific Computing* 20, no. 1 (January): 33–61.

Cheung, M. C. M., L. van Driel-Gesztelyi, V. Martıénez Pillet, and M. J. Thompson. 2017. 'The Life Cycle of Active Region Magnetic Fields.' *SSR* 210, nos. 1-4 (September): 317–349.

Chou, Dean-Yi, and De-Chang Dai. 2001. 'Solar Cycle Variations of Subsurface Meridional Flows in the Sun.' *ApJl* 559, no. 2 (October): L175–L178.

Choudhuri, Arnab Rai. 2021. 'The meridional circulation of the Sun: Observations, theory and connections with the solar dynamo.' *Science China Physics, Mechanics, and Astronomy* 64, no. 3 (March): 239601.

Collier Cameron, A., E. B. Ford, S. Shahaf, et al. 2021. 'Separating planetary reflex Doppler shifts from stellar variability in the wavelength domain.' *MNRAS* 505, no. 2 (August): 1699–1717.

Collier Cameron, A., A. Mortier, D. Phillips, et al. 2019. 'Three years of Sun-as-a-star radial-velocity observations on the approach to solar minimum.' *MNRAS* 487, no. 1 (July): 1082–1100.

Collier Cameron, Andrew. 2016. 'Extrasolar Planetary Transits.' In *Methods of Detecting Exoplanets: 1st Advanced School on Exoplanetary Science,* edited by Valerio Bozza, Luigi Mancini, and Alessandro Sozzetti, 428:89. Astrophysics and Space Science Library. January.

Cosentino, Rosario, Christophe Lovis, Francesco Pepe, et al. 2012. 'Harps-N: the new planet hunter at TNG.' In *Ground-based and Airborne Instrumentation for Astronomy IV,* edited by Ian S. McLean, Suzanne K. Ramsay, and Hideki Takami, 8446:84461V. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. September.

Cosentino, Rosario, Christophe Lovis, Francesco Pepe, et al. 2014. 'HARPS-N @ TNG, two year harvesting data: performances and results.' In *Ground-based and Airborne Instrumentation for Astronomy V,* edited by Suzanne K. Ramsay, Ian S. McLean, and Hideki Takami, 9147:91478C. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July.

Costes, Jean C., Christopher A. Watson, Ernst de Mooij, et al. 2021. 'Long-term stellar activity variations and their effect on radial-velocity measurements.' *MNRAS* 505, no. 1 (July): 830–850.

Couvidat, S., J. Schou, J. T. Hoeksema, et al. 2016. 'Observables Processing for the Helioseismic and Magnetic Imager Instrument on the Solar Dynamics Observatory.' *Solar Physics* 291, no. 7 (August): 1887–1938.

Crass, Jonathan, B. Scott Gaudi, Stephanie Leifer, et al. 2021. 'Extreme Precision Radial Velocity Working Group Final Report.' *arXiv e-prints* (July): arXiv:2107.14291.

Cretignier, M., X. Dumusque, S. Aigrain, and F. Pepe. 2023. 'YARARA V2: Reaching sub-m s^{-1} precision over a decade using PCA on line-by-line radial velocities.' *A&A* 678 (October): A2.

Cretignier, M., X. Dumusque, R. Allart, F. Pepe, and C. Lovis. 2020a. 'Measuring precise radial velocities on individual spectral lines. II. Dependance of stellar activity signal on line depth.' *A&A* 633 (January): A76.

Cretignier, M., X. Dumusque, N. C. Hara, and F. Pepe. 2021. 'YARARA: Significant improvement in RV precision through post-processing of spectral time series.' *A&A* 653 (September): A43.

Cretignier, M., X. Dumusque, and F. Pepe. 2022. 'Stellar activity correction using PCA decomposition of shells.' *A&A* 659 (March): A68.

Cretignier, M., J. Francfort, X. Dumusque, R. Allart, and F. Pepe. 2020b. 'RASSINE: Interactive tool for normalising stellar spectra. I. Description and performance of the code.' *A&A* 640 (August): A42.

Cretignier, M., A. G. M. Pietrow, and S. Aigrain. 2024. 'Stellar surface information from the Ca II H&K lines - I. Intensity profiles of the solar activity components.' *MNRAS* 527, no. 2 (January): 2940–2962.

Crossfield, Ian J. M., Erik Petigura, Joshua E. Schlieder, et al. 2015. 'A Nearby M Star with Three Transiting Super-Earths Discovered by K2.' *ApJ* 804, no. 1 (May): 10.

Cumming, Andrew, Geoffrey W. Marcy, and R. Paul Butler. 1999. 'The Lick Planet Search: Detectability and Mass Thresholds.' *ApJ* 526, no. 2 (December): 890–915.

Cutri, R. M., E. L. Wright, T. Conrow, et al. 2013. *Explanatory Supplement to the AllWISE Data Release Products.* Explanatory Supplement to the AllWISE Data Release Products, by R. M. Cutri et al. November.

Dalal, S., R. D. Haywood, A. Mortier, W. J. Chaplin, and N. Meunier. 2023. 'Predicting convective blueshift and radial-velocity dispersion due to granulation for FGK stars.' *MNRAS* 525, no. 3 (November): 3344–3353.

Dalal, S., F. Rescigno, M. Cretignier, et al. 2024. 'Trio of super-Earth candidates orbiting K-dwarf HD 48948: A new habitable zone candidate.' Submitted to MNRAS in January 2024.

Damasso, M., A. S. Bonomo, N. Astudillo-Defru, et al. 2018. 'Eyes on K2-3: A system of three likely sub-Neptunes characterized with HARPS-N and HARPS.' *A&A* 615 (July): A69.

Danielski, Camilla, Anna Brucalassi, Serena Benatti, et al. 2022. 'The homogeneous characterisation of Ariel host stars.' *Experimental Astronomy* 53, no. 2 (April): 473–510.

Davenport, James R. A., Kevin R. Covey, Riley W. Clarke, et al. 2019. 'The Evolution of Flare Activity with Stellar Age.' *ApJ* 871, no. 2 (February): 241.

De Beurs, Z. L., A. Vanderburg, C. J. Shallue, and Harps-N Collaboration. 2021. 'Identifying Exoplanets with Deep Learning. IV. Removing Stellar Activity Signals from Radial Velocity Measurements Using Neural Networks,' American Astronomical Society Meeting Abstracts, 53 (January): 332.04.

Del Moro, D. 2004. 'Solar granulation properties derived from three different time series.' *A&A* 428 (December): 1007–1015.

Delchambre, L. 2015. 'Weighted principal component analysis: a weighted covariance eigendecomposition approach.' *MNRAS* 446, no. 4 (February): 3545–3555.

Delisle, J. -B., N. Unger, N. C. Hara, and D. Ségransan. 2022. 'Efficient modeling of correlated noise. III. Scalable methods for jointly modeling several observables' time series with Gaussian processes.' *A&A* 659 (March): A182.

Demin, S. A., Y. A. Nefedyev, A. O. Andreev, N. Y. Demina, and S. F. Timashev. 2018. 'Non-stationarity and cross-correlation effects in the MHD solar activity.' *Advances in Space Research* 61, no. 2 (January): 639–644.

Desort, M., A. -M. Lagrange, F. Galland, S. Udry, and M. Mayor. 2007. 'Search for exoplanets with the radial-velocity technique: quantitative diagnostics of stellar activity.' *A&A* 473, no. 3 (October): 983–993.

Di Mauro, M. P. 2016. 'A review on Asteroseismology.' In *Frontier Research in Astrophysics II (FRAPWS2016),* 29. May.

DiTomasso, Victoria, Chantanelle Nava, Mercedes López-Morales, et al. 2023. 'Independent Validation of the Temperate Super-Earth HD 79211 b using HARPS-N.' *AJ* 165, no. 2 (February): 38.

Dodson-Robinson, Sarah E. 2022. *NWelch: Spectral analysis of time series with nonuniform observing cadence.* Astrophysics Source Code Library, record ascl:2202.002, February.

Dodson-Robinson, Sarah E., Victor Ramirez Delgado, Justin Harrell, and Charlotte L. Haley. 2022. 'Magnitude-squared Coherence: A Powerful Tool for Disentangling Doppler Planet Discoveries from Stellar Activity.' *AJ* 163, no. 4 (April): 169.

Domıénguez Sánchez, H., M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer. 2018. 'Improving galaxy morphologies for SDSS with Deep Learning.' *MNRAS* 476, no. 3 (February): 3661–3676.

Donati, J. -F., and A. Collier Cameron. 1997. 'Differential rotation and magnetic polarity patterns on AB Doradus.' *MNRAS* 291, no. 1 (October): 1–19.

Donati, J. -F., D. Kouach, C. Moutou, et al. 2020. 'SPIRou: NIR velocimetry and spectropolarimetry at the CFHT.' *MNARS* 498, no. 4 (November): 5684–5703.

Donati, J. -F., and J. D. Landstreet. 2009. 'Magnetic Fields of Nondegenerate Stars.' *ARA&A* 47, no. 1 (September): 333–370.

Donati, J. -F., L. T. Lehmann, P. I. Cristofari, et al. 2023. 'Magnetic fields and rotation periods of M dwarfs from SPIRou spectra.' *MNRAS* 525, no. 2 (October): 2015–2039.

Donati, J. -F., M. Semel, B. D. Carter, D. E. Rees, and A. Collier Cameron. 1997. 'Spectropolarimetric observations of active stars.' *MNRAS* 291, no. 4 (November): 658–682.

Dong, Subo, Boaz Katz, and Aristotle Socrates. 2013. 'Directly Imaging Tidally Powered Migrating Jupiters.' *ApJl* 762, no. 2 (January): L26.

Dravins, D. 1999. 'Stellar Surface Convection, Line Asymmetries, and Wavelength Shifts.' In *IAU Colloq. 170: Precise Stellar Radial Velocities,* edited by J. B. Hearnshaw and Colin David Scarfe, 185:268. Astronomical Society of the Pacific Conference Series. January.

Dressing, Courtney D., and David Charbonneau. 2013. 'The Occurrence Rate of Small Planets around Small Stars.' *ApJ* 767, no. 1 (April): 95.

Dumusque, X. 2018. 'Measuring precise radial velocities on individual spectral lines. I. Validation of the method and application to mitigate stellar activity.' *A&A* 620 (November): A47.

Dumusque, X., M. Cretignier, D. Sosnowska, et al. 2021. 'Three years of HARPS-N high-resolution spectroscopy and precise radial velocity data for the Sun.' *A&A* 648 (April): A103.

Dumusque, X., S. Udry, C. Lovis, N. C. Santos, and M. J. P. F. G. Monteiro. 2011. 'Planetary detection limits taking into account stellar noise. I. Observational strategies to reduce stellar oscillation and granulation effects.' *A&A* 525 (January): A140.

Dumusque, Xavier, Alex Glenday, David F. Phillips, et al. 2015. 'HARPS-N Observes the Sun as a Star.' *ApJl* 814, no. 2 (December): L21.

Eastman, Jason, B. Scott Gaudi, and Eric Agol. 2013. 'EXOFAST: A Fast Exoplanetary Fitting Suite in IDL.' *PASP* 125, no. 923 (January): 83.

Edelson, R. A., and J. H. Krolik. 1988. 'The Discrete Correlation Function: A New Method for Analyzing Unevenly Sampled Variability Data.' *ApJ* 333 (October): 646.

Ervin, Tamar, Samuel Halverson, Abigail Burrows, et al. 2022a. 'Leveraging Space-based Data from the Nearest Solar-type Star to Better Understand Stellar Activity Signatures in Radial Velocity Data.' *AJ* 163, no. 6 (June): 272.

Ervin, Tamar, Samuel Halverson, Abigail Burrows, et al. 2022b. *SolAster: 'Sun-as-a-star' radial velocity variations.* Astrophysics Source Code Library, record ascl:2207.009, July.

Evershed, J. 1909. 'Radial movement in sun-spots.' *MNRAS* 69 (March): 454.

Fabrycky, Daniel, and Scott Tremaine. 2007. 'Shrinking Binary and Planetary Orbits by Kozai Cycles with Tidal Friction.' *ApJ* 669, no. 2 (November): 1298–1315.

Faria, J. P., R. D. Haywood, B. J. Brewer, et al. 2016. 'Uncovering the planets and stellar activity of CoRoT-7 using only radial velocities.' *A&A* 588 (April): A31.

Faria, J. P., N. C. Santos, P. Figueira, and B. J. Brewer. 2018. 'kima: Exoplanet detection in radial velocities.' *The Journal of Open Source Software* 3, no. 26 (June): 487.

Figueira, P., N. C. Santos, F. Pepe, C. Lovis, and N. Nardetto. 2013. 'Line-profile variations in radial-velocity measurements. Two alternative indicators for planetary searches.' *A&A* 557 (September): A93.

Fischer, Debra A., Guillem Anglada-Escude, Pamela Arriagada, et al. 2016. 'State of the Field: Extreme Precision Radial Velocities.' *PASP* 128, no. 964 (June): 066001.

Fletcher, L., B. R. Dennis, H. S. Hudson, et al. 2011. 'An Observational Overview of Solar Flares.' *SSR* 159, nos. 1-4 (September): 19–106.

Ford, Eric B. 2006. 'Improving the Efficiency of Markov Chain Monte Carlo for Analyzing the Orbits of Extrasolar Planets.' *ApJ* 642, no. 1 (May): 505–522.

Ford, Eric B., and Frederic A. Rasio. 2008. 'Origins of Eccentric Extrasolar Planets: Testing the Planet-Planet Scattering Model.' *ApJ* 686, no. 1 (October): 621–636.

Foreman-Mackey, Daniel, Eric Agol, Sivaram Ambikasaran, and Ruth Angus. 2017. 'Fast and Scalable Gaussian Process Modeling with Applications to Astronomical Time Series.' *ApJ* 154, no. 6 (December): 220.

Foreman-Mackey, Daniel, Alex Conley, Will Meierjurgen Farr, et al. 2013. *emcee: The MCMC Hammer,* March.

Foreman-Mackey, Daniel, Weixiang Yu, Sachin Yadav, et al. 2024. *dfm/tinygp: The tiniest of Gaussian Process libraries.* V. v0.3.0, January.

Fortney, Jonathan J., Channon Visscher, Mark S. Marley, et al. 2020. 'Beyond Equilibrium Temperature: How the Atmosphere/Interior Connection Affects the Onset of Methane, Ammonia, and Clouds in Warm Transiting Giant Planets.' *AJ* 160, no. 6 (December): 288.

Foukal, Peter. 1998. 'What Determines the Relative Areas of Spots and Faculae on Sun-like Stars?' *ApJ* 500, no. 2 (June): 958–965.

Fouqué, P., E. Martioli, J. -F. Donati, et al. 2023. 'The SPIRou legacy survey. Rotation period of quiet M dwarfs from circular polarization in near-infrared spectral lines: The SPIRou APERO analysis.' *A&A* 672 (April): A52.

Fox, Chris, and Paul Wiegert. 2022. 'NEOSSat Observations of TESS Objects of Interest.' In *Bulletin of the American Astronomical Society,* 54:102.25. June.

Fressin, F., T. Guillot, and L. Nesta. 2009. 'Interpreting the yield of transit surveys: are there groups in the known transiting planets population?' *A&A* 504, no. 2 (September): 605–615.

Fröhlich, Claus, Bo N. Andersen, Thierry Appourchaux, et al. 1997. 'First Results from VIRGO, the Experiment for Helioseismology and Solar Irradiance Monitoring on SOHO.' *SolPhys* 170, no. 1 (January): 1–25.

Fröhlich, Claus, José Romero, Hansjörg Roth, et al. 1995. 'VIRGO: Experiment for Helioseismology and Solar Irradiance Monitoring.' *SolPhys* 162, nos. 1-2 (December): 101–128.

Fulton, Benjamin J., Erik A. Petigura, Andrew W. Howard, et al. 2017. 'The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets.' *AJ* 154, no. 3 (September): 109.

Gaia Collaboration. 2020. 'VizieR Online Data Catalog: Gaia EDR3 (Gaia Collaboration, 2020).' *VizieR Online Data Catalog* (November): I/350.

Gaia Collaboration, A. Vallenari, A. G. A. Brown, et al. 2023. 'Gaia Data Release 3. Summary of the content and survey properties.' *A&A* 674 (June): A1.

García, Rafael A., and Jérôme Ballot. 2019. 'Asteroseismology of solar-type stars.' *Living Reviews in Solar Physics* 16, no. 1 (September): 4.

Gaudi, B. Scott, Sara Seager, Bertrand Mennesson, et al. 2020. 'The Habitable Exoplanet Observatory (HabEx) Mission Concept Study Final Report.' *arXiv e-prints* (January): arXiv:2001.06683.

Gelman, Andrew, and Donald B Rubin. 1992. 'Inference from iterative simulation using multiple sequences.' *Statistical science,* 457–472.

Giclas, H. L., Jr. Burnham R., and N. G. Thomas. 1979. 'The Lowell Propoer motion Survey.' *Lowell Observatory Bulletin* 8 (January): 145–154.

Gilbertson, Christian, and Eric Ford. 2022. 'Jointly Modeling Telluric Features and Stellar Variability with StellarSpectraObservationFitting.jl.' In *Bulletin of the American Astronomical Society,* 54:102.196. June.

Giles, Helen A. C., Andrew Collier Cameron, and Raphaëlle D. Haywood. 2017. 'A Kepler study of starspot lifetimes with respect to light-curve amplitude and spectral type.' *MNRAS* 472, no. 2 (December): 1618–1627.

Gilliland, Ronald L., and Sallie L. Baliunas. 1987. 'Objective Characterization of Stellar Activity Cycles. I. Methods and Solar Cycle Analyses.' *ApJ* 314 (March): 766.

Gomes da Silva, J., N. C. Santos, X. Bonfils, et al. 2011. 'Long-term magnetic activity of a sample of M-dwarf stars from the HARPS program. I. Comparison of activity indices.' *A&A* 534 (October): A30.

Gray, David F. 1989. 'The Morphology of Reversed Spectral Line Bisectors.' *PASP* 101 (September): 832.

———. 2005. 'Shapes of Spectral Line Bisectors for Cool Stars.' PASP 117, no. 833 (July): 711–720.

Gregory, Phil. 2010. *Bayesian Logical Data Analysis for the Physical Sciences.* Chap. 12.

Grieves, N., J. Ge, N. Thomas, et al. 2018. 'Chemo-kinematics of the Milky Way from the SDSS-III MARVELS survey.' *MNRAS* 481, no. 3 (December): 3244–3265.

Grigor'ev, V. M., and M. L. Demidov. 1987. 'Observations of the solar mean magnetic field at the Sayan Observatory during 1982-1984.' *SolPhys* 114, no. 1 (January): 147–163.

Grunblatt, Samuel K., Andrew W. Howard, and Raphaëlle D. Haywood. 2015. 'Determining the Mass of Kepler-78b with Nonparametric Gaussian Process Estimation.' *ApJ* 808, no. 2 (August): 127.

Guerrero, Natalia M., S. Seager, Chelsea X. Huang, et al. 2021. 'The TESS Objects of Interest Catalog from the TESS Prime Mission.' *ApJs* 254, no. 2 (June): 39.

Hadden, Sam, and Yoram Lithwick. 2014. 'Densities and Eccentricities of 139 Kepler Planets from Transit Time Variations.' *ApJ* 787, no. 1 (May): 80.

Hale, S. J., W. J. Chaplin, G. R. Davies, and Y. P. Elsworth. 2020. 'A next generation upgraded observing platform for the automated Birmingham Solar Oscillations Network (BiSON).' In *Software and Cyberinfrastructure for Astronomy VI,* edited by Juan C. Guzman and Jorge Ibsen, 11452:1145222. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. December.

Hall, Jeffrey C. 2008. 'Stellar Chromospheric Activity.' *Living Reviews in Solar Physics* 5, no. 1 (December): 2.

Hamilton, Douglas P., and Joseph A. Burns. 1992. 'Orbital stability zones about asteroids II. The destabilizing effects of eccentric orbits and of solar radiation.' *Icarus* 96, no. 1 (March): 43–64.

Hara, Nathan C., G. Boué, J. Laskar, and A. C. M. Correia. 2017. 'Radial velocity data analysis with compressed sensing techniques.' *MNRAS* 464, no. 1 (January): 1220–1246.

Hara, Nathan C., and Eric B. Ford. 2023. 'Statistical Methods for Exoplanet Detection with Radial Velocities.' *Annual Review of Statistics and Its Application* 10, no. 1 (March): 623–649.

Hara, Nathan C., and Alessandro R. Mari. 2021. *l1p: Python implementation of the l1 periodogram.* Astrophysics Source Code Library, record ascl:2112.024, December.

Hara, Nathan C., Nicolas Unger, Jean-Baptiste Delisle, Rodrigo F. Dıéaz, and Damien Ségransan. 2022. 'Detecting exoplanets with the false inclusion probability. Comparison with other detection criteria in the context of radial velocities.' *A&A* 663 (July): A14.

Harada, Caleb K., Courtney D. Dressing, Stephen R. Kane, and Bahareh Adami Ardestani. 2024. 'Setting the stage for the search for life with the Habitable Worlds Observatory: Properties of 164 promising planet survey targets.' *arXiv e-prints* (January): arXiv:2401.03047.

Hart, A. B. 1954. 'Motions in the Sun at the photospheric level. IV. The equatorial rotation and possible velocity fields in the photosphere.' *MNRAS* 114 (January): 17.

———. 1956. 'Motions in the Sun at the photospheric level. VI. Large-scale motions in the equatorial region.' *MNRAS* 116 (January): 38.

Hathaway, David H. 2010. ‘The Solar Cycle.’ *Living Reviews in Solar Physics* 7, no. 1 (December): 1.

———. 2015. ‘The Solar Cycle.’ *Living Reviews in Solar Physics* 12, no. 1 (September): 4.

Hathaway, David H., and Lisa Rightmire. 2010. ‘Variations in the Sun’s Meridional Flow over a Solar Cycle.’ *Science* 327, no. 5971 (March): 1350.

Hatzes, A. P., R. Dvorak, G. Wuchterl, et al. 2010. ‘An investigation into the radial velocity variations of CoRoT-7.’ *A&A* 520 (September): A93.

Haywood, R. D., A. Collier Cameron, D. Queloz, et al. 2014. ‘Planets and stellar activity: hide and seek in the CoRoT-7 system.’ *MNRAS* 443, no. 3 (September): 2517–2531.

Haywood, R. D., A. Collier Cameron, Y. C. Unruh, et al. 2016. ‘The Sun as a planet-host star: proxies from SDO images for HARPS radial-velocity variations.’ *MNRAS* 457, no. 4 (April): 3637–3651.

Haywood, R. D., T. W. Milbourne, S. H. Saar, et al. 2022. ‘Unsigned Magnetic Flux as a Proxy for Radial-velocity Variations in Sun-like Stars.’ *ApJ* 935, no. 1 (August): 6.

Hébrard, É. M., J. -F. Donati, X. Delfosse, et al. 2016. ‘Modelling the RV jitter of early-M dwarfs using tomographic imaging.’ *MNRAS* 461, no. 2 (September): 1465–1497.

Hébrard, G., F. Bouchy, F. Pont, et al. 2008. ‘Misaligned spin-orbit in the XO-3 planetary system?’ *A&A* 488, no. 2 (September): 763–770.

Hekker, S., and J. Christensen-Dalsgaard. 2017. ‘Giant star seismology.’ *A&A* 25, no. 1 (June): 1.

Henry, Gregory W., Geoffrey W. Marcy, R. Paul Butler, and Steven S. Vogt. 2000. ‘A Transiting “51 Peg-like” Planet.’ *ApJl* 529, no. 1 (January): L41–L44.

Herbst, William, and Jeffrey R. Miller. 1989. ‘Hα Photometry of Dwarf K and M Stars: Chromospheric Activity.’ *AJ* 97 (March): 891.

Hildebrand, A. R., K. A. Carroll, E. F. Tedesco, et al. 2004. ‘Advantages of Searching for Asteroids from Low Earth Orbit: the NEOSSat Mission.’ *Earth Moon and Planets* 95, nos. 1-4 (December): 33–40.

Hoeksema, J. Todd, Yang Liu, Keiji Hayashi, et al. 2014. 'The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: Overview and Performance.' *SolPhys* 289, no. 9 (September): 3483–3530.

Houllé, M., A. Vigan, A. Carlotti, et al. 2021. 'Direct imaging and spectroscopy of exoplanets with the ELT/HARMONI high-contrast module.' *A&A* 652 (August): A67.

Howard, Russell A., Angelos Vourlidas, and Guillermo Stenborg. 2023. 'The evolution of our understanding of coronal mass ejections.' *Frontiers in Astronomy and Space Sciences* 10 (November): 1264226.

Howell, Steve B., Charlie Sobeck, Michael Haas, et al. 2014. 'The K2 Mission: Characterization and Early Results.' *PASP* 126, no. 938 (April): 398.

Huang, Chelsea, Yanqin Wu, and Amaury H. M. J. Triaud. 2016. 'Warm Jupiters Are Less Lonely than Hot Jupiters: Close Neighbors.' *ApJ* 825, no. 2 (July): 98.

Huang, Chelsea X., Andrew Vanderburg, Andras Pál, et al. 2020a. 'Photometry of 10 Million Stars from the First Two Years of TESS Full Frame Images: Part I.' *Research Notes of the American Astronomical Society* 4, no. 11 (November): 204.

———. 2020b. 'Photometry of 10 Million Stars from the First Two Years of TESS Full Frame Images: Part II.' *Research Notes of the American Astronomical Society* 4, no. 11 (November): 206.

Huang, G. -H., C. -H. Lin, and L. C. Lee. 2017. 'Solar Open Flux Migration from Pole to Pole: Magnetic Field Reversal.' *Scientific Reports* 7 (August): 9488.

Hughes, P. A., H. D. Aller, and M. F. Aller. 1992. 'The University of Michigan Radio Astronomy Data Base. I. Structure Function Analysis and the Relation between BL Lacertae Objects and Quasi-stellar Objects.' *ApJ* 396 (September): 469.

Irving, Zackery A., Steven H. Saar, Bradford J. Wargelin, and José-Dias do Nascimento. 2023. 'Stellar Cycles in Fully Convective Stars and a New Interpretation of Dynamo Evolution.' *ApJ* 949, no. 2 (June): 51.

Irwin, Alan W., Bruce Campbell, Christopher L. Morbey, G. A. H. Walker, and S. Yang. 1989. 'Long-Period Radial-Velocity Variations of Arcturus.' *PASP* 101 (February): 147.

Isaacson, Howard, and Debra Fischer. 2010. 'Chromospheric Activity and Jitter Measurements for 2630 Stars on the California Planet Search.' *ApJ* 725, no. 1 (December): 875–885.

Jacob, W. S. 1855. 'On certain Anomalies presented by the Binary Star 70 Ophiuchi.' *MNRAS* 15 (June): 228.

Jeffers, S. V., and C. U. Keller. 2009. 'An analytical model to demonstrate the reliability of reconstructed 'active longitudes'.' In *15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun,* edited by Eric Stempels, 1094:664–667. American Institute of Physics Conference Series. February.

Jeffers, Sandra V., René Kiefer, and Travis S. Metcalfe. 2023. 'Stellar Activity Cycles.' *Sapce Sci. Rev.* 219, no. 7 (October): 54.

Jenkins, Jon M. 2002. 'The Impact of Solar-like Variability on the Detectability of Transiting Terrestrial Planets.' *ApJ* 575, no. 1 (August): 493–505.

Jenkins, Jon M., Douglas A. Caldwell, Hema Chandrasekaran, et al. 2010. 'Overview of the Kepler Science Processing Pipeline.' *ApJl* 713, no. 2 (April): L87–L91.

Jenkins, Jon M., Joseph D. Twicken, Sean McCauliff, et al. 2016. 'The TESS science processing operations center.' In *Software and Cyberinfrastructure for Astronomy IV,* edited by Gianluca Chiozzi and Juan C. Guzman, vol. 9913, 99133E. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. August.

Jiang, Jonathan H., Xuan Ji, Nicolas Cowan, Renyu Hu, and Zonghong Zhu. 2019. 'Empirical Predictions for the Period Distribution of Planets to Be Discovered by the Transiting Exoplanet Survey Satellite.' *AJ* 158, no. 2 (August): 96.

Johnson, John Asher, Kimberly M. Aller, Andrew W. Howard, and Justin R. Crepp. 2010. 'Giant Planet Occurrence in the Stellar Mass-Metallicity Plane.' *PASP* 122, no. 894 (August): 905.

Jones, David E., David C. Stenning, Eric B. Ford, et al. 2017. 'Improving Exoplanet Detection Power: Multivariate Gaussian Process Models for Stellar Activity.' *arXiv e-prints* (November): arXiv:1711.01318.

Jontof-Hutter, Daniel, Eric B. Ford, Jason F. Rowe, et al. 2016. 'Secure Mass Measurements from Transit Timing: 10 Kepler Exoplanets between 3 and 8 $M_\oplus$ with Diverse Densities and Incident Fluxes.' *ApJ* 820, no. 1 (March): 39.

Jurgenson, C., D. Fischer, T. McCracken, et al. 2016. 'EXPRES: a next generation RV spectrograph in the search for earth-like worlds.' In *Ground-based and Airborne Instrumentation for Astronomy VI,* edited by Christopher J. Evans, Luc Simard, and Hideki Takami, vol. 9908, 99086T. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. August.

Kalos, Malvin H. 1986. *Monte Carlo methods. Vol.1: Basics.*

Katsova, M. M., V. N. Obridko, D. D. Sokoloff, and I. M. Livshits. 2022. 'Solar and Stellar Flares: Frequency, Active Regions, and Stellar Dynamo.' *ApJ* 936, no. 1 (September): 49.

Kempton, Eliza M. -R., Jacob L. Bean, Dana R. Louie, et al. 2018. 'A Framework for Prioritizing the TESS Planetary Candidates Most Amenable to Atmospheric Characterization.' *PASP* 130, no. 993 (November): 114401.

Kipping, David M. 2013. 'Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws.' *MNRAS* 435, no. 3 (November): 2152–2160.

———. 2014. 'Bayesian priors for the eccentricity of transiting planets.' *MNRAS* 444, no. 3 (November): 2263–2269.

Klein, B., S. Aigrain, M. Cretignier, et al. 2024. 'Investigating activity-induced shape distortions in the absorption lines of the Sun.' Submitted to MNRAS in March 2024.

Klein, Baptiste, Jean-François Donati, Claire Moutou, et al. 2021. 'Investigating the young AU Mic system with SPIRou: large-scale stellar magnetic field and close-in planet mass.' *MNRAS* 502, no. 1 (March): 188–205.

Kochukhov, Oleg. 2016. 'Doppler and Zeeman Doppler Imaging of Stars.' In *Lecture Notes in Physics, Berlin Springer Verlag,* edited by Jean-Pierre Rozelot and Coralie Neiner, 914:177.

Komm, R., I. González Hernández, R. Howe, and F. Hill. 2015. 'Solar-Cycle Variation of Subsurface Meridional Flow Derived with Ring-Diagram Analysis.' *SolPhys* 290, no. 11 (November): 3113–3136.

Kopparapu, Ravi Kumar, Ramses M. Ramirez, James SchottelKotte, et al. 2014. 'Habitable Zones around Main-sequence Stars: Dependence on Planetary Mass.' *ApJl* 787, no. 2 (June): L29.

Kotov, V. A., and L. S. Levitskii. 1983. 'The discreteness of the rotation periods of the solar and interplanetary magnetic fields.' *Izvestiya Ordena Trudovogo Krasnogo Znameni Krymskoj Astrofizicheskoj Observatorii* 68 (January): 56–68.

Kreidberg, Laura. 2015. 'batman: BAsic Transit Model cAlculatioN in Python.' *PASP* 127, no. 957 (November): 1161.

Kutsenko, A. S., V. I. Abramenko, and V. B. Yurchyshyn. 2017. 'Contribution to the Solar Mean Magnetic Field from Different Solar Regions.' *SolPhys* 292, no. 9 (September): 121.

Lakeland, Ben S., Tim Naylor, Raphaëlle D. Haywood, et al. 2024. 'The magnetically quiet solar surface dominates HARPS-N solar RVs during low activity.' *MNRAS* 527, no. 3 (January): 7681–7691.

Landstreet, J. D. 1982. 'A search for magnetic fields in normal upper-main-sequence stars.' *ApJ* 258 (July): 639–650.

———. 1992. 'Magnetic fields at the surfaces of stars.' *A&A* 4, no. 1 (June): 35–77.

Lanza, A. F. 2010. 'Stellar magnetic cycles.' In *Solar and Stellar Variability: Impact on Earth and Planets,* edited by Alexander G. Kosovichev, Alexandre H. Andrei, and Jean-Pierre Rozelot, 264:120–129. February.

Lanza, A. F., A. Collier Cameron, and R. D. Haywood. 2019. 'Reducing activity-induced variations in a radial-velocity time series of the Sun as a star.' MNRAS 486, no. 3 (July): 3459–3464.

Lanza, A. F., L. Malavolta, S. Benatti, et al. 2018. 'The GAPS Programme with HARPS-N at TNG. XVII. Line profile indicators and kernel regression as diagnostics of radial-velocity variations due to stellar activity in solar-like stars.' *A&A* 616 (September): A155.

Laskar, J. 1988. 'Secular evolution of the solar system over 10 million years.' *A&A* 198 (June): 341–362.

———. 1990. 'The chaotic motion of the solar system: A numerical estimate of the size of the chaotic zones.' *Icarus* 88, no. 2 (December): 266–291.

Laskar, Jacques. 1993. 'Frequency analysis for multi-dimensional systems. Global dynamics and diffusion.' *Physica D Nonlinear Phenomena* 67, nos. 1-3 (August): 257–281.

Leighton, Robert B. 1963. 'The Solar Granulation.' *Annual Review of Astronomy and Astrophysics* 1 (January): 19.

Leighton, Robert B., Robert W. Noyes, and George W. Simon. 1962. 'Velocity Fields in the Solar Atmosphere. I. Preliminary Report.' *ApJ* 135 (March): 474.

Liang, Yan, Peter Melchior, Sicong Lu, Andy Goulding, and Charlotte Ward. 2023. 'Autoencoding Galaxy Spectra. II. Redshift Invariance and Outlier Detection.' *AJ* 166, no. 2 (August): 75.

Liang, Yan, Joshua N. Winn, and Peter Melchior. 2024. 'AESTRA: Deep Learning for Precise Radial Velocity Estimation in the Presence of Stellar Activity.' *AJ* 167, no. 1 (January): 23.

Lienhard, F., A. Mortier, L. Buchhave, et al. 2022. 'Multi-mask least-squares deconvolution: extracting RVs using tailored masks.' *MNRAS* 513, no. 4 (July): 5328–5343.

Lienhard, F., A. Mortier, H. M. Cegla, et al. 2023. 'Unsigned magnetic flux proxy from solar optical intensity spectra.' *MNRAS* 522, no. 4 (July): 5862–5878.

Lin, Andrea S. J., Andrew Monson, Suvrath Mahadevan, et al. 2022. 'Observing the Sun as a Star: Design and Early Results from the NEID Solar Feed.' *AJ* 163, no. 4 (April): 184.

Lindegren, Lennart, and Dainis Dravins. 2003. 'The fundamental definition of "radial velocity".' *A&A* 401 (April): 1185–1201.

Llama, Joe, Debra Fischer, John Brewer, Lily Zhao, and Andrew Szymkowiak. 2022. 'Observing the Sun with EXPRES and the Lowell Observatory Solar Telescope.' In *Bulletin of the American Astronomical Society,* 54:102.102. June.

Löhner-Böttcher, Johannes, Wolfgang Schmidt, H. Doerr, et al. 2017. 'LARS - An Absolute Reference Spectrograph for solar observations, Upgrade from a prototype to a turnkey system.' *A&A* 607 (July).

López-Morales, Mercedes, Raphaëlle D. Haywood, Jeffrey L. Coughlin, et al. 2016. 'Kepler-21b: A Rocky Planet Around a V = 8.25 Magnitude Star.' *AJ* 152, no. 6 (December): 204.

Louie, Dana R., Drake Deming, Loic Albert, et al. 2018. 'Simulated JWST/NIRISS Transit Spectroscopy of Anticipated Tess Planets Compared to Select Discoveries from Space-based and Ground-based Surveys.' *PASP* 130, no. 986 (April): 044401.

Lovis, C., M. Mayor, F. Bouchy, et al. 2009. 'Towards the characterization of the hot Neptune/super-Earth population around nearby bright stars.' In *Transiting Planets,* edited by Frédéric Pont, Dimitar Sasselov, and Matthew J. Holman, 253:502–505. February.

Lovis, Christophe, X. Dumusque, N. C. Santos, S. Udry, and M. Mayor. 2011. 'HARPS Long-Term Results: Effects of Stellar Magnetic Cycles, and Planet Detection Limits.' In *AAS/Division for Extreme Solar Systems Abstracts,* 2:2.02. AAS/Division for Extreme Solar Systems Abstracts. September.

Lubin, Dan, David Tytler, and David Kirkman. 2012. 'Frequency of Maunder Minimum Events in Solar-type Stars Inferred from Activity and Metallicity Observations.' *ApJl* 747, no. 2 (March): L32.

Luhn, J., J. Wright, A. Howard, and H. Isaacson. 2020. 'Astrophysical Insights into Radial Velocity Jitter from an Analysis of 650 Planet-Search Stars.' In *American Astronomical Society Meeting Abstracts #235,* 235:374.06. American Astronomical Society Meeting Abstracts. January.

Luque, R., L. M. Serrano, K. Molaverdikhani, et al. 2021. 'A planetary system with two transiting mini-Neptunes near the radius valley transition around the bright M dwarf TOI-776.' *A&A* 645 (January): A41.

Malavolta, L., C. Lovis, F. Pepe, C. Sneden, and S. Udry. 2017. 'Atmospheric stellar parameters from cross-correlation functions.' *MNRAS* 469, no. 4 (August): 3965–3975.

Malavolta, L., V. Nascimbeni, G. Piotto, et al. 2016. 'The GAPS programme with HARPS-N at TNG. XI. Pr 0211 in M 44: the first multi-planet system in an open cluster.' *A&A* 588 (April): A118.

Malavolta, Luca, Andrew W. Mayo, Tom Louden, et al. 2018. 'An Ultra-short Period Rocky Super-Earth with a Secondary Eclipse and a Neptune-like Companion around K2-141.' *AJ* 155, no. 3 (March): 107.

Maldonado, J., D. F. Phillips, X. Dumusque, et al. 2019. 'Temporal evolution and correlations of optical activity indicators measured in Sun-as-a-star observations.' *A&A* 627 (July): A118.

Mamajek, Eric E., and Lynne A. Hillenbrand. 2008. 'Improved Age Estimation for Solar-Type Dwarfs Using Activity-Rotation Diagnostics.' *APJ* 687, no. 2 (November): 1264–1293.

Mandel, Kaisey, and Eric Agol. 2002. 'Analytic Light Curves for Planetary Transit Searches.' *ApJl* 580, no. 2 (December): L171–L175.

Mann, Andrew W., Trent Dupuy, Adam L. Kraus, et al. 2019. 'How to Constrain Your M Dwarf. II. The Mass-Luminosity-Metallicity Relation from 0.075 to 0.70 Solar Masses.' *ApJ* 871, no. 1 (January): 63.

Mann, Andrew W., Gregory A. Feiden, Eric Gaidos, Tabetha Boyajian, and Kaspar von Braun. 2015. 'How to Constrain Your M Dwarf: Measuring Effective Temperature, Bolometric Luminosity, Mass, and Radius.' *ApJ* 804, no. 1 (May): 64.

Mann, Andrew W., Elisabeth R. Newton, Aaron C. Rizzuto, et al. 2016. 'Zodiacal Exoplanets in Time (ZEIT). III. A Short-period Planet Orbiting a Pre-main-sequence Star in the Upper Scorpius OB Association.' *AJ* 152, no. 3 (September): 61.

Mantovan, G., L. Malavolta, S. Desidera, et al. 2024. 'The GAPS programme at TNG. XLIX. TOI-5398, the youngest compact multi-planet system composed of an inner sub-Neptune and an outer warm Saturn.' *A&A* 682 (February): A129.

Markov, Yu. G., L. V. Rykhlova, and I. N. Sinitsyn. 2010. 'Some improved methods for modeling the Earth's polar motion.' *Astronomy Reports* 54, no. 9 (September): 861–870.

Marsden, S. C., P. Petit, S. V. Jeffers, et al. 2014. 'A BCool magnetic snapshot survey of solar-type stars.' *MNRAS* 444, no. 4 (November): 3517–3536.

Martıénez-Arnáiz, R., J. Maldonado, D. Montes, C. Eiroa, and B. Montesinos. 2010. 'Chromospheric activity and rotation of FGK stars in the solar vicinity. An estimation of the radial velocity jitter.' *A&A* 520 (September): A79.

Mayo, Andrew W., Vinesh M. Rajpaul, Lars A. Buchhave, et al. 2019. 'An 11 Earth-mass, Long-period Sub-Neptune Orbiting a Sun-like Star.' *AJ* 158, no. 4 (October): 165.

Mayor, M., F. Pepe, D. Queloz, et al. 2003. 'Setting New Standards with HARPS.' *The Messenger* 114 (December): 20–24.

Mayor, M., and S. Udry. 2008. 'The quest for very low-mass planets.' *Physica Scripta Volume T* 130 (August): 014010.

Mayor, Michel, and Didier Queloz. 1995. 'A Jupiter-mass companion to a solar-type star.' *Nature* 378, no. 6555 (November): 355–359.

McLaughlin, D. B. 1924. 'Some results of a spectrographic study of the Algol system.' *ApJ* 60 (July): 22–31.

McQuillan, A., S. Aigrain, and T. Mazeh. 2013. 'Measuring the rotation period distribution of field M dwarfs with Kepler.' *MNRAS* 432, no. 2 (June): 1203–1216.

McQuillan, A., T. Mazeh, and S. Aigrain. 2014. 'Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample.' *ApJS* 211, no. 2 (April): 24.

Medina, Amber A., David Charbonneau, Jennifer G. Winters, Jonathan Irwin, and Jessica Mink. 2022. 'Variability Timescales of Hα on Active Mid-to-late M dwarfs.' *ApJ* 928, no. 2 (April): 185.

Melchior, Peter, Yan Liang, ChangHoon Hahn, and Andy Goulding. 2023. 'Autoencoding Galaxy Spectra. I. Architecture.' *AJ* 166, no. 2 (August): 74.

Mengel, M. W., S. C. Marsden, B. D. Carter, et al. 2017. 'A BCool survey of the magnetic fields of planet-hosting solar-type stars.' *MNRAS* 465, no. 3 (March): 2734–2747.

Meunier, N., M. Desort, and A. -M. Lagrange. 2010. 'Using the Sun to estimate Earth-like planets detection capabilities . II. Impact of plages.' *A&A* 512 (March): A39.

Meunier, N., and A. -M. Lagrange. 2020. 'Radial-velocity variations due to meridional flows in the Sun and solar-type stars: impact on exoplanet detectability.' *A&A* 638 (June): A54.

Meunier, N., A. -M. Lagrange, S. Borgniet, and M. Rieutord. 2015. 'Using the Sun to estimate Earth-like planet detection capabilities. VI. Simulation of granulation and supergranulation radial velocity and photometric time series.' *A&A* 583 (November): A118.

Meunier, N., A. -M. Lagrange, T. Boulet, and S. Borgniet. 2019. 'Activity time series of old stars from late F to early K. I. Simulating radial velocity, astrometry, photometry, and chromospheric emission.' *A&A* 627 (July): A56.

Meunier, Nadège. 2021. 'Stellar variability in radial velocity.' *arXiv e-prints* (April): arXiv:2104.06072.

Middelkoop, F. 1982. 'Magnetic structure in cool stars. IV - Rotation and CA II H and K emission of main-sequence stars.' *A&A* 107, no. 1 (March): 31–35.

Miklos, M., T. W. Milbourne, R. D. Haywood, et al. 2020. 'Testing the Spectroscopic Extraction of Suppression of Convective Blueshift.' *ApJ* 888, no. 2 (January): 117.

Milbourne, T. W., R. D. Haywood, D. F. Phillips, et al. 2019. 'HARPS-N Solar RVs Are Dominated by Large, Bright Magnetic Regions.' *ApJ* 874, no. 1 (March): 107.

Morbidelli, A., K. Baillié, K. Batygin, et al. 2022. 'Contemporary formation of early Solar System planetesimals at two distinct radial locations.' *Nature Astronomy* 6 (January): 72–79.

Mordasini, C., R. van Boekel, P. Mollière, Th. Henning, and Björn Benneke. 2016. 'The Imprint of Exoplanet Formation History on Observable Present-day Spectra of Hot Jupiters.' *ApJ* 832, no. 1 (November): 41.

Morrell, Sam, and Tim Naylor. 2019. 'Exploring the M-dwarf Luminosity-Temperature-Radius relationships using Gaia DR2.' MNRAS 489, no. 2 (October): 2615–2633.

———. 2020. 'Erratum: Exploring the M-dwarf Luminosity-Temperature-Radius relationships using Gaia DR2.' MNRAS 498, no. 4 (November): 5048–5050.

Morris, Robert L., Joseph D. Twicken, Jeffrey C. Smith, et al. 2020. *Kepler Data Processing Handbook: Photometric Analysis.* Kepler Science Document KSCI-19081-003, March.

Mortier, A., and A. Collier Cameron. 2017. 'Stacked Bayesian general Lomb-Scargle periodogram: Identifying stellar activity signals.' *A&A* 601 (May): A110.

Mortier, A., J. P. Faria, C. M. Correia, A. Santerne, and N. C. Santos. 2015. 'BGLS: A Bayesian formalism for the generalised Lomb-Scargle periodogram.' *A&A* 573 (January): A101.

Mortier, A., S. G. Sousa, V. Zh. Adibekyan, I. M. Brandão, and N. C. Santos. 2014. 'Correcting the spectroscopic surface gravity using transits and asteroseismology. No significant effect on temperatures or metallicities with ARES and MOOG in local thermodynamic equilibrium.' *A&A* 572 (December): A95.

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

Motalebi, F., S. Udry, M. Gillon, et al. 2015. 'The HARPS-N Rocky Planet Search. I. HD 219134 b: A transiting rocky planet in a multi-planet system at 6.5 pc from the Sun.' *A&A* 584 (December): A72.

Murphy, Kevin P. 2013. *Machine Learning: a Probabilistic Perspective.* Cambridge, Mass: MIT Press.

Nardetto, N., D. Mourard, P. Kervella, et al. 2006. 'High resolution spectroscopy for Cepheids distance determination. I. Line asymmetry.' *A&A* 453, no. 1 (July): 309–319.

Nardiello, D., L. Malavolta, S. Desidera, et al. 2022. 'The GAPS Programme at TNG. XXXVII. A precise density measurement of the young ultra-short period planet TOI-1807 b.' *A&A* 664 (August): A163.

Nava, Chantanelle, Mercedes López-Morales, Raphaëlle D. Haywood, and Helen A. C. Giles. 2020. 'Exoplanet Imitators: A Test of Stellar Activity Behavior in Radial Velocity Signals.' *AJ* 159, no. 1 (January): 23.

Nava, Chantanelle, Mercedes López-Morales, Annelies Mortier, et al. 2022. 'K2-79b and K2-222b: Mass Measurements of Two Small Exoplanets with Periods beyond 10 days that Overlap with Periodic Magnetic Activity Signals.' *AJ* 163, no. 2 (February): 41.

Nayek, Rajdip, Souvik Chakraborty, and Sriram Narasimhan. 2019. 'A Gaussian process latent force model for joint input-state estimation in linear structural systems.' *Mechanical Systems and Signal Processing* 128 (August): 497–530.

Ng, Kim Kwee. 2016. 'Prediction Methods in Solar Sunspots Cycles.' *Scientific Reports* 6 (February): 21028.

Nicholson, B. A., and S. Aigrain. 2022. 'Quasi-periodic Gaussian processes for stellar activity: From physical to kernel parameters.' *MNRAS* 515, no. 4 (October): 5251–5266.

Norris, Charlotte M., Benjamin Beeck, Yvonne C. Unruh, et al. 2017. 'Spectral variability of photospheric radiation due to faculae. I. The Sun and Sun-like stars.' *A&A* 605 (September): A45.

Norton, Aimee, Rachel Howe, Lisa Upton, and Ilya Usoskin. 2023. 'Solar Cycle Observations.' *SSR* 219, no. 8 (December): 64.

Noyes, R. W., N. O. Weiss, and A. H. Vaughan. 1984. 'The relation between stellar rotation rate and activity cycle periods.' *ApJ* 287 (December): 769–773.

Obridko, V. N., and B. D. Shelting. 1992. 'Cyclic Variation of the Global Magnetic Field Indices.' *SolPhys* 137, no. 1 (January): 167–177.

Oláh, K., Z. Kolláth, T. Granzer, et al. 2009. 'Multiple and changing cycles of active stars. II. Results.' *A&A* 501, no. 2 (July): 703–713.

Osborn, H. P., M. Ansdell, Y. Ioannou, et al. 2020. 'Rapid classification of TESS planet candidates with convolutional neural networks.' *A&A* 633 (January): A53.

Otegi, J. F., F. Bouchy, and R. Helled. 2020. 'Revisited mass-radius relations for exoplanets below 120 $M_{\oplus}$.' *A&A* 634 (February): A43.

Ould-Elhkim, M., C. Moutou, J. -F. Donati, et al. 2023. 'Wapiti: A data-driven approach to correct for systematics in RV data. Application to SPIRou data of the planet-hosting M dwarf GJ 251.' *A&A* 675 (July): A187.

Owen, James E., and Fred C. Adams. 2019. 'Effects of magnetic fields on the location of the evaporation valley for low-mass exoplanets.' *MNRAS* 490, no. 1 (November): 15–20.

Palumbo, III, Michael L., Eric B. Ford, Jason T. Wright, et al. 2022. 'GRASS: Distinguishing Planet-induced Doppler Signatures from Granulation with a Synthetic Spectra Generator.' *AJ* 163, no. 1 (January): 11.

Papaloizou, J., and J. E. Pringle. 1978. 'Non-radial oscillations of rotating stars and their relevance to the short-period oscillations of cataclysmic variables.' *MNRAS* 182 (February): 423–442.

Pavliotis, G.A. 2015. *Stochastic Processes and applications.*

Pepe, F., S. Cristiani, R. Rebolo, et al. 2021. 'ESPRESSO at VLT. On-sky performance and first results.' *A&A* 645 (January): A96.

Pepe, F., M. Mayor, G. Rupprecht, et al. 2002. 'HARPS: ESO's coming planet searcher. Chasing exoplanets with the La Silla 3.6-m telescope.' *The Messenger* 110 (December): 9–14.

Pepper, Joshua, Stephen R. Kane, Joseph E. Rodriguez, et al. 2020. 'TESS Reveals HD 118203 b to be a Transiting Planet.' *AJ* 159, no. 6 (June): 243.

Perruchot, S., D. Kohler, F. Bouchy, et al. 2008. 'The SOPHIE spectrograph: design and technical key-points for high throughput and high stability.' In *Ground-based and Airborne Instrumentation for Astronomy II,* edited by Ian S. McLean and Mark M. Casali, vol. 7014, 70140J. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July.

Perryman, Michael. 2018. *The Exoplanet Handbook.*

Pesnell, W. Dean, B. J. Thompson, and P. C. Chamberlin. 2012. 'The Solar Dynamics Observatory (SDO).' *SolPhys* 275, nos. 1-2 (January): 3–15.

Petit, P., B. Dintrans, S. K. Solanki, et al. 2008. 'Toroidal versus poloidal magnetic fields in Sun-like stars: a rotation threshold.' *MNRAS* 388, no. 1 (July): 80–88.

Petrovich, Cristobal. 2015. 'Hot Jupiters from Coplanar High-eccentricity Migration.' *ApJ* 805, no. 1 (May): 75.

Petrovich, Cristobal, and Scott Tremaine. 2016. 'Warm Jupiters from Secular Planet-Planet Interactions.' *ApJ* 829, no. 2 (October): 132.

Phillips, David F., Alex G. Glenday, Xavier Dumusque, et al. 2016. 'An astro-comb calibrated solar telescope to search for the radial velocity signature of Venus.' In *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II,* edited by Ramón Navarro and James H. Burge, vol. 9912, 99126Z. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July.

Pinamonti, M., M. Damasso, F. Marzari, et al. 2018. 'The HADES RV Programme with HARPS-N at TNG. VIII. GJ15A: a multiple wide planetary system sculpted by binary interaction.' *A&A* 617 (September): A104.

Piskunov, N., and O. Kochukhov. 2002. 'Doppler Imaging of stellar magnetic fields. I. Techniques.' *A&A* 381 (January): 736–756.

Pollacco, D., I. Skillen, A. Collier Cameron, et al. 2008. 'WASP-3b: a strongly irradiated transiting gas-giant planet.' *MNRAS* 385, no. 3 (April): 1576–1584.

Pollacco, D. L., I. Skillen, A. Collier Cameron, et al. 2006. 'The WASP Project and the SuperWASP Cameras.' *PASP* 118, no. 848 (October): 1407–1418.

Proctor, M. R. E., and N. O. Weiss. 1982. 'Magnetoconvection.' *Reports on Progress in Physics* 45, no. 11 (November): 1317–1379.

Provost, J., G. Berthomieu, and A. Rocca. 1981. 'Low Frequency Oscillations of a Slowly Rotating Star - Quasi Toroidal Modes.' *A&A* 94 (January): 126.

Quanz, S. P., M. Ottiger, E. Fontanet, et al. 2022. 'Large Interferometer For Exoplanets (LIFE). I. Improved exoplanet detection yield estimates for a large mid-infrared space-interferometer mission.' *A&A* 664 (August): A21.

Queloz, D., F. Bouchy, C. Moutou, et al. 2009. 'The CoRoT-7 planetary system: two orbiting super-Earths.' *A&A* 506, no. 1 (October): 303–319.

Queloz, D., A. Eggenberger, M. Mayor, et al. 2000. 'Detection of a spectroscopic transit by the planet orbiting the star HD209458.' *A&A* 359 (July): L13–L17.

Queloz, D., G. W. Henry, J. P. Sivan, et al. 2001. 'No planet for HD 166435.' *A&A* 379 (November): 279–287.

Quirrenbach, A., P. J. Amado, J. A. Caballero, et al. 2016. 'CARMENES: an overview six months after first light.' In *Ground-based and Airborne Instrumentation for Astronomy VI,* edited by Christopher J. Evans, Luc Simard, and Hideki Takami, 9908:990812. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. August.

Rajpaul, V., S. Aigrain, M. A. Osborne, S. Reece, and S. Roberts. 2015. 'A Gaussian process framework for modelling stellar activity signals in radial velocity data.' *MNRAS* 452, no. 3 (September): 2269–2291.

Rajpaul, V. M., S. Aigrain, and L. A. Buchhave. 2020. 'A robust, template-free approach to precise radial velocity extraction.' *MNRAS* 492, no. 3 (March): 3960–3983.

Rajpaul, V. M., L. A. Buchhave, G. Lacedelli, et al. 2021. 'A HARPS-N mass for the elusive Kepler-37d: a case study in disentangling stellar activity and planetary signals.' *MNRAS* 507, no. 2 (October): 1847–1868.

Rasmussen, Carl Edward, and Christopher K. I. Williams. 2006. *Gaussian Processes for Machine Learning.*

Rein, H., and S. -F. Liu. 2012. 'REBOUND: an open-source multi-purpose N-body code for collisional dynamics.' *A&A* 537 (January): A128.

Rein, Hanno, and David S. Spiegel. 2015. 'IAS15: a fast, adaptive, high-order integrator for gravitational dynamics, accurate to machine precision over a billion orbits.' *MNRAS* 446, no. 2 (January): 1424–1437.

Reiners, A. 2009. 'Activity-induced radial velocity jitter in a flaring M dwarf.' *A&A* 498, no. 3 (May): 853–861.

Reiners, Ansgar. 2012. 'Observations of Cool-Star Magnetic Fields.' *Living Reviews in Solar Physics* 9, no. 1 (February): 1.

Reiners, Ansgar, Nandan Joshi, and Bertrand Goldman. 2012. 'A Catalog of Rotation and Activity in Early-M Stars.' *AJ* 143, no. 4 (April): 93.

Remes, Sami, Markus Heinonen, and Samuel Kaski. 2017. 'Non-Stationary Spectral Kernels.' *arXiv e-prints* (May): arXiv:1705.08736.

Rempel, M., M. Schüssler, R. H. Cameron, and M. Knölker. 2009. 'Penumbral Structure and Outflows in Simulated Sunspots.' *Science* 325, no. 5937 (July): 171.

Rescigno, F., B. Dixon, and R. D. Haywood. 2023a. *MAGPy-RV: Gaussian Process regression pipeline with MCMC parameter searching.* Astrophysics Source Code Library, record ascl:2310.006, October.

Rescigno, F., G. Hébrard, A. Vanderburg, et al. 2023b. 'A hot mini-Neptune and a temperate, highly eccentric sub-Saturn around the bright K-dwarf TOI-2134.' *MNRAS* (October).

Rescigno, F., A. Mortier, X. Dumusque, et al. 2024. 'The Mean Longitudinal Magnetic Field and its Uses in Radial-Velocity Surveys.' Sumbitted to MNRAS in March 2024.

Rice, K., L. Malavolta, A. Mayo, et al. 2019. 'Masses and radii for the three super-Earths orbiting GJ 9827, and implications for the composition of small exoplanets.' *MNRAS* 484, no. 3 (April): 3731–3745.

Ricker, George R., Joshua N. Winn, Roland Vanderspek, et al. 2015. 'Transiting Exoplanet Survey Satellite (TESS).' *Journal of Astronomical Telescopes, Instruments, and Systems* 1 (January): 014003.

Rincon, François, and Michel Rieutord. 2018. 'The Sun's supergranulation.' *Living Reviews in Solar Physics* 15, no. 1 (September): 6.

Robert, Christian P., and George Casella. 2005. 'Monte Carlo Statistical Methods (Springer Texts in Statistics),' 113. Berlin, Heidelberg: Springer-Verlag.

Roberts, S., M. Osborne, M. Ebden, et al. 2013. 'Gaussian processes for time-series modelling.' *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (1984): 20110550.

Robertson, D. R. S., L. C. Gallo, A. Zoghbi, and A. C. Fabian. 2015. 'Searching for correlations in simultaneous X-ray and UV emission in the narrow-line Seyfert 1 galaxy 1H 0707-495.' *MNRAS* 453, no. 4 (November): 3455–3460.

Robertson, Paul, Suvrath Mahadevan, Michael Endl, and Arpita Roy. 2014. 'Stellar activity masquerading as planets in the habitable zone of the M dwarf Gliese 581.' *Science* 345, no. 6195 (July): 440–444.

Rogers, Leslie A. 2015. 'Most 1.6 Earth-radius Planets are Not Rocky.' *ApJ* 801, no. 1 (March): 41.

Rossiter, R. A. 1924. 'On the detection of an effect of rotation during eclipse in the velocity of the brigher component of beta Lyrae, and on the constancy of velocity of this system.' *ApJ* 60 (July): 15–21.

Roy, Arpita, Abhijit Chakraborty, Suvrath Mahadevan, et al. 2016. 'Precision velocimetry planet hunting with PARAS: current performance and lessons to inform future extreme precision radial velocity instruments.' In *Ground-based and Airborne Instrumentation for Astronomy VI,* edited by Christopher J. Evans, Luc Simard, and Hideki Takami, vol. 9908, 99086R. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. August.

Saar, Steven H., and Robert A. Donahue. 1997. 'Activity-Related Radial Velocity Variation in Cool Stars.' *ApJ* 485, no. 1 (August): 319–327.

Santerne, A., R. F. Dıéaz, C. Moutou, et al. 2012. 'SOPHIE velocimetry of Kepler transit candidates. VII. A false-positive rate of 35% for Kepler close-in giant candidates.' *A&A* 545 (September): A76.

Santos, A. R. G., S. Mathur, R. A. Garcıéa, M. S. Cunha, and P. P. Avelino. 2021. 'On the relation between active-region lifetimes and the autocorrelation function of light curves.' *MNRAS* 508, no. 1 (November): 267–278.

Santos, N. C., A. Mortier, J. P. Faria, et al. 2014. 'The HARPS search for southern extra-solar planets. XXXV. The interesting case of HD 41248: stellar activity, no planets?' *A&A* 566 (June): A35.

Scargle, J. D. 1982. 'Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data.' *ApJ* 263 (December): 835–853.

Scherrer, P. H., J. Schou, R. I. Bush, et al. 2012. 'The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO).' *SolPhys* 275, nos. 1-2 (January): 207–227.

Scherrer, Philip H., John M. Wilcox, and Robert Howard. 1972. 'The Mean Photospheric Magnetic Field from Solar Magnetograms: Comparisons with the Interplanetary Magnetic Field.' *SolPhys* 22, no. 2 (February): 418–424.

Schou, J., P. H. Scherrer, R. I. Bush, et al. 2012. 'Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) Instrument on the Solar Dynamics Observatory (SDO).' *SolPhys* 275, nos. 1-2 (January): 229–259.

Schrijver, C. J. 2002. 'Solar spots as prototypes for stellar spots.' *Astronomische Nachrichten* 323 (July): 157–164.

Schrijver, Carolus J., and Cornelis Zwaan. 2000. *Solar and Stellar Magnetic Activity.*

Schwab, C., A. Rakich, Q. Gong, et al. 2016. 'Design of NEID, an extreme precision Doppler spectrograph for WIYN.' In *Ground-based and Airborne Instrumentation for Astronomy VI,* edited by Christopher J. Evans, Luc Simard, and Hideki Takami, vol. 9908, 99087H. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. August.

See, T. J. J. 1896. 'Researches on the orbit of 70 Ophiuchi, and on a periodic perturbation in the motion of the system arising from the action of an unseen body.' *AJ* 16 (January): 17–23.

Semel, M. 1989. 'Zeeman-Doppler imaging of active stars. I - Basic principles.' *A&A* 225 (November): 456–466.

Sergison, Darryl J., Tim Naylor, S. P. Littlefair, Cameron P. M. Bell, and C. D. H. Williams. 2020. 'Characterizing the I-band variability of YSOs over six orders of magnitude in time-scale.' *MNRAS* 491, no. 4 (February): 5035–5055.

Serrano, L. M., S. C. C. Barros, M. Oshagh, et al. 2018. 'Distinguishing the albedo of exoplanets from stellar activity.' *A&A* 611 (March): A8.

Severny, A. B. 1971. 'Time Fluctuations of the General Magnetic Field of the Sun.' *QJRAS* 12 (March): 363.

Shallue, Christopher J., and Andrew Vanderburg. 2018. 'Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90.' *AJ* 155, no. 2 (February): 94.

Sheeley, Jr., N. R., and Y. -M. Wang. 2015. 'The Recent Rejuvenation of the Sun's Large-scale Magnetic Field: A Clue for Understanding Past and Future Sunspot Cycles.' *ApJ* 809, no. 2 (August): 113.

Shepherd, Simon J., Sergei I. Zharkov, and Valentina V. Zharkova. 2014. 'Prediction of Solar Activity from Solar Background Magnetic Field Variations in Cycles 21-23.' *ApJ* 795, no. 1 (November): 46.

Siegel, Jared C., Ryan A. Rubenzahl, Samuel Halverson, and Andrew W. Howard. 2022. 'Into the Depths: A New Activity Metric for High-precision Radial Velocity Measurements Based on Line Depth Variations.' *AJ* 163, no. 6 (June): 260.

Simola, U., X. Dumusque, and J. Cisewski-Kehe. 2019. 'Measuring precise radial velocities and cross-correlation function line-profile variations using a Skew Normal density.' *A&A* 622 (February): A131.

Simon, G. W., and R. B. Leighton. 1964. 'Velocity Fields in the Solar Atmosphere. III. Large-Scale Motions, the Chromospheric Network, and Magnetic Fields.' *ApJ* 140 (October): 1120.

Simonetti, J. H., J. M. Cordes, and D. S. Heeschen. 1985. 'Flicker of extragalactic radio sources at two frequencies.' *ApJ* 296 (September): 46–59.

Skrutskie, M. F., R. M. Cutri, R. Stiening, et al. 2006. ‘The Two Micron All Sky Survey (2MASS).’ *AJ* 131 (February): 1163–1183.

Snodgrass, Herschel B., and Roger K. Ulrich. 1990. ‘Rotation of Doppler Features in the Solar Photosphere.’ *ApJ* 351 (March): 309.

Solanki, Sami K. 2003. ‘Sunspots: An overview.’ *A&AR* 11, nos. 2-3 (January): 153–286.

Stalport, M., M. Cretignier, S. Udry, et al. 2023. ‘A review of planetary systems around HD 99492, HD 147379, and HD 190007 with HARPS-N.’ *A&A* 678 (October): A90.

Stalport, M., J. -B. Delisle, S. Udry, et al. 2022. ‘A general stability-driven approach for the refinement of multi-planet systems.’ *A&A* 664 (August): A53.

Stassun, Keivan G., Ryan J. Oelkers, Martin Paegert, et al. 2019. ‘The Revised TESS Input Catalog and Candidate Target List.’ *AJ* 158, no. 4 (October): 138.

Stassun, Keivan G., Ryan J. Oelkers, Joshua Pepper, et al. 2018. ‘The TESS Input Catalog and Candidate Target List.’ *AJ* 156, no. 3 (September): 102.

Stein, Robert F. 2012. ‘Solar Surface Magneto-Convection.’ *Living Reviews in Solar Physics* 9, no. 1 (July): 4.

Stephenson, C. B. 1986. ‘Dwarf K and M stars of high proper motion found in a hemispheric survey.’ *AJ* 92 (July): 139–165.

Strassmeier, Klaus G. 2009. ‘Starspots.’ *A&AR* 17, no. 3 (September): 251–308.

Strous, Louis H., and Cornelis Zwaan. 1999. ‘Phenomena in an Emerging Active Region. II. Properties of the Dynamic Small-Scale Structure.’ *ApJ* 527, no. 1 (December): 435–444.

Struve, O. 1952. ‘Proposal for a project of high-precision stellar radial velocity work.’ *The Observatory* 72 (October): 199–200.

Suárez Mascareño, A., J. P. Faria, P. Figueira, et al. 2020. ‘Revisiting Proxima with ESPRESSO.’ *A&A* 639 (July): A77.

Sugiura, Nariaki. 1978. ‘Further analysts of the data by Akaike’ s information criterion and the finite corrections.’ *Communications in Statistics - Theory and Methods* 7 (1): 13–26.

SunPy Community, The, Stuart J. Mumford, Steven Christe, et al. 2015. 'SunPy—Python for solar physics.' *Computational Science and Discovery* 8, no. 1 (January): 014009.

Suzuki, Y. 1967. 'On the Wilson Effect of the Sunspots.' *PASPJ* 19 (January): 220.

Ter Braak, Cajo JF. 2006. 'A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces.' *Statistics and Computing* 16 (3): 239–249.

The LUVOIR Team. 2019. 'The LUVOIR Mission Concept Study Final Report.' *arXiv e-prints* (December): arXiv:1912.06219.

Thomas, John H., and Nigel O. Weiss. 2008. *Sunspots and Starspots.*

Thompson, A. P. G., C. A. Watson, E. J. W. de Mooij, and D. B. Jess. 2017. 'The changing face of α Centauri B: probing plage and stellar activity in K dwarfs.' *MNRAS* 468, no. 1 (June): L16–L20.

Thompson, Samantha J., Didier Queloz, Isabelle Baraffe, et al. 2016. 'HARPS3 for a roboticized Isaac Newton Telescope.' In *Ground-based and Airborne Instrumentation for Astronomy VI,* edited by Christopher J. Evans, Luc Simard, and Hideki Takami, vol. 9908, 99086F. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. August.

Thompson, W. T. 2006. 'Coordinate systems for solar image data.' *A&A* 449, no. 2 (April): 791–803.

Thorngren, Daniel P., Jonathan J. Fortney, Ruth A. Murray-Clay, and Eric D. Lopez. 2016. 'The Mass-Metallicity Relation for Giant Planets.' *ApJ* 831, no. 1 (November): 64.

Tran, Quang H., Megan Bedell, Daniel Foreman-Mackey, and Rodrigo Luger. 2023. 'Joint Modeling of Radial Velocities and Photometry with a Gaussian Process Framework.' *ApJ* 950, no. 2 (June): 162.

Trifonov, Trifon, Lev Tal-Or, Mathias Zechmeister, et al. 2020. 'Public HARPS radial velocity database corrected for systematic errors.' *A&A* 636 (April): A74.

Trippe, Sascha. 2014. 'Polarization and Polarimetry: a Review.' *Journal of Korean Astronomical Society* 47, no. 1 (February): 15–39.

Turmon, M., J. T. Hoeksema, X. Sun, and M. Bobra. 2012. 'HARPs: Tracked Active Region Patch Data Product from SDO/HMI.' In *AGU Fall Meeting Abstracts,* vol. 2012, SH13A–2246. December.

Twicken, J. D., B. D. Clarke, S. T. Bryson, et al. 2010. 'Photometric Analysis in the Kepler Science Operations Center Pipeline.' In *Software and Cyberinfrastructure for Astronomy,* 7740:774023. Proc. SPIE. July.

Udalski, A., K. Zebrun, M. Szymanski, et al. 2002. 'The Optical Gravitational Lensing Experiment. Search for Planetary and Low- Luminosity Object Transits in the Galactic Disk. Results of 2001 Campaign – Supplement.' *ACTAA* 52 (June): 115–128.

Udry, S., M. Mayor, and N. C. Santos. 2003. 'Statistical properties of exoplanets. I. The period distribution: Constraints for the migration scenario.' *A&AP* 407 (August): 369–376.

Ulmer-Moll, S., M. Lendl, S. Gill, et al. 2022. 'Two long-period transiting exoplanets on eccentric orbits: NGTS-20 b (TOI-5152 b) and TOI-5153 b.' *A&A* 666 (October): A46.

Van Driel-Gesztelyi, Lidia, and Lucie May Green. 2015. 'Evolution of Active Regions.' *Living Reviews in Solar Physics* 12, no. 1 (September): 1.

Vanderburg, Andrew, David W. Latham, Lars A. Buchhave, et al. 2016. 'Planetary Candidates from the First Year of the K2 Mission.' *ApJ* 222, no. 1 (January): 14.

Vanderburg, Andrew, Andrew W. Mann, Aaron Rizzuto, et al. 2018. 'Zodiacal Exoplanets in Time (ZEIT). VII. A Temperate Candidate Super-Earth in the Hyades Cluster.' *AJ* 156, no. 2 (August): 46.

VanderPlas, Jacob T. 2018. 'Understanding the Lomb-Scargle Periodogram.' *ApJs* 236, no. 1 (May): 16.

Vaughan, A. H. 1983. 'The Mount Wilson program for stellar activity cycles.' In *Solar and Stellar Magnetic Fields: Origins and Coronal Effects,* edited by J. O. Stenflo, 102:113–131. January.

Walton, Stephen R. 1987. 'Flux-Tube Models of Solar Plages.' *ApJ* 312 (January): 909.

Wang, Jie. 2023. 'An Intuitive Tutorial to Gaussian Process Regression.' *Computing in Science and Engineering* 25, no. 4 (July): 4–11.

Wang, Y. -M., and E. Robbrecht. 2011. 'Asymmetric Sunspot Activity and the Southward Displacement of the Heliospheric Current Sheet.' *ApJ* 736, no. 2 (August): 136.

Welch, P. D. 1967. 'The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms.' *IEEE Transactions on Audio and Electroacoustics* 15 (2): 70–73.

West, Richard G., Edward Gillen, Daniel Bayliss, et al. 2019. 'NGTS-4b: A sub-Neptune transiting in the desert.' *MNRAS* 486, no. 4 (July): 5094–5103.

Wilson, Andrew Gordon, David A. Knowles, and Zoubin Ghahramani. 2012. 'Gaussian process regression networks' [in English (US)]. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012,* 599–606. Proceedings of the 29th International Conference on Machine Learning, ICML 2012. 29th International Conference on Machine Learning, ICML 2012 ; Conference date: 26-06-2012 Through 01-07-2012.

Wilson, D. M., M. Gillon, C. Hellier, et al. 2008. 'WASP-4b: A 12th Magnitude Transiting Hot Jupiter in the Southern Hemisphere.' *ApJ* 675, no. 2 (March): L113.

Wilson, O. C. 1968. 'Flux Measurements at the Centers of Stellar H- and K-Lines.' *ApJ* 153 (July): 221.

———. 1978. 'Chromospheric variations in main-sequence stars.' *ApJ* 226 (December): 379–396.

Wilson, Thomas G., Elisa Goffo, Yann Alibert, et al. 2022. 'A pair of sub-Neptunes transiting the bright K-dwarf TOI-1064 characterized with CHEOPS.' *MNRAS* 511, no. 1 (March): 1043–1071.

Winn, Joshua N. 2011. 'The Rossiter-McLaughlin effect for exoplanets.' In *The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution,* edited by Alessandro Sozzetti, Mario G. Lattanzi, and Alan P. Boss, 276:230–237. November.

Winn, Joshua N., Daniel Fabrycky, Simon Albrecht, and John Asher Johnson. 2010. 'Hot Stars with Hot Jupiters Have High Obliquities.' *ApJl* 718, no. 2 (August): L145–L149.

Winters, Jennifer G., Todd J. Henry, John C. Lurie, et al. 2015. 'The Solar Neighborhood. XXXV. Distances to 1404 m Dwarf Systems Within 25 pc in the Southern Sky.' *AJ* 149, no. 1 (January): 5.

Wise, A. W., S. E. Dodson-Robinson, K. Bevenour, and A. Provini. 2018. 'New Methods for Finding Activity-sensitive Spectral Lines: Combined Visual Identification and an Automated Pipeline Find a Set of 40 Activity Indicators.' *AJ* 156 (October): 180.

Wise, Alexander, Peter Plavchan, Xavier Dumusque, Heather Cegla, and Duncan Wright. 2022. 'Spectral Line Depth Variability in Radial Velocity Spectra.' *ApJ* 930, no. 2 (May): 121.

Wittenmyer, Robert A., Simon J. O'Toole, H. R. A. Jones, et al. 2010. 'The Frequency of Low-mass Exoplanets. II. The "Period Valley".' *ApJ* 722, no. 2 (October): 1854–1863.

Wolff, C. L., and J. B. Blizard. 1986. 'Properties of R-Modes in the Sun.' *SolPhys* 105, no. 1 (May): 1–15.

Wolszczan, A., and D. A. Frail. 1992. 'A planetary system around the millisecond pulsar PSR1257 + 12.' *Nature* 355, no. 6356 (January): 145–147.

Wright, Edward L., Peter R. M. Eisenhardt, Amy K. Mainzer, et al. 2010. 'The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance.' *AJ* 140, no. 6 (December): 1868–1881.

Xiang, N. B., and Z. N. Qu. 2016. 'Ensemble Empirical Mode Decomposition of the Magnetic Field of the Sun as a Star.' *AJ* 151, no. 3 (March): 76.

Xie, J. L., X. J. Shi, and J. C. Xu. 2017. 'Temporal Variation of the Rotation of the Solar Mean Magnetic Field.' *AJ* 153, no. 4 (April): 171.

Yeo, K. L., S. K. Solanki, and N. A. Krivova. 2013. 'Intensity contrast of solar network and faculae.' *A&A* 550 (February): A95.

Yuan, Ming, and Yi Lin. 2005. 'Model Selection and Estimation in Regression with Grouped Variables.' *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68, no. 1 (December): 49–67.

Zarro, D. M. 1983. 'The dependence of Hα on chromospheric activity in G and K main-sequence stars.' *ApJl* 267 (April): L61–L63.

Zechmeister, M., and M. Kürster. 2009. 'The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms.' *A&A* 496, no. 2 (March): 577–584.

Zechmeister, M., A. Reiners, P. J. Amado, et al. 2018. 'Spectrum radial velocity analyser (SERVAL). High-precision radial velocities and two alternative spectral indicators.' *A&A* 609 (January): A12.

Zeng, Li, Dimitar D. Sasselov, and Stein B. Jacobsen. 2016. 'Mass-Radius Relation for Rocky Planets Based on PREM.' *ApJ* 819, no. 2 (March): 127.

Zhang, Michael, Fei Dai, Jacob L. Bean, Heather A. Knutson, and Federica Rescigno. 2023. 'Outflowing Helium from a Mature Mini-Neptune.' *ApJl* 953, no. 2 (August): L25.

Zhao, J., Eric B. Ford, and C. G. Tinney. 2022. 'FIESTA II. Disentangling Stellar and Instrumental Variability from Exoplanetary Doppler Shifts in the Fourier Domain.' *ApJ* 935, no. 2 (August): 75.

Zhao, J., and C. G. Tinney. 2020. 'FIESTA - disentangling stellar variability from exoplanets in the Fourier domain.' *MNRAS* 491, no. 3 (January): 4131–4146.

Zhao, Lily L., Xavier Dumusque, Eric B. Ford, et al. 2023. 'The Extreme Stellar-signals Project. III. Combining Solar Data from HARPS, HARPS-N, EXPRES, and NEID.' *AJ* 166, no. 4 (October): 173.

Zhu, Chunming, Rui Liu, David Alexander, and R. T. James McAteer. 2016. 'Observation of the Evolution of a Current Sheet in a Solar Flare.' *ApJl* 821, no. 2 (April): L29.

Zhu, Wei, and Yanqin Wu. 2018. 'The Super Earth-Cold Jupiter Relations.' *AJ* 156, no. 3 (September): 92.

Zirin, Harold, and Robert Howard. 1966. 'The Structure of the Solar Chromosphere. II. Spectroheliograms in λ 10830 Å and Their Interpretation.' *ApJ* 146 (November): 367.