The mean longitudinal magnetic field and its uses in radial-velocity surveys

F. Rescigno [1]★ A. Mortier [2] X. Dumusque,[3] B. S. Lakeland [1] R. Haywood,[1] N. Piskunov,[4]
B. A. Nicholson,[5] M. López-Morales,[6] S. Dalal [1] M. Cretignier [7] B. Klein [7]
A. Collier Cameron [8,9] A. Ghedina,[10] M. Gonzalez,[10] R. Cosentino,[10] A. Sozzetti [11] and S. H. Saar[6]

[1]*Department of Astrophysics, University of Exeter, Stocker Rd, Exeter EX4 4QL, UK*
[2]*Department of Astrophysics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*
[3]*Observatoire de Geneéve, Université de Geneéve, Chemin de Pegasi, 51, CH-1290 Versoix, Switzerland*
[4]*Department of Physics and Astronomy, Uppsala University, Box 516, SE-75120 Uppsala, Sweden*
[5]*Centre for Astrophysics, University of Southern Queensland, West St, Toowoomba, QLD 4350, Australia*
[6]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[7]*Department of Physics, Sub-department of Astrophysics, University of Oxford, Oxford OX1 3RH, UK*
[8]*SUPA, School of Physics & Astronomy, University of St Andrews, North Haugh, St Andrews KY169SS, UK*
[9]*Centre for Exoplanet Science, University of St Andrews, North Haugh, St Andrews KY169SS, UK*
[10]*Fundación Galileo Galilei – INAF (Telescopio Nazionale Galileo), Rambla J. A. F. Perez 7, E-38712 Breña Baja (La Palma), Canary Islands, Spain*
[11]*INAF – Osservatorio Astrofisico di Torino, Strada Osservatorio, 20 I-10025 Pino Torinese (TO), Italy*

ABSTRACT

This work focuses on the analysis of the mean longitudinal magnetic field as a stellar activity tracer in the context of small exoplanet detection and characterization in radial-velocity (RV) surveys. We use Solar Dynamics Observatory/Helioseismic and Magnetic Imager filtergrams to derive Sun-as-a-star magnetic field measurements, and show that the mean longitudinal magnetic field is an excellent rotational period detector and a useful tracer of the solar magnetic cycle. To put these results into context, we 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. 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. We also test the mean longitudinal magnetic field as a 'stellar proxy' on a reduced solar data set to simulate stellar-like observational sampling. With a Gaussian Process regression analysis, we confirm that the solar mean longitudinal magnetic field is the most effective of the considered indicators, and is the most efficient rotational period indicator over different levels of stellar activity. This work highlights the need for polarimetric time series observations of stars.

Key words: methods: data analysis – techniques: radial velocities – Sun: activity – Sun: magnetic fields – planets and satellites: detection – stars: activity.

1 INTRODUCTION

In the last 20 yr, the radial-velocity (RV) method has been used to successfully detect and characterize hundreds of exoplanets, from blazing hot giants to rocky super-Earths. With the aim of finding potential Earth analogues that future missions such as the Habitable Worlds Observatory[1] (Harada et al. 2024) or the Large Interferometer for Exoplanets (LIFE: Quanz et al. 2022) can observe in search of biosignatures, the community is now more than ever targeting rocky exoplanets in their stellar habitable zone. The RV signals imprinted

by these planets on the light of their host stars are, however, of the order of tens of cm s^{-1} and until recently have been out of reach for most precise spectrographs.

As instrumentation reaches the required precision, the biggest challenge we now face for the detection of Earth analogues and the accurate measurements of their masses is stellar variability (Saar & Donahue 1997; Lindegren & Dravins 2003; Meunier, Desort & Lagrange 2010; Dumusque et al. 2011; Fischer et al. 2016; Crass et al. 2021; Meunier 2021). The RV variations generated by activity on the surface of host stars are often of the order of several m s^{-1} and they can easily mimic or completely drown planetary oscillations. Stellar-induced signals, in fact, often dominate the RVs of Sun-like stars. These signals are still challenging to model as they affect the time series over multiple time-scales, from minutes to years. Moreover, the longer baselines required for disentangling Earth-like signals introduce a further source of 'noise': magnetic cycles. Over

★ E-mail: f.rescigno@bham.ac.uk
[1]Based on the studies for the Large UV/Optical/IR Surveyor (LUVOIR: The LUVOIR Team 2019), and the Habitable Exoplanet Observatory (Gaudi et al. 2020).

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 characterization 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).

Great care is then required when accounting for and modelling stellar activity in order to obtain accurate orbital solutions and to precisely determine planetary masses. A common approach is to use Gaussian Processes (GP). GP regression has proven to be an effective tool for modelling stellar activity (e.g. Haywood et al. 2014; Rajpaul et al. 2015; Faria et al. 2016; Serrano et al. 2018; Barros et al. 2020). It, however, has its limitations. In particular, its ability to predict stellar variability is strongly reliant on an accurate detection of the stellar rotational period (Nicholson & Aigrain 2022). A precise determination of the periodicity of the stellar-induced RVs is vital to correctly differentiate them from Keplerian signals (Bortle et al. 2021), and to compute accurate masses (Blunt et al. 2023; Dalal et al. 2024). 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). 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 are therefore much harder to interpret correctly.

The accurate detection of the stellar rotation period is vital for many other areas of astrophysics beside exoplanet characterization. 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, Mazeh & Aigrain (2014) highlighted a bimodality in the rotation period–temperature relation of more than 30 000 *Kepler* targets. Amard, Roquette & Matt (2020) studied the contribution of stellar metallicity to the decay of 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, especially in times of low activity (Aigrain et al. 2015). We therefore use activity proxies as extra suppliers of information. Useful indicators are not sensitive to the Doppler shifts caused by the presence of planets. They instead map the spectral line distortions generated by the activity on the surface of the star or are generated by chromospheric emissions. Most commonly used activity proxies are either extracted from the same spectra as the RVs, such as the S-index, or are derived from the shape of their cross-correlation function (CCF), such as the full width at half-maximum (FWHM) and the bisector span (BIS). Nevertheless, even an analysis of these 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.

1.1 The mean longitudinal magnetic field

Quantities that directly measure stellar magnetic field properties may be more useful for measuring stellar rotation and activity-induced RV variability. 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 absolute 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).

In the model of light as an electromagnetic (EM) wave, light polarization is characterized by the four Stokes parameters: total intensity, I, two linear polarization parameters, Q and U, and one circular polarization parameter, V. Stokes Q corresponds to an EM wave propagating vertically or horizontally with respect to the line of sight. Stokes U is an EM wave propagating on a plane on a $\pm 45°$ angle from the Q axis. Stokes V describes light where there is a phase offset between the electric and magnetic parts of the EM wave, producing either right hand (clockwise) or left hand (anticlockwise). Typically, a single observation yields two Stokes parameters: I and one of polarizations. Light produced in the presence of a magnetic field becomes polarized through the Zeeman effect. In the case of measuring the polarization of light produced by stars, V traces the line-of-sight (longitudinal) component of the large-scale magnetic field, and Q and U measure the transverse components. While all components are needed to fully describe the magnetic field state, the measurement of linear polarization (Q and U) are more challenging for stars than circular (V). The observation of a single polarization state (Q, U, or V) requires four sub-exposures, meaning that linear polarization takes twice the telescope time to complete. Additionally, linear polarization is a second-order effect in the wavelength domain and, especially in the case of weak-fields, is a smaller component of the total intensity compared to V (Bagnulo & Landstreet 2015), and is more prone to spurious signals from surface reflections. While progress has been made for the estimation of magnetic fields (and in particular of unsigned magnetic fluxes) for other stars (see e.g. Lienhard et al. 2023; Kochukhov, Hackman & Lehtinen 2023), this remains challenging.

Given these difficulties, Stokes V observations are currently the more efficient way study changes in large-scale stellar magnetic fields by measuring the net line-of-sight signed magnetic flux, also called longitudinal magnetic flux. The mean longitudinal magnetic field, B_l (sometimes also referred to as $<B_z>$), 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 polarization as (e.g. Landstreet 1982):

$$\frac{V}{I} = g_{\text{eff}} C_z \lambda_0^2 \frac{1}{I} \frac{dI}{d\lambda} B_l, \tag{1}$$

where g_{eff} is the effective Landé factor, I is the intensity at wavelength λ, λ_0 is the average wavelength, and $C_z = 4.67 \times 10^{-13}$ Å^{-1} G^{-1}. The mean longitudinal field can then be expressed as the first-order moment of the Stokes V parameter as (Donati & Collier Cameron 1997):

$$B_l = -2.14 \times 10^{11} \frac{\int \nu V(\nu) d\nu}{\lambda_{\text{av}} g_{\text{av}} c \int [I_c - I(\nu)] d\nu}, \tag{2}$$

where λ_{av} and g_{av} refer to the average wavelength and the average Landé factor of the lines used to compute B_l, respectively, and I_c is the continuum intensity. 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.

Analyses of B_l time series have successfully determined rotation periods using traditional methods such as Lomb–Scargle periodograms. Studies done with B_l data from the near-infrared SPectropolarimétre InfraROUge (Donati et al. 2020) have detected rotational periods of chemically peculiar stars (Babcock 1949), and M-dwarfs (e.g. Landstreet 1992; Donati & 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; Nicholson et al. 2021; Marsden et al. 2023).Yu et al. (2019) and later Donati et al. (2023) also introduced Gaussian Process regression to the modelling of B_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 ($v \sin i < 2 \, \mathrm{km \, s^{-1}}$) makes the detection of 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 field. 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. They were able to detect the magnetic field in one-third of the sample, and of these stars 21 were Sun-like. With mostly a single observation per star, the survey reached precisions in B_l of 0.2 G, and demonstrated that B_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. We therefore lack the data to analyse the longitudinal magnetic field as a stellar activity tracer. This work will present proof to the need of polarimetric time series stellar observation. In order to circumvent this limitation and as previous studies have done to better understand stellar variability and its dependence to other measurable quantities (Haywood et al. 2016; Collier Cameron et al. 2019; Haywood et al. 2022), we turned to the best observed FGK star: the Sun.

The origin of B_l, in solar science often also called solar mean magnetic field or general magnetic field, is still strongly debated. Some attribute the largest contribution to the signal of B_l to the weak, large-scale magnetic flux over the entire visible disc (i.e. the background flux, or the 'quiet' Sun flux) (e.g. Severny 1971; Xiang & Qu 2016). In fact, Bose & Nagaraju (2018) claim that 80 per cent of the signal of B_l is generated by the background magnetic flux. Using resolved full-disc solar data, they partition the solar surface between sunspots, faculae, and background. They then calculate 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 we will further explore in Section 3). These results are in opposition to the ones of others. For example, Scherrer, Wilcox & Howard (1972) show 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, Abramenko & Yurchyshyn (2017) used a similar technique to Bose & Nagaraju (2018) on similar data, but recovered different results. They considered a magnitude threshold of 30 Mx cm^{-2} (equal to 30 G) and found that the B_l component derived from active regions

contributed from 65 to 95 per cent of the total field. They therefore claim that B_l is directly generated by magnetic flux concentrations, meaning spots, faculae, and network. They assert 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) section the solar surface with a magnetic flux concentration mask on the magnetogram, while Bose & Nagaraju (2018) separate 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 paper. We will instead focus on addressing how its behaviours can help us understand stellar activity in the Extreme Precision RV (EPRV) regime and pinpoint stellar characteristics needed for activity modelling.

In this work, we use resolved solar observations to extract the mean longitudinal magnetic field of the Sun, and compare it to the RVs and the common activity proxies of Sun-as-a-star observations. Since previous works highlighted the effectiveness of B_l to measure stellar characteristics such as differential rotation (Bruning 1991), the aim of this analysis is to determine whether B_l can be a useful tracer of stellar activity in Sun-like stars in the context of RV surveys. This paper is organized as follows: we describe the data in Section 2. We analyse the derived time series to better identify the properties and periodicities of B_l in Section 3. Section 4 covers how we 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. We conclude in Section 5.

2 DATA

2.1 HARPS-n Sun-as-a-star data

The High Accuracy Radial-velocity Planet Searcher for the Northern hemisphere (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 fibre into the HARPS-N spectrograph (Cosentino et al. 2012, 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 min 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 (Dumusque et al. 2021) which computes the CCF using a G2 stellar mask. From the CCFs, we also calculate the standard activity indicators: the FWHM and the BIS. Using the Ca H&K lines, we also measure the S-index.

Further corrections were applied to the data for them to better represent stellar observations. The details of these corrections, which we summarize here, are given 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. Next, the effects of differential extinction (noticeable due to the Sun being resolved on the sky) are then removed. The FWHM is corrected for the effects of the Earth's orbital eccentricity and obliquity, which makes the observed projected rotational velocity, and thus the spectral line widths, change over time. Finally, the S-index is corrected for ghosts on the CCD.
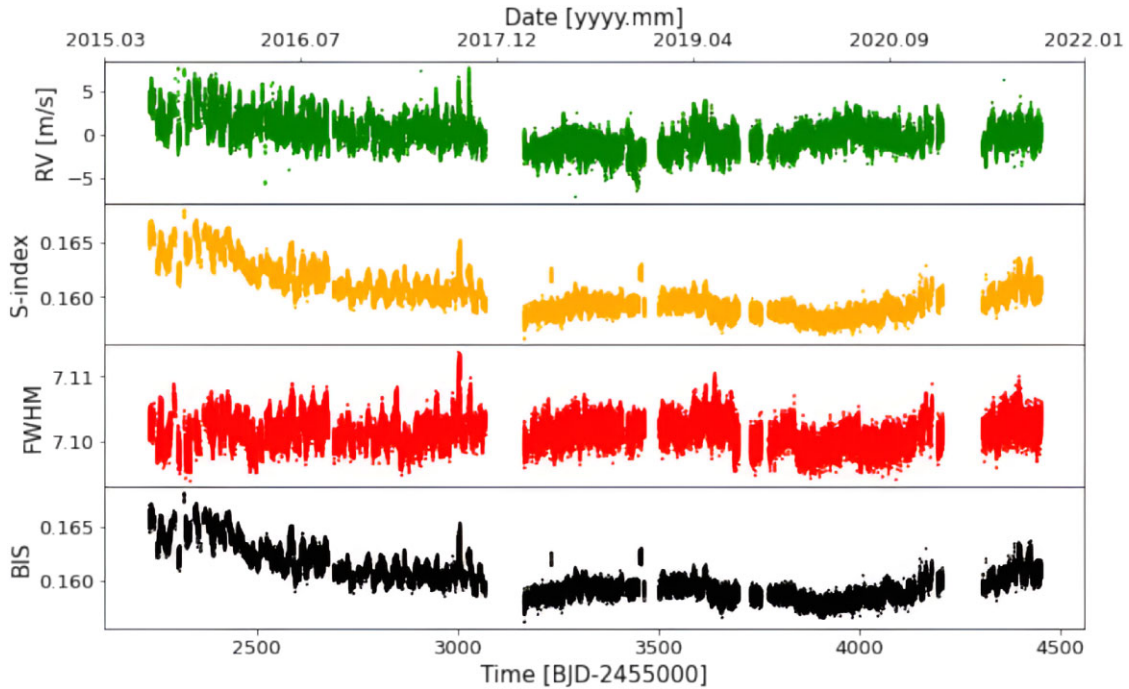
Figure 1. HARPS-N Solar telescope data. From the top panel, the corrected RVs in green, the S-index in orange, the FWHM in red, and the BIS in black. Uncertainties are included but too small to be visible.

Some of our observations will be affected by clouds or other bad weather. To select the best data, we apply strict cuts. The first cut is based on a data quality factor, Q_f, which is computed using a mixture-model where $Q_f = 0$ indicates the worst affected data, and $Q_f = 1$ the data not affected by clouds (Collier Cameron et al. 2019; Al Moulla et al. 2023). We use only data where $Q_f > 0.99$. We make a second cut directly using a metric of the exposure meter: the ratio, R, of the maximum and mean counts for every observation. Given the large flux of the solar light, the slight variations and delays in shutter speed of the HARPS-N instrument that are negligible for stellar observations result in a variation in total flux between exposures. This gives a Gaussian distribution in values of R over the full time series. We thus fit the distribution of R with a Gaussian and remove all data where R is higher than the mean plus three standard deviations ($R = 1.5$). As a last conservative cut, we perform a 5σ clipping on the remaining RVs.

In total, we considered 64 332 data points from BJD 2457232.873 (2015 July 29) to BJD 2459449.104 (2021 August 22). All the considered time series are shown in Fig. 1.

2.2 SDO/HMI resolved-sun images

The Solar Dynamics Observatory (SDO: Pesnell, Thompson & Chamberlin 2012) was launched in late 2010 by NASA's Living With a Star Program, a program designed to understand the causes of solar variability and its impacts on Earth. The aim of the SDO was specifically to study the solar atmosphere on short time-scales over many wavelengths. One of its three scientific instruments, the Helioseismic and Magnetic Imager (HMI: Schou et al. 2012; Scherrer et al. 2012) has been taking continuous full-disc observations of the solar surface with its two cameras of 4096×4096 pixels nearly without interruption since mid-2010. The SDO/HMI instrument has near single-granule resolution (Pesnell et al. 2012; Schou et al. 2012). It takes polarized filtergrams of the visible solar disc in two polarization

states by measuring six wavelengths centred in the 6173.3 Å neutral Fe I line (Couvidat et al. 2016). Observations are taken every 45 s, as well as compiled in 12 min (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.

In this work, we use the 720-s integrated SDO/HMI exposures of the continuum photometric intensity, the Dopplergrams, and the magnetograms reduced with the Vector Pipeline. An example is shown in Fig. 2. While the telescope produces near-continuous observations, we choose a cadence of four hours, yielding six images per 24-h period and 31 755 images spanning nearly 13 yr from BJD 2455318 (2010 May 1) to BJD 2459945 (2022 December 31).

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

While the data are corrected to account for most instrumental effects, long baseline analysis of the solar Doppler velocities was not the original aim of the SDO mission. For this reason long-term stability of the instrument was not prioritized. Therefore, studying the evolution of the solar activity over multiple months or years is not straightforward. To do so a framework was developed that scrambles full-disc images to Sun-as-a-star-like observations coupled with a weighting-based baseline (Haywood et al. 2016). A full pipeline named SOLASTER was introduced by Ervin et al. (2022). In this work, we will briefly cover how time series of RVs and B_l were generated. For more information refer to Meunier et al. (2010), Haywood et al. (2016, 2022), and Milbourne et al. (2019).
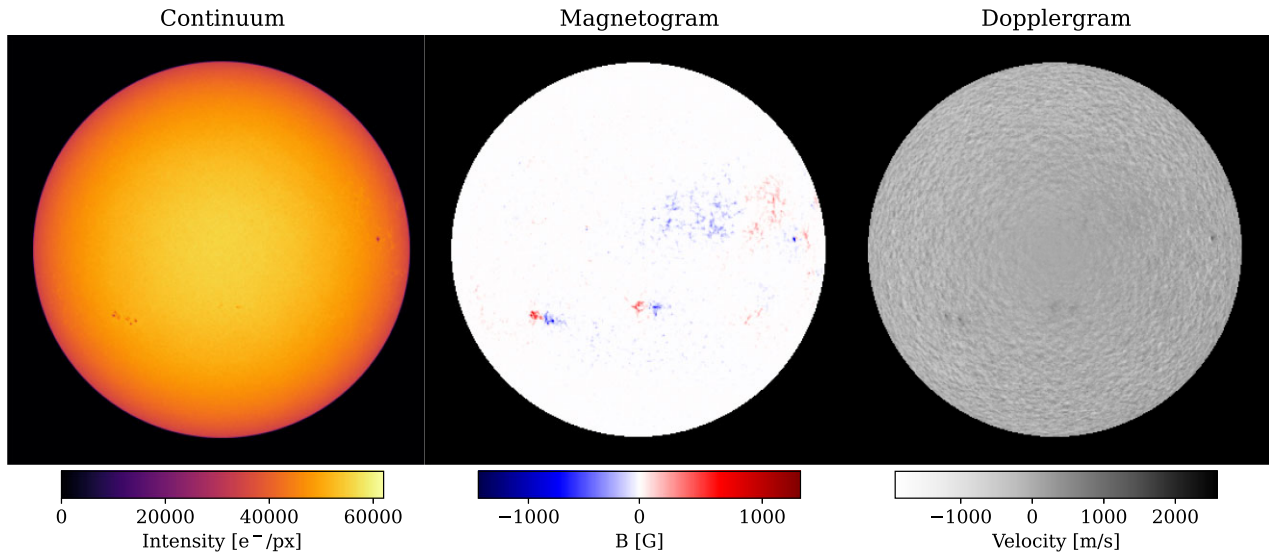
Figure 2. Example SDO/HMI images from 2015 July 29. From the left to right panels: the continuum intensity (uncorrected for limb darkening), the line-of-sight magnetogram, and the Dopplergram (corrected for the solar rotation and spacecraft motion).

Mean longitudinal magnetic field: 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 polarizations, V'_{LCP} and V'_{RCP}:

$$B_{los} = (V'_{LCP} - V'_{RCP})K_m, \qquad (3)$$

in which $K_m = 0.231405$ 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. 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, we compute the disc-averaged, longitudinal magnetic field of the Sun 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:

$$B_l = \frac{\sum_{ij} B_{los,ij} I_{ij}}{\sum_{ij} I_{ij}}, \qquad (4)$$

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. 3 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, Solanki & Krivova 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 4.1. The maximum field strength derived for the solar magnetic cycle (Cycle 24 and the beginning of Cycle 25) included in the data are 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 our comparison.

RV variations: given the lack of long-term stability for the Dopplergram data, we compute the disc-averaged RVs starting from a physically motivated model. We define the 'quiet' Sun average Doppler velocity as baseline and compute the RVs of each image relative to their respective 'quiet' Sun value. We define as active all solar surface larger than 20 ppm ($60 \, Mm^2$) with absolute foreshortening-corrected magnetic field larger than 24 G (see Haywood et al. 2016 and Milbourne et al. 2019). All remaining pixels are then defined 'quiet'. We can then compute the baseline value after correcting for the movement of the spacecraft by summing over all 'quiet' pixels. With this technique, all RV signals not directly induced by active regions, such as p modes, granulation or supergranulation, are not included. We are therefore only looking at active region-induced, rotationally modulated RV variations. These ΔRVs are computed as the linear combination of two active-regions contributions: δRV_{phot} (the signal generated by active regions breaking the red- and blueshift balance of the rotating disc by enhancing or diminishing the photometric intensity of their area) and δRV_{conv} (the signal generated by the suppression of convection in magnetic areas yielding to a decrease of the overall convective blueshift on the solar surface). The resulting ΔRVs are plotted in blue on the bottom of Fig. 3.

3 FULL TIME SERIES ANALYSIS: HOW DOES B_l RELATE TO THE RV VARIATIONS?

We begin the analysis by assessing the basic properties of the mean longitudinal magnetic field compared to the other derived time series, in order to investigate the behaviour of B_l over the Solar magnetic cycle. With this study, we aim to answer the following questions: Is B_l a direct proxy of activity-induced RV variations? How does B_l change with activity and does it behave like the RVs? What information can we extract from analysing B_l that cannot be derived from other common indicators? Can we use B_l to trace and model the solar magnetic cycle?

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, we compute the Spearman rank-order correlation coefficient of the two time series. When considering all 13 yr of
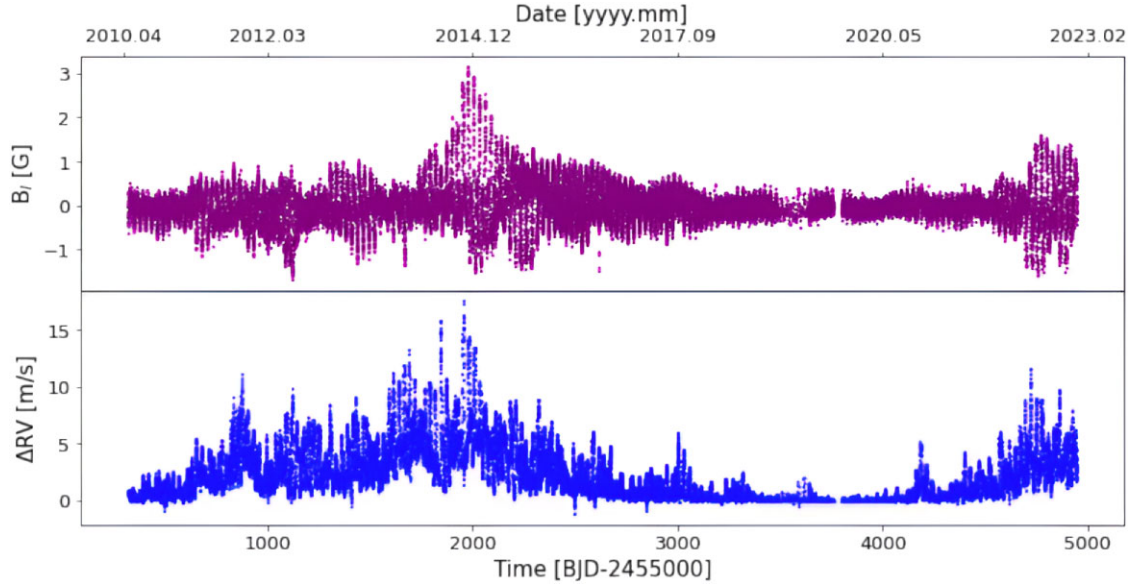
Figure 3. SDO/HMI-derived mean longitudinal magnetic field on the top panel, and the model RVs on the bottom panel. Uncertainties are not included as they would be too small to be visible.
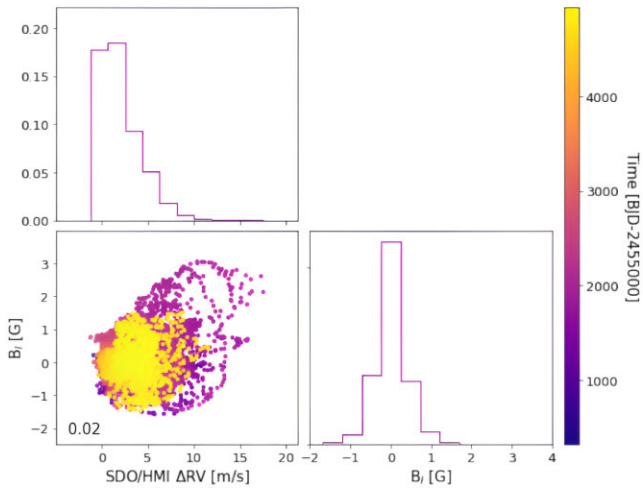

Figure 4. Correlation plot between the SDO/HMI-derived RVs and the mean longitudinal magnetic field. The colour indicates the Julian date of each data point. The computed Spearman rank correlation factor is also included.

data, we calculate a correlation coefficient of 0.02, as shown in Fig. 4, indicating that B_l does not correlate with the contemporaneous SDO/HMI-derived ΔRVs. We 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.

We 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 BIS, as plotted in the Appendix in Fig. A1. With the exception of the FWHM, the HARPS-N RVs 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 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 we will address in Section 4.2. This behaviour can be attributed to the fact that these proxies are sensitive to a mixture of different active regions, such as spots, faculae, and network (Cretignier, Pietrow & Aigrain 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, we 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 radial velocity over all time-scales. 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 between positive and negative values around a mean value of 0.02 G that is stable in time. The mean value of the ΔRVs changes with magnetic cycle phase, going roughly from 2.4 m s^{-1} at high activity, to 0.3 m s^{-1} during minimum.

We, however, notice a general trend shared between the two time series through the solar cycle. We 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 we extract two new time series: we 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, we repeat the same steps with window lengths between 1 d and 1 yr. The results of this analysis are plotted in the top panel of Fig. 5 as a purple solid line. We 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 d 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. 6. 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. They can therefore be used to correct for the long-term magnetic activity signal via techniques such
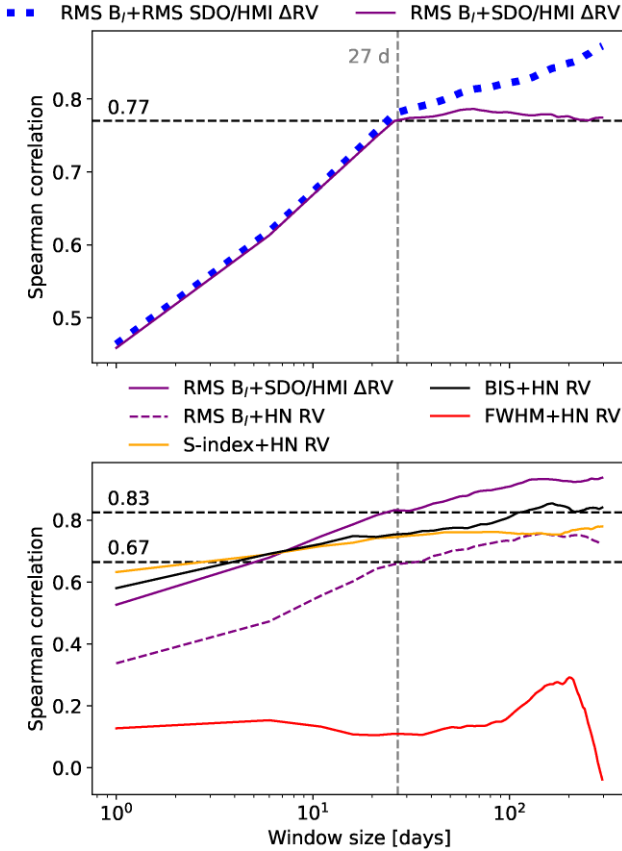
Figure 5. Spearman rank correlation coefficient between two time series against the size of the window (in days) used to smooth the signal (in logarithmic scale). Top panel: correlation between the RMS of B_l and the RMS of the ΔRVs in blue dotted line, and between the RMS of B_l and the time-aware mean of ΔRVs as a purple solid line. All considered time series are derived from SDO/HMI data and include all available observations. Bottom panel: Spearman correlation coefficient with varying window size. The time series considered have been matched following the method in Section 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 BIS (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 grey-dashed line. Horizontal black-dashed lines indicate the correlation coefficient achieved when smoothing over this window.
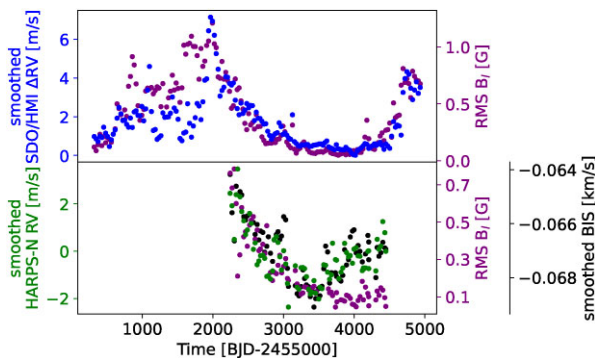
Figure 6. Top panel: time series of the time-aware mean over an averaging window of 27 d of the SDO/HMI ΔRVs in blue. In purple the RMS of B_l over the same window. Bottom panel: time series of the time-aware mean of the matched HARPS-N RVs in green, of the matched BIS in black, and the RMS of the matched B_l in purple over the a window of 27 d.

as contemporaneous fit, or can be selected as training set for a squared exponential kernel in a GP regression framework. As a simple test, we find the best-fitting sine function to the RMS over a window of 27 d of B_l. We then use the derived parameters to subtract the magnetic cycle long-term trend in the ΔRVs. This very rudimentary method is able to flatted the ΔRVs and reduce their RMS by more than 60 per cent.

3.2 Structure functions

We also compute the structure functions of all time series. The structure function measures the variability in a time series at each time-scale. It is calculated as

$$\mathrm{SF}(\tau) = \left\langle \left(f(t) - f(t-\tau) \right)^2 \right\rangle, \tag{5}$$

where the average is taken over all pairs of observations $f(t)$ separated by time τ. Lakeland et al. (2024) show that, for a continuous, uncorrelated signal, $\mathrm{SF}(\tau) = 2 \times \mathrm{RMS}^2$. We therefore follow the example of Lakeland et al. (2024) and use $\sqrt{\frac{1}{2}\mathrm{SF}}$ to quantify the variability at each time-scale, to better draw analogy with the RMS. We highlight here a few key properties of structure functions; for a more comprehensive review, see Simonetti, Cordes & Heeschen (1985), Sergison et al. (2020), Lakeland & Naylor (2022), and references therein. First, structure functions typically increase with time-scale. That is to say that two observations taken further apart are likely to be more different than two observations taken closer together. A notable exception is for a periodic signal, in which observations separated by the period or multiples thereof have similar values. In a structure function, such a periodicity manifests as sharp dips. The second important feature of a structure function is the location of the transition between the region of increasing variability and the plateau. This transition occurs once observations separated by greater time-scale no longer show more variation. The time-scale at which this occurs is the characteristic variability time-scale of the signal. For time-scales longer than this characteristic τ, no additional intrinsic variability is present. In practical terms, to effectively sample a specific source of variability, observations should be taken with time lags within the increasing region of the SF. In Fig. 7, we plot the structure functions of the time series shown in Figs 1 and 3. To ensure each structure function is well sampled (i.e. with many pairs of observations contributing to each SF calculation), we require at least 50 pairs of observations in each τ bin.

To allow for direct comparison, we only consider the SDO/HMI data over the overlapping years with HARPS-N (2015 to 2021). In Fig. 7, the higher cadence of HARPS-N is highlighted by the presence of data at shorter time-scales, while the SDO/HMI-derived time series have a minimum τ of 4 h. Both RV time series (HARPS-N in green, and SDO/HMI in blue) have similar structure function behaviours. They both increase until a time-scale of $\sim$10 d, they then grow at a significantly slower rate (forming plateaus of sorts), to finally start increasing more strongly after 100 d, 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 $\sim$10 d, and a plateau at a $\sqrt{\frac{1}{2}\mathrm{SF}}$ of $\sim$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. We 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 B_l at the rise and the decline
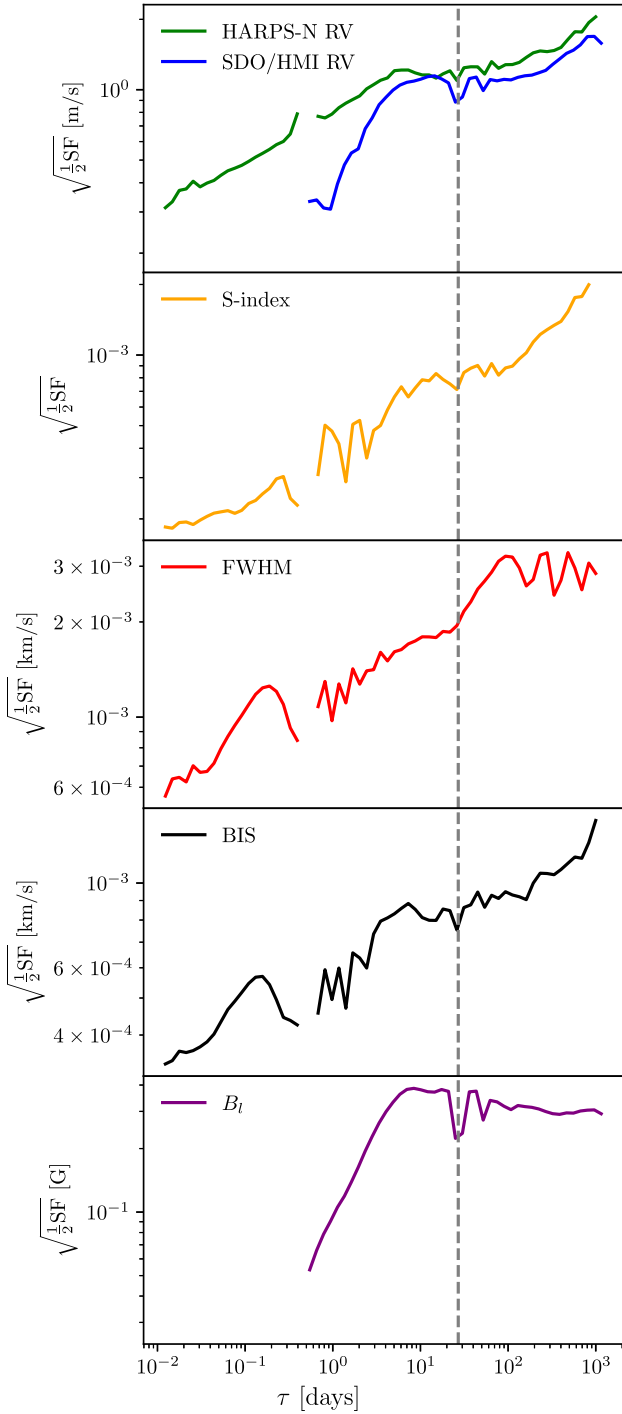
Figure 7. Structure functions of the time series shown in Figs 1 and 3. See the main text for more details. From top to bottom panel: the structure functions for the HARPS-N and SDO/HMI ΔRVs, the S-index, the CCF FWHM, CCF BIS, and the mean longitudinal magnetic field. The higher cadence of the HARPS-N data is visible in the structure functions as the smaller minimum time-scale. Likewise, the diurnal cycle of the ground-based observations gives rise to a gap in the structure function at ~ 0.5 d, since there are no pairs of observations separated by this time-scale. The solar rotation period at 27 d is indicated by a grey-dashed line.

of Cycle 24 versus 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 d (and a second smaller one at ~55 d). These dips highlight the strong modulation of the time series at the solar rotation period. The SDO/HMI-derived RVs 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 calculated using thousands of spectral lines and are sensitive to variability caused by additional physical processes on the Sun beyond those directly linked to magnetic activity, whereas the RVs 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 time-scales. They all present dips at ~1.5 and 2.5 d, which are not replicated in their RVs. The lack of a plateau region in any of the HARPS-N activity proxies means that no characteristic time-scale of the variability can be retrieved: the activity indicators are affected by multiple physical processes all with different time-scales. On the other hand, B_l shows a characteristic time-scale 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.

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 observations of the two time series in order to investigate the relationship between B_l and the RVs from HARPS-N. To do this, we interpolate the SDO/HMI data on to the timestamp of the nearest HARPS-N observation if the time between the two is less than 1 h. If the time between an SDO/HMI observation and its closest match in the HARPS-N data set is more than 1 h, the data point is omitted. A justification for this approach is provided in Appendix B. The resulting time series has 2891 data points and includes the diurnal cycle and realistic poor-weather breaks from the HARPS-N data, and the maximum of six observations per 24-h period of the selected data set for SDO/HMI.

3.4 Correlation analysis

We 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, as shown in Fig. 8. As derived previously in Section 3.1, B_l does not correlate with either of the RVs. It is interesting to note, however, that correlations between the HARPS-N activity indicators and the RVs derived with the same instrument are lower than (or in the case of the BIS, comparable to) their correlation to the SDO/HMI RVs. As mentioned previously, the SDO/HMI ΔRVs are only sensitive to rotationally modulated active region-induced RV variations, while the HARPS-N RVs are additionally influenced by all other physical processes on the solar surface as well as instrumental systematics. These results highlight why these indicators are not able to successfully map the RV signals imprinted by processes such as granulation or supergranulation. We also replicate the same averaging window study undertaken in Section 3.1, as shown in the bottom panel of Fig. 5. We 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 1 d and 1 yr. The results are plotted as a purple solid
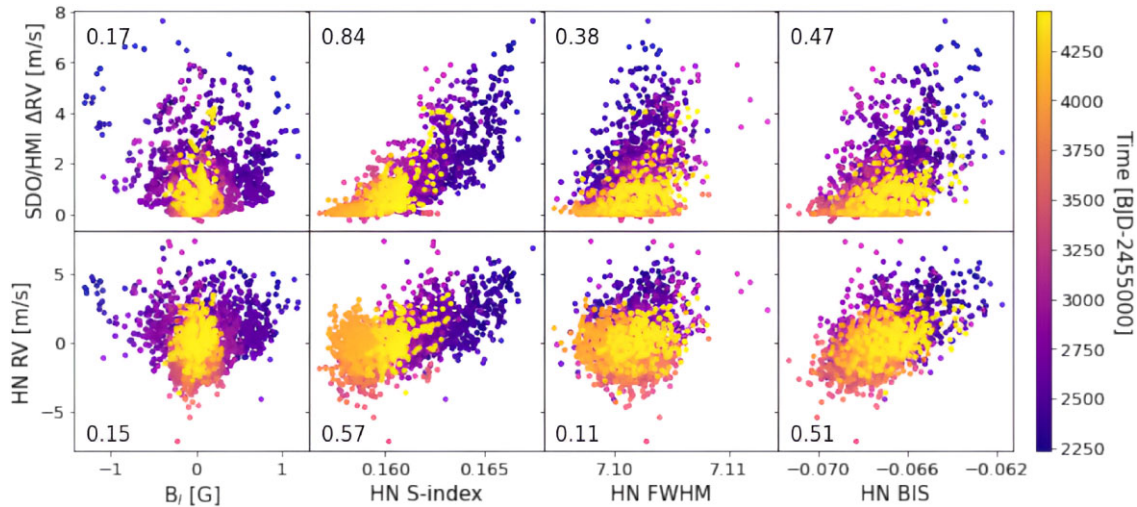
Figure 8. Correlation plots between the matched HARPS-N (here HN) and SDO/HMI time series. The SDO/HMI ΔRVs are in the first row, while the HARPS-N RVs are in the second. From the leftmost to the rightmost column, we plot the mean longitudinal magnetic field, the S-index, the FWHM, and the BIS. The data are colour-coded based on observation time and the Spearman rank correlation coefficients for each set are also included.

line. With a window size of roughly the solar rotation period, the two time series reach a strong correlation of 0.83, reconfirming our earlier conclusion. For comparison, we 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 $\sim$27 d. 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, we 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. 6, we 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 matches 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 2458500, 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.

3.5 Periodogram analysis

We compute the Generalized Lomb–Scargle periodograms (GLS: Zechmeister & Kürster 2009) of all the time series for the complete and the matched data sets. Both produce similar results. For this analysis, we focus on signals with periods smaller than 100 d, as longer magnetic cycle periodicity would not be reliably picked up with the available baseline. In particular, we are interested in assessing the ability of B_l to systematically recover the solar rotation period. For these reasons, we remove all long-term signals with a low-pass filter. Via this comparison, we are 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, we only include the periodograms for the matched data sets and plot them in Fig. 9. In the figure, the Carrington solar rotational period of $P_{rot} = 27.2753$ d as seen from the Earth is highlighted as a reference with a grey dash–dotted line. We also include the first and second harmonic of the Carrington period as dotted lines. By quick visual 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 & Levitskii 1983; Grigor'ev & Demidov 1987; Obridko & Shelting 1992) and is similar to the results obtained by Xie, Shi & Xu (2017) via wavelet transformation.

Both RVs are slightly more sensitive to half rotational period than the full one, although they present wide forests of peaks at P_{rot} and $P_{rot}/2$ both. The S-index has its most significant peak around 100 d, followed by one at $\sim$29 d. Even considering all peaks above the 0.1 per cent 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 $\sim$29 d. It also shows peaks at $\sim P_{rot}/2$, as well as $\sim$19 d. The periodogram of the BIS is nicely peaked around $P_{rot}/2$, with some signal around P_{rot}, as well as $\sim$32 and 22 d. The periodograms of all time series excluding B_l are complex at high frequencies, and have power at longer periods. Finally, the true solar P_{rot} cannot be recovered to a reasonable level of accuracy or precision from this analysis. 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, we should not expect any power at low frequencies. 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 RVs 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
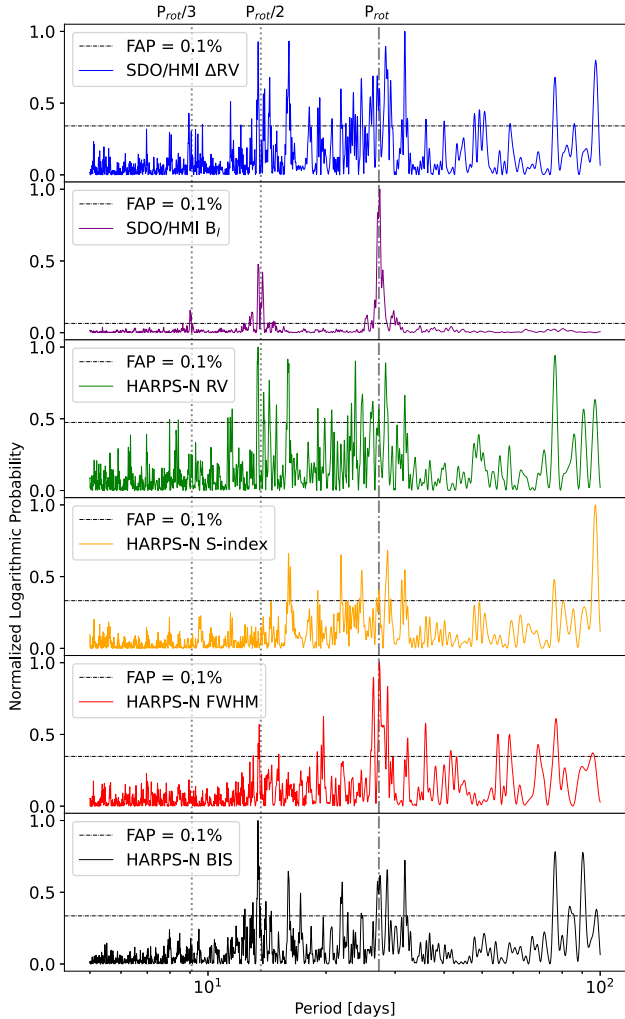
Figure 9. Generalized Lomb–Scargle Periodogram of the matched time series. On the *x*-axis the period in days, on the *y*-axis the normalized logarithmic Zechmeister–Kürster power (or probability). From the top to bottom panels, the matched time series of SDO/HMI ΔRVs, and mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS. The Carrington Solar rotation period is indicated by a grey dash–dotted line. Its half- and third-period harmonics are also highlighted by dotted lines. The false alarm probability (FAP) equal to 0.1 per cent are included as dashed grey horizontal lines.

value of B_l stays constant in time. The magnetic cycle only affects the amplitude of the oscillations, not their mid-point, meaning that they cannot be fit by a long-period sine curve. This effect yields a much simpler periodogram. We can overall conclude that a Fourier analysis of B_l is significantly more sensitive to P_{rot} and allows for a much more precise and accurate identification of the solar rotation period than all other analysed time series.

3.6 Autocorrelation function analysis

Another way of isolating the rotational period is to compute the autocorrelation function (ACF) of the time series (Giles, Collier Cameron & Haywood 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. We use the method developed by Edelson & Krolik (1988) and
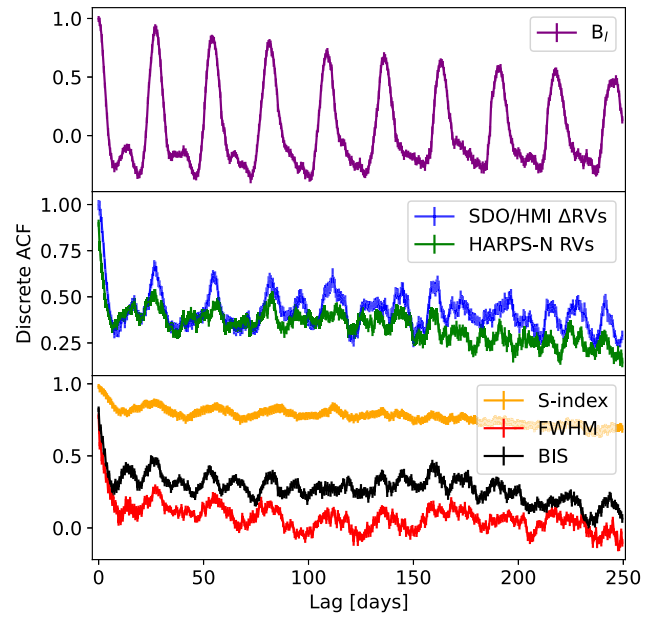
Figure 10. ACF over a lag window of 250 d of the mean longitudinal magnetic field (top row in purple), the SDO/HMI and HARPS-N RVs (middle row in blue and green, respectively), and the HARPS-N activity proxies S-index, FWHM and BIS (bottom row in orange, red, and black). Uncertainties are included as error bars.

updated in Robertson et al. (2015) to compute the ACF for unevenly sampled data sets. 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. Our code is adapted from PYDCG.[2] Assuming significant rotational modulation, the solar rotation period can be extracted as the lag between each major peak in the ACF. We obtain the ACF for all matched time series, as shown in Fig. 10. As in Section 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. We compute the half-life of the autocorrelation to be 2.74±0.02 d. While not 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 than the HARPS-N ones, as they are derived with a model that considers only rotationally modulated components.

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, Mortier et al., in preparation). We therefore also investigate the presence of any significant lag between the mean longitudinal magnetic field and both RV time series. We 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. We limit our investigation to lag values between −30 and 30 d. The results of this analysis are plotted in the top panel of Fig. 11. We first examine the CCF between B_l and the total SDO/HMI ΔRVs, plotted in dark blue. When considering the entirety of the available 13 yr of data, no significant lag can be

---

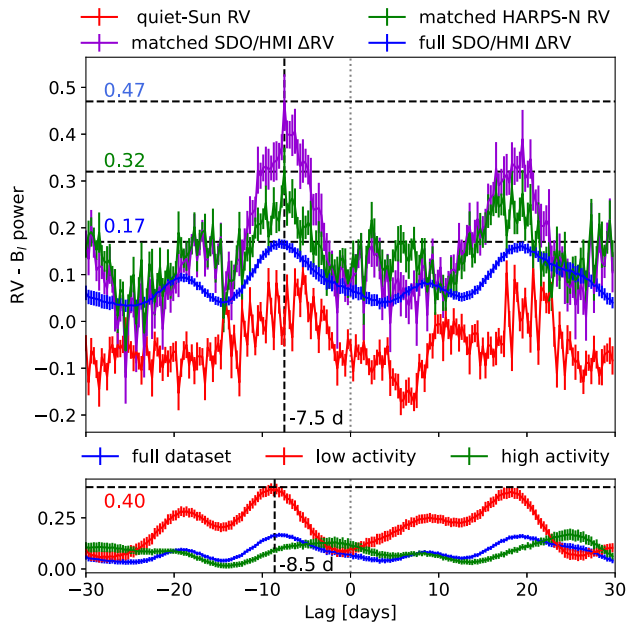[2]Available at: https://github.com/astronomerdamo/pydcf

Figure 11. Top panel: lag plot between B_l and the RV time series. The lag against the matched SDO/HMI ΔRVs is plotted in pale blue, while the lag against the HARPS-N matched RVs is plotted in green. We include the lag between B_l and the full SDO/HMI ΔRVs time series in dark blue. In red, we plot the lag between the matched B_l and the 'quiet-Sun' RVs, 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 error bars. The best correlation achieved and the best-fitting lag are highlighted by black-dashed lines. The 0 lag is highlighted with a grey 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 data set, in red the low-activity section of the same RVs (2015 December–2021 January), in green the high-activity section of the RVs (2010 May–2015 November). The best-fitting lag of the low-activity RVs and its respective correlation are highlighted with black-dashed lines. The 0 d lag is identified by a vertical dotted grey line.

found. A best correlation of 0.17 is recorded at ~ -7 d a next step we compute the CCFs between time-matched data sets. Starting from the matched SDO/HMI ΔRVs, we find a lag of -7.5 ± 0.5 d 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 d with a correlation coefficient of 0.32±0.05. We also note second possible peaks for all RV time series at ~20 d. We interpret this as the repetition of the same lag in the 'next' rotational period (assuming a $P_{rot} \sim 27$ d). The derived best lag is comparable to a fourth of the solar rotation, or roughly the difference between disc centre and limb. We 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, we 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, we also plot in red and green respectively in the bottom panel of Fig. 11 the CCF between B_l and the SDO/HMI ΔRVs during high activity (when the active region filling factors are maximized) 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 d with a 0.40±0.05 peak correlation coefficient. The cross-correlation also peaks at ~20 d, but differently from before, it also presents somewhat significant peaks at ~9 d (and the related rotation peak at ~ −19 d). 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 our 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 solar minimum. We 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. 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 does not reach the necessary correlation threshold of 0.5 to be considered significant. In fact, we are able to prove that the best-fitting lag is not constant in time. To confirm this, we divide the SDO/HMI data in rolling 200 d 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 to 5 d. Roughly 80 per cent of best-fitting lags have correlation coefficients under 0.4. We retrieve no significant trends with time. The −7.5 d result is only recovered from the distributions of all the best-fitting lags when considering only results with correlation coefficient above 0.4.

Looking back to Fig. 3, 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 & Wang (2015) show that during most observed magnetic cycles the solar large-scale field undergoes sudden rejuvenation only after the Sun has reached its maximum. They in fact state that a significant increase in 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 variability in the RVs (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
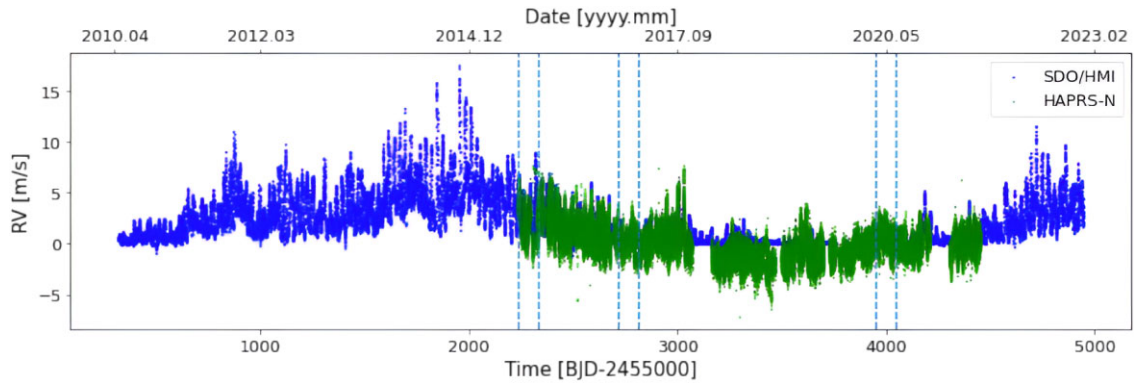
Figure 12. The selected stellar season-like 100 d chunks for the analysis in Section 4 are shown by vertical-dashed lines. The SDO/HMI RVs are plotted in the darker blue, and the HARPS-N RVs in the lighter green. Uncertainties are included but may be too small to be visible.

(Huang, Lin & Lee 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 & Robbrecht (2011) find 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 RVs, the mean value of B_l and its evolution with time inform 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 & Kochukhov 2002).

4 STELLAR-LIKE OBSERVATIONS: CAN WE USE B_l TO MEASURE P_{rot}?

In Section 3, we have demonstrated that with high cadence and long baseline, the mean longitudinal magnetic field is an effective 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 time-scales. 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 data sets? 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? We therefore test the mean longitudinal magnetic field as a 'stellar' activity tracer. We limit our data to the average stellar season length, 100 d. We 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 available in the HARPS-N data set.

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, we pass through a second data selection process. A typical cadence for stars is maximum of an observation a night. We therefore select only one observation taken each day of data. We do not average all data points 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. We instead randomly select one observation over each 24 h window. In this analysis, we do not account for the difference of integration time per exposure. The 5 min 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. We select three 100-d chunks over three stages of stellar activity. A high stellar activity case is selected for BJD 2457235 to 2457335 (2015 July 31 to 2015 November 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 2457716 to 2457816 (2016 November 23 to 2017 March 3). A low stellar activity case is selected during the extended minimum Cycle 24 for BJD 2458950 to 2459050 (2020 April 10 to 2020 July 19). The selected chunks are shown in Fig. 12 with vertical-dashed lines. From here onward, only the observed HARPS-N RVs will be considered in the analysis. As a reminder, given the matching method summarized in Section 3.3, bad-weather breaks are already included. We note that this already significantly reduced data set 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 described in Section 2.1 will eliminate data taken under not ideal conditions, these requirements are still more relaxed than what would be expected of a EPRV target. Moreover, we 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, we have reduced the data set considerably to a cadence similar to what new missions with dedicated Earth-like targets such as the Terra Hunting Experiment (THE: Thompson et al. 2016) are aiming to achieve.

As mentioned in Section 2.2.1, the uncertainties on the mean longitudinal magnetic field derived directly from SDO/HMI errors
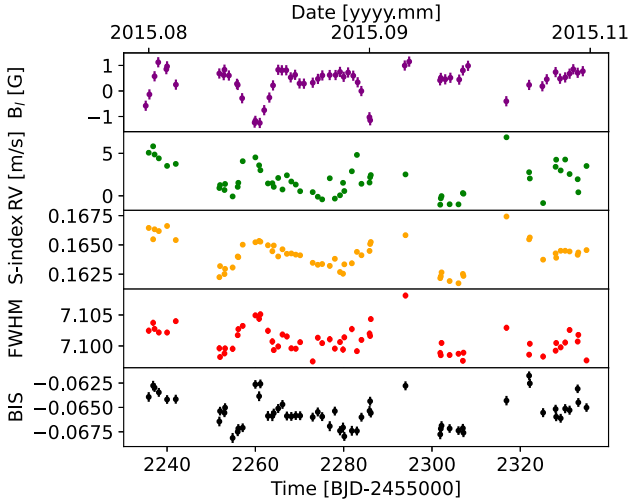
Figure 13. Correlation plots of the (from the top to bottom panels) high-, medium-, and low-activity data selection of the HARPS-N (here HN) RVs against the considered activity proxies. The data are colour-coded based on date. The Spearman rank-order correlation coefficient of each pair is also included.
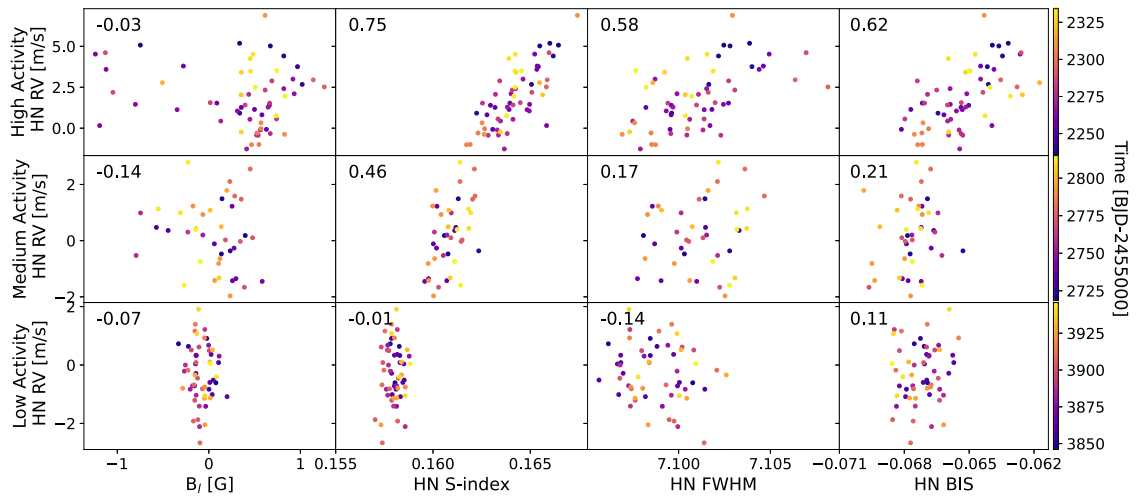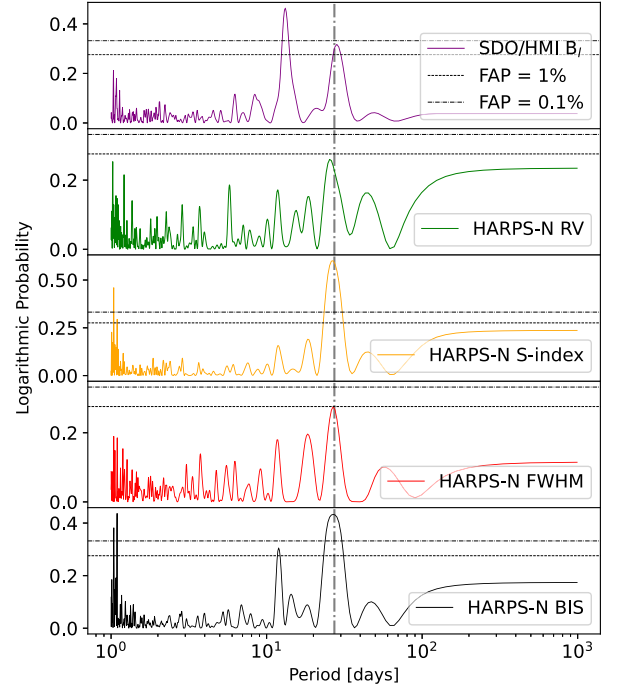

Figure 15. GLS periodograms of the high-activity data. From the top to bottom panels: mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS. 1 and 0.1 per cent FAPs are shown as dotted and dashed grey lines. The vertical dash–dotted black line highlights the Carrington solar rotational period.

4.2 Preliminary analysis

High stellar activity case: we start with the chronologically first 100-d chunk: the high stellar activity case. Over this window, we match 69 data points. The selected data are plotted in Fig. 14. Even with a quick visual inspection, a clear periodic signal can be identified in the mean longitudinal magnetic field. Similarly to Section 3, we compute the correlation between each considered activity indicator and the HARPS-N RVs, shown in the first row of Fig. 13. At this stage, the solar activity is strong and dominated by rotationally modulated
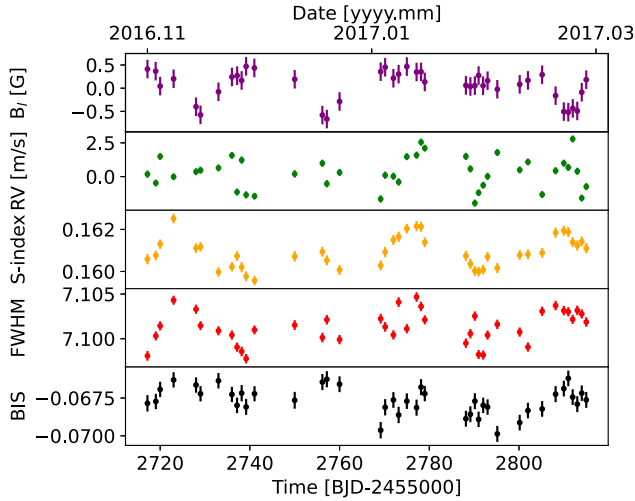
are very small. No existing or planned polarimetric survey of far-away stars could reach those levels. To better represent the stellar case, we instead inflate the uncertainties of B_l to the best achieved uncertainty level of 0.2 G on fully detected mean longitudinal fields for Sun-like star by the 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. We therefore use a constant error on B_l measurements of 0.2 G. To match this uncertainty and to truly represent the inflated error, we also inject into the data set white noise randomly extracted from a Gaussian distribution with a σ of 0.2 G.

Figure 14. High-activity data set selection. From the top to bottom panels: mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS. Some uncertainties may be too small to be clearly visible.

Figure 16. Medium activity data set selection. From the top to bottom panels: mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS. Some uncertainties may be too small to be clearly visible.

Figure 17. GLS periodograms of the medium activity data. From the top to bottom panels: mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS. 1 and 0.1 per cent FAPs are shown as dotted and dashed grey lines. The vertical dash–dotted black line highlights the Carrington solar rotational period.

effects, as shown by the high correlation between the RVs and the indicators S-index, FWHM and BIS. We then compute the ACF of each of the considered time series, as shown in the Appendix in Fig. C1. In this case, the magnetic field ACF does show a clear peak at the solar rotation period, and is the only time series for which a period can be systematically retrieved. Of the activity proxies, only the S-index includes hints to the ~ 27 d period, but the peaks are too wide for a proper rotation period analysis. We also plot the GLS periodograms of the all the time series for this chunk in Fig. 15. In this case, as expected, all HARPS-N proxies and to a lower degree the RVs themselves have power at the solar rotation period. Once again the B_l is sensitive to $P_{\rm rot}$ and $P_{\rm rot}/2$ signals.

Medium stellar activity case: A similar analysis is then undertaken for the medium activity case. Over this window we match 47 data points. We plot the derived time series in Fig. 16. The correlation relationships between the RVs and the activity indicators are plotted in the second row of Fig. 13. The computed Spearman rank correlation coefficient are now significantly lower for all proxies and no correlation above 0.5 can now be found. We can postulate that most of the rotationally modulated effects are now reduced in significance with respect to other photospheric and chromospheric variability. As before, we compute the ACF, plotted in the Appendix in Fig. C2. While some information regarding the solar rotation period could be extracted from the ACF of B_l, at this stage of activity no systematic period extraction can be applied to any of the considered time series. We also perform a GLS periodogram analysis, as shown in Fig. 17. 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 FAP of 1 per cent. The S-index and B_l are the only ones that present power at the rotational period over the 0.1 per cent FAP level. It is of note that all HARPS-N activity indicators are now also presenting a peak at ~ 40 d (not an harmonic of the rotational period or one of its aliases). This peak exceeds the 0.1 per cent 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.
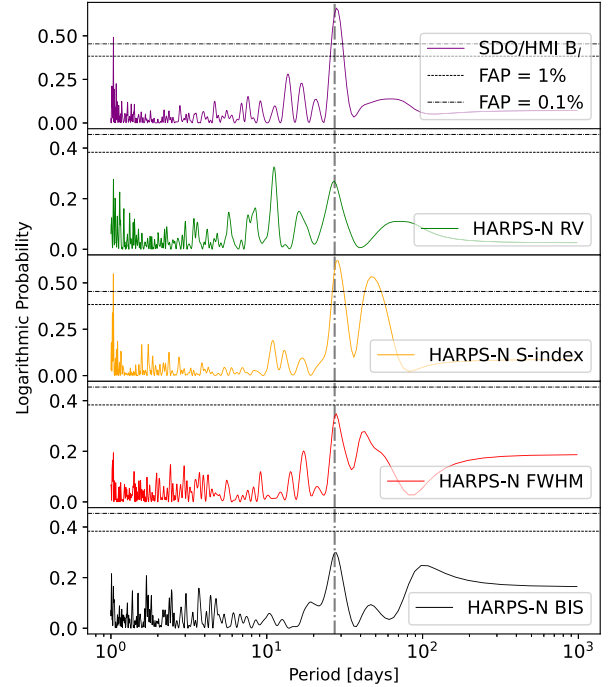
Low stellar activity case: We repeat the same analysis once more with the last selected data set over the extended solar minimum. We match 79 data points. The extracted time series are plotted in Fig. 18. At this level of activity the considered uncertainty of 0.2 G is comparable to the B_l signal itself. There is now no correlation between any of the activity proxies and the RVs, as shown in the bottom row of Fig. 13. Most of the rotationally modulated effects are now overshadowed by other sources of activity. Both an ACF (included in the Appendix in Fig. C3) and a periodogram analyses yield no information regarding the periodicity of the Sun. Most of the signal seems to in fact be aperiodic, as illustrated in Fig. 19.

4.3 Gaussian process regression analysis

The usual next step in the analysis of RV 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. This is done in order to identify the hyperparameters that better fit the stellar signal, which can then be used to inform priors in a second GP analysis of the RVs themselves. In this work, we undertake the most uninformative Gaussian process regression analysis in order to simulate a preliminary analysis, or a 'worse case scenario', in a typical exoplanet detection. To model the stellar activity, we use a quasi-period (QP) kernel (Haywood et al. 2014), with an added white noise 'jitter' term in the form:

$$k(t_i, t_j) = A^2 \cdot \exp\left[-\frac{|t_i - t_j|^2}{\tau^2} - \frac{\sin^2\left(\frac{\pi \cdot |t_i - t_j|}{P_{\rm rot}}\right)}{\mu^2} \right] + \delta_{i,j}\beta^2, \quad (6)$$

where $k(t_i, t_j)$ is the covariance function between times t_i and t_j, A is the amplitude of the signal, τ is the time-scale over which
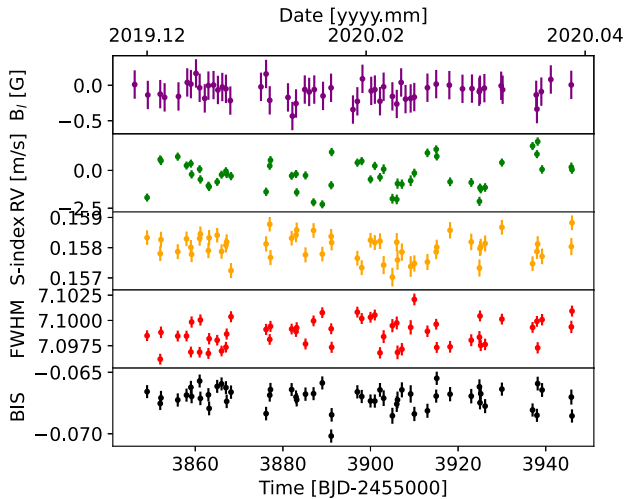
Figure 18. Low-activity data set selection. From top to bottom panels: mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS. Some uncertainties may be too small to be clearly visible.
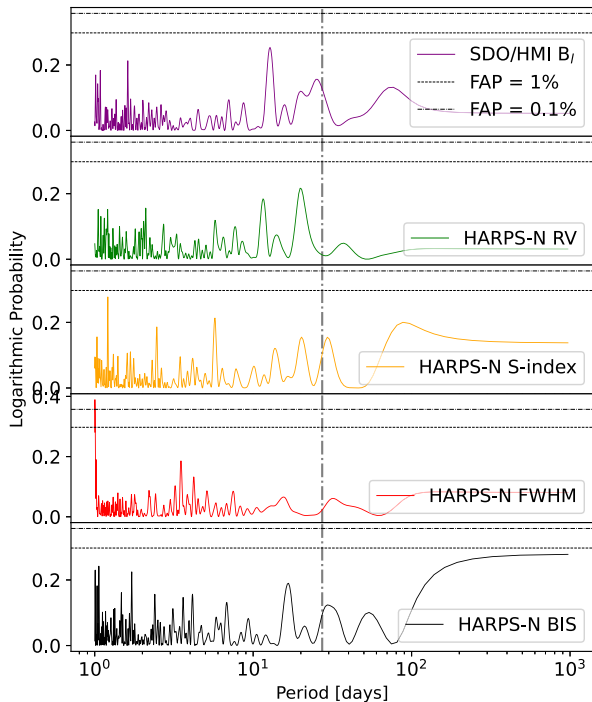

Figure 19. GLS periodograms of the low-activity data. From top to bottom panels: mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS. 1 and 0.1 per cent FAPs are shown as dotted and dashed grey lines. The vertical dash–dotted black line highlights the Carrington solar rotational period.

the quasi-periodicity evolves (related to the evolution time-scale of active regions), P_{rot} is the period of the stellar rotation, and μ is the harmonic complexity of the fit. β is a jitter term and is modelling the white noise contribution to the data derived from their inherent precision. It is only applied to the diagonal of the matrix (via the Kronoker Delta function $\delta_{i,j}$).

The QP kernel has been successfully employed to model stellar activity in both RV (e.g. Rajpaul et al. 2015; Barros et al. 2020) and
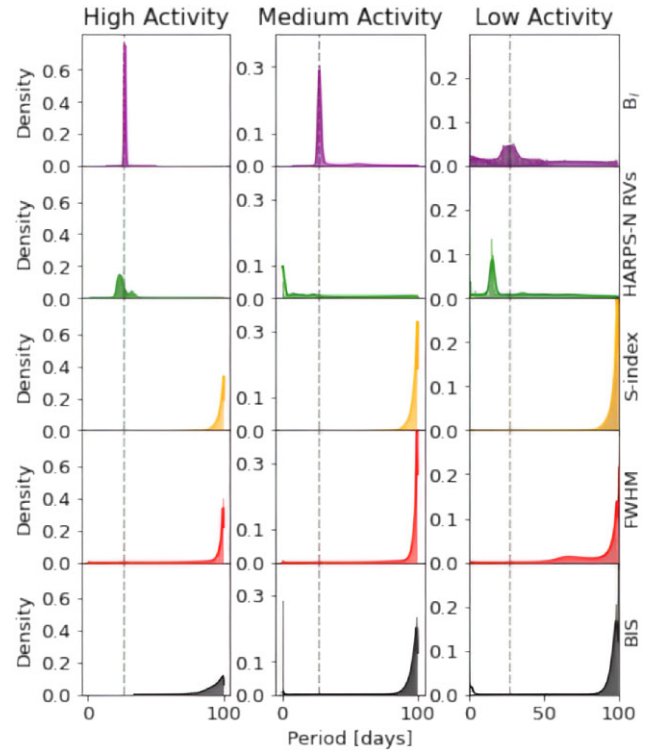
Figure 20. Collection of posteriors for the period P_{rot} of the quasi-periodic (QP) kernel after GP regression. From the left to right panels, we consider the high-, medium-, and low-activity cases. From the top to bottom panels, we see the posteriors of mean longitudinal magnetic field, HARPS-N RVs, 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.

stellar activity proxy analyses (e.g. Haywood et al. 2014; Grunblatt, Howard & Haywood 2015; Dalal et al. 2024). In this work, we test whether a similar analysis could be undertaken with B_l and whether it could be more successful than the same analysis on other activity proxies. To do so we use MAGPY_RV[3] (Rescigno, Dixon & Haywood 2023; Rescigno et al. 2024). MAGPY_RV is a GP regression pipeline with affine invariant Markov chain Monte Carlo (MCMC) parameter searching algorithm. We run the same analysis for all time series: B_l, HARPS-N RVs, S-index, FWHM, and BIS. 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]. We also bind both the period P_{rot} and the evolution time-scale with uniform priors between [0, 100] given the length of the selected window. We bind the white noise β with a Gaussian prior centred in the mean value of the uncertainties of the considered data set and of width equal to 25 per cent 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 we are able to derive from initial analysis in the low-activity case. For ease of comparison, we use the same priors in all runs. For all analyses, we simultaneously evolve 200 chains over 50 000 iterations each, with a discarded burn-in phase of 10 000 steps. We 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

[3]Available at: https://magpy-rv.readthedocs.io/en/latest/

full convergence, we select an investigation length over which most of the parameter space for all data sets is investigated, and over which all B_l chains are fully converged. We note that no detrending has been done to remove the magnetic cycle signal from any of the data sets. While we are aware of the presence of an overall descending trend in the RVs and some of the HARPS-N activity proxies due to the decline of the magnetic activity, this effect is not especially visible in the shortened 100-d baseline. This method is meant to reproduce the results of the same uninformed process of a first analysis of stellar data. We therefore do not want to introduce any pre-whitening due to information that are not directly derived from the selected data sets.

In this work, we focus on the hyperparameters useful for subsequent RV fitting: the period of the solar activity signal, its evolution time-scale and its harmonic complexity. The amplitudes and jitters of each time series are not comparable. We plot the posterior distributions after MCMC analysis of each considered hyperparameter in Figs 20–22. 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: We start our assessment from the period of the solar activity, in Fig. 20. 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 was 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 the half-period. 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 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 d. They model the activity in the data as a long period (longer than the data set) with shorter evolution time-scale τ and higher μ. They therefore 'assign' more of the signal to other time-dependent hyperparameter. This is another confirmation of the sensitivity of B_l to the solar rotational period. RVs 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 fore-shortening, but 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 time-scale: similarly to the period, we plot the posterior distributions of the evolution time-scale τ in Fig. 21. All common activity proxies as well as the mean longitudinal magnetic field prefer longer evolution time-scales than the radial velocity. The τ posterior distributions of the RVs peak at values comparable to the rotation period. This result is also supported by previous analysis (e.g. Camacho, Faria & Viana 2022). The longer time-scale recovered by B_l does not oppose the result of the RV analysis. In fact, even if we assume that the main source of the variations in B_l is the magnetic
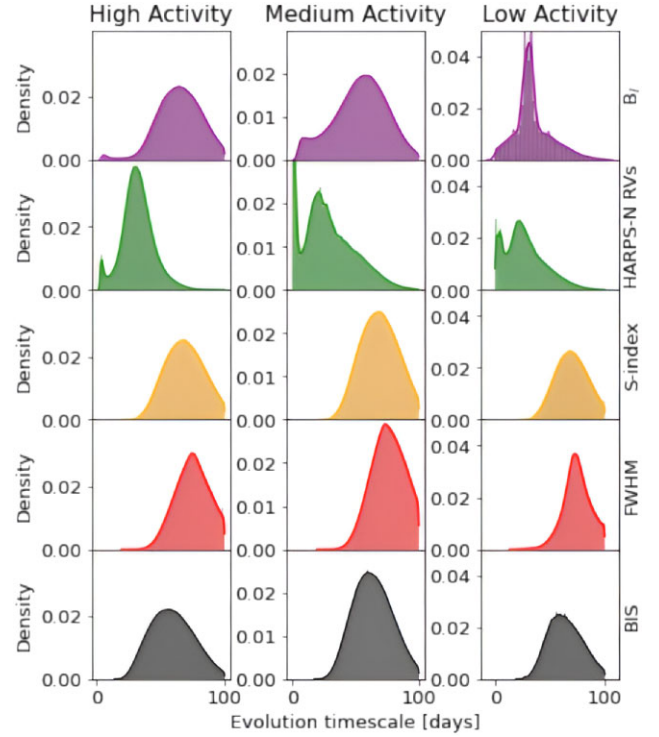

Figure 21. Collection of posteriors for the evolution time-scale after τ of the QP kernel GP regression. From left to right panels, we consider the high-, medium-, and low-activity cases. From the top to bottom panels, we see the posteriors of mean longitudinal magnetic field, HARPS-N RVs, S-index, FWHM, and BIS in their respective colours. Note the shared y-axis for each column.

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 & Zwaan 1999). After 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 & Howard 1965), and convective collapse begins. It is then reasonable for the S-index also to converge to comparable evolution time-scales to B_l. Opposite magnetic polarities separate after 24 h and areas of the same polarity migrate towards each other to coalesce into larger features such as pores (van Driel-Gesztelyi & 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 of 30 to 60 d) photometric active regions gradually disappear. In this process opposite polarity fragments magnetically reconnect and the flux slowly cancel itself. By the time coronal heating decreases, the plage (the coronal counterpart of faculae) 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 region is proportional to the magnetic flux it reaches at maximum development (van Driel-Gesztelyi & Green 2015). The results of τ in B_l can therefore be reliably employed as upper bounds in following RV GP regression analyses.

The harmonic complexity: When looking at the posterior distributions for the harmonic complexity μ, shown in Fig. 22, we
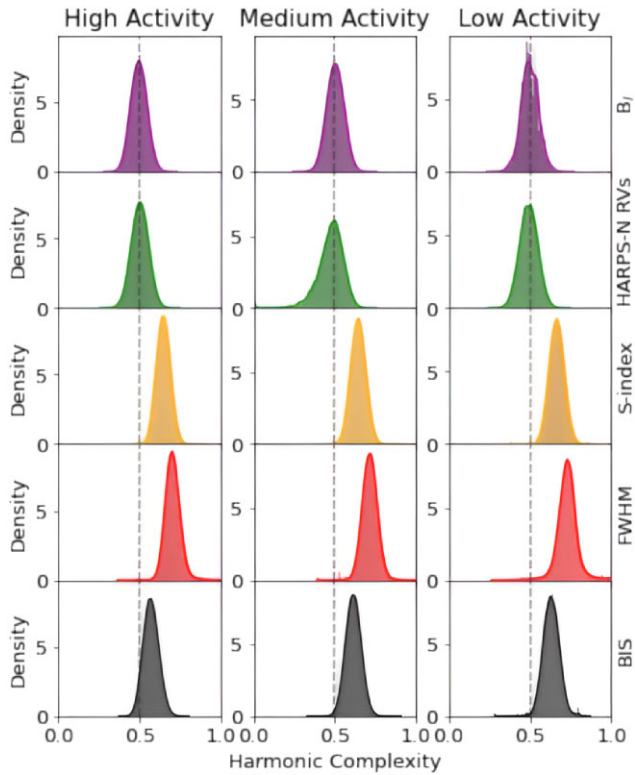
Figure 22. Collection of posteriors for the harmonic complexity μ of the QP kernel after GP regression. From the left to right panels, we consider the high-, medium- and low-activity cases. From the top to bottom panels, we see the posteriors of mean longitudinal magnetic field, HARPS-N RVs, S-index, 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.

note how 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. This μ value yields a covariance that prefers GP models with one extra 'bump' per period. In a very simplistic view, this could be physically equated to two active regions on opposite sides of the solar sphere with respect to the observer rotating in and out of view, or to an active region distribution that produces a similar signal. This result is in line with the conclusions in Jeffers & Keller (2009). However, we caution that the harmonic complexity is the only hyperparameter of the QP kernel that cannot be reliably and systematically tied to a specific physical property (Nicholson & Aigrain 2022). From the formulation of the QP kernel in equation (6), 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 3, that the mean longitudinal magnetic field exhibits less complex signals than the other proxies. However first, the posterior distributions in Fig. 22 need to be considered within the larger context of the Gaussian Process regression. We cannot do a direct comparison of the extracted best-fitting 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 note that for all levels of activity, the best-fitting jitter term extracted by the GP after MCMC optimization 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. We 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, we 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 considered 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 QP 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 we have proven that they are in agreement over all activity levels. Moreover, the evolution time-scale derived for B_l can inform the upper bound of the same hyperparamter for the RVs.

5 CONCLUSIONS

In this work, we 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. We considered the longitudinal magnetic field extracted from SDO/HMI observations alongside the ΔRVs derived with a model from the same data. We performed correlation analysis, we computed their structure functions, GLS periodograms and ACFs, and tested for the presence of any time lag between the two time series. In parallel, we duplicated all our analyses with Sun-as-a-star observations taken by the HARPS-N spectrograph (we considered the HARPS-N derived RVs, the S-index, and the FWHM and the BIS of the CCF). We find the following:

(i) 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 RVs. We however find that the RMS of B_l computed over a window comparable to the solar rotation period do 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, we are able to reduce the RV scatter by more than 60 per cent. B_l can therefore be used to successfully model out the long-term RV signal due to the magnetic cycle of the Sun.

(ii) B_l has a significantly simpler structure function than all other considered time series, with a characteristic time-scale of $\sim$10 d.

(iii) 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 RVs are as simple, and in most cases no clear rotation period can be isolated. This point is further confirmed by the autocorrelation analysis, in which the rotational period signal of B_l stays clear and strong over multiple rotations.

(iv) A lag analysis was performed and a minorly relevant lag between B_l and the RVs was found at circa -7.5 d. 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, we therefore have shown 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 our understanding of the physical processes happening on

the surface of the Sun. This is, however, not representative of the type of observational time series we have for exoplanet detection. Therefore, we also tested the B_l as a 'stellar activity tracer'. We reduced the time series to 100-d windows, with a single observation per night, and inflated the uncertainties in B_l to those achieved by previous polarimetric surveys. We then performed a typical preliminary analysis followed by a Gaussian process regression with a QP kernel. We performed the same analysis for three chunks of data over high, medium and low-activity levels. We found the following:

(i) The mean longitudinal magnetic field outperforms the other activity indicators in a preliminary periodogram analysis for the medium- and low-activity levels.

(ii) After one-dimensional GP regression using a QP kernel, B_l is the only time series (compared against HARPS-N RVs, S-index, FWHM, and BIS) 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.

(iii) B_l (as well as all other proxies) recovers a longer evolution time-scale than the RVs.

(iv) The best-fitting harmonic complexities of B_l and the RVs strongly agree within uncertainties.

With this analysis, we have reconfirmed 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-fitting 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, in more fields than simply exoplanetology.

DATA AVAILABILITY

The observational data included in this publication are all either currently or soon-to-be public. The SDO/HMI images are available at https://sdo.gsfc.nasa.gov/data/aiahmi/. The first half of the HARPS-N solar data is available at https://dace.unige.ch/observationSearch/?observationType=[%22solarSpectroscopy%22]. The second half is in the process of release and will be described and made available in an upcoming publication.

REFERENCES

Aigrain S. et al., 2015, MNRAS, 450, 3211
Al Moulla K., Dumusque X., Figueira P., Lo Curto G., Santos N. C., Wildi F., 2023, A&A, 669, A39
Amard L., Roquette J., Matt S. P., 2020, MNRAS, 499, 3481
Babcock H. W., 1949, ApJ, 110, 126
Bagnulo S., Landstreet J. D., 2015, Stellar magnetic fields. Cambridge Univ. Press, Cambridge, p. 224
Barros S. C. C., Demangeon O., Díaz R. F., Cabrera J., Santos N. C., Faria J. P., Pereira F., 2020, A&A, 634, A75
Blunt S. et al., 2023, AJ, 166, 62
Boisse I., Bouchy F., Hébrard G., Bonfils X., Santos N., Vauclair S., 2011, A&A, 528, A4
Bortle A., Fausey H., Ji J., Dodson-Robinson S., Ramirez Delgado V., Gizis J., 2021, AJ, 161, 230
Bose S., Nagaraju K., 2018, ApJ, 862, 35
Brown S. F., Donati J. F., Rees D. E., Semel M., 1991, A&A, 250, 463
Bruning D. H., 1991, Publ. Astron. Soc. Pac., 103, 368
Bumba V., Howard R., 1965, ApJ, 141, 1492
Camacho J. D., Faria J. P., Viana P. T. P., 2022, MNRAS, 519, 5439
Cheung M. C. M., van Driel-Gesztelyi L., Martínez Pillet V., Thompson M. J., 2017, Space Sci. Rev., 210, 317
Collier Cameron A. et al., 2019, MNRAS, 487, 1082
Cosentino R. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham. p. 84461V
Cosentino R. et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham. p. 91478C
Costes J. C. et al., 2021, MNRAS, 505, 830
Couvidat S. et al., 2016, Sol. Phys., 291, 1887
Crass J. et al., 2021, preprint (arXiv:2107.14291)
Cretignier M., Pietrow A. G. M., Aigrain S., 2024, MNRAS, 527, 2940
Dalal S. et al., 2024, MNRAS, 531, 4464
Donati J. F., Collier Cameron A., 1997, MNRAS, 291, 1
Donati J. F., Landstreet J. D., 2009, ARA&A, 47, 333
Donati J. F. et al., 2020, MNARS, 498, 5684
Donati J. F. et al., 2023, MNRAS, 525, 2015
Dumusque X., Udry S., Lovis C., Santos N. C., Monteiro M. J. P. F. G., 2011, A&A, 525, A140
Dumusque X. et al., 2015, ApJ, 814, L21
Dumusque X. et al., 2021, A&A, 648, A103
Edelson R. A., Krolik J. H., 1988, ApJ, 333, 646
Ervin T. et al., 2022, AJ, 163, 272

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

Fischer D. A. et al., 2016, Publ. Astron. Soc. Pac., 128, 066001

Fouqué P. et al., 2023, A&A, 672, A52

Gaudi B. S. et al., 2020, preprint (arXiv:2001.06683)

Giles H. A. C., Collier Cameron A., Haywood R. D., 2017, MNRAS, 472, 1618

Grigor'ev V. M., Demidov M. L., 1987, Sol. Phys., 114, 147

Grunblatt S. K., Howard A. W., Haywood R. D., 2015, ApJ, 808, 127

Harada C. K., Dressing C. D., Kane S. R., Ardestani B. A., 2024, ApJS, 272, 30

Hathaway D. H., 2015, Liv. Rev. Sol. Phys., 12, 4

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

Haywood R. D. et al., 2016, MNRAS, 457, 3637

Haywood R. D. et al., 2022, ApJ, 935, 6

Hébrard É. M., Donati J. F., Delfosse X., Morin J., Moutou C., Boisse I., 2016, MNRAS, 461, 1465

Hoeksema J. T. et al., 2014, Sol. Phys., 289, 3483

Huang G. H., Lin C. H., Lee L. C., 2017, Sci. Rep., 7, 9488

Irving Z. A., Saar S. H., Wargelin B. J., do Nascimento J.-D., 2023, ApJ, 949, 51

Jeffers S. V., Keller C. U., 2009, in Stempels E., ed., AIP Conf. Proc. Ser. Vol. 1094, 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun. Am. Inst. Phys., New York. p. 664

Klein B. et al., 2021, MNRAS, 502, 188

Kochukhov O., Hackman T., Lehtinen J. J., 2023, A&A, 680, L17

Kotov V. A., Levitskii L. S., 1983, Izvestiya Ordena Trudovogo Krasnogo Znameni Astrofizicheskoj Observatorii, 68, 56

Kutsenko A. S., Abramenko V. I., Yurchyshyn V. B., 2017, Sol. Phys., 292, 121

Lakeland B. S., Naylor T., 2022, MNRAS, 514, 2736

Lakeland B. S. et al., 2024, MNRAS, 527, 7681

Landstreet J. D., 1982, ApJ, 258, 639

Landstreet J. D., 1992, A&A, 4, 35

Lienhard F., Mortier A., Cegla H. M., Cameron A. C., Klein B., Watson C. A., 2023, MNRAS, 522, 5862

Lindegren L., Dravins D., 2003, A&A, 401, 1185

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

Marsden S. C. et al., 2014, MNRAS, 444, 3517

Marsden S. C. et al., 2023, MNRAS, 522, 792

Mengel M. W. et al., 2017, MNRAS, 465, 2734

Meunier N., 2021, preprint (arXiv:2104.06072)

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

Milbourne T. W. et al., 2019, ApJ, 874, 107

Morbidelli A., Baillié K., Batygin K., Charnoz S., Guillot T., Rubie D. C., Kleine T., 2022, Nat. Astron., 6, 72

Nava C., López-Morales M., Haywood R. D., Giles H. A. C., 2020, AJ, 159, 23

Nava C. et al., 2022, AJ, 163, 41

Nicholson B. A., Aigrain S., 2022, MNRAS, 515, 5251

Nicholson B. A., Hussain G., Donati J. F., Wright D., Folsom C. P., Wittenmyer R., Okumura J., Carter B. D., 2021, MNRAS, 504, 2461

Norton A., Howe R., Upton L., Usoskin I., 2023, Space Sci. Rev., 219, 64

Obridko V. N., Shelting B. D., 1992, Sol. Phys., 137, 167

Oláh K. et al., 2009, A&A, 501, 703

Pesnell W. D., Thompson B. J., Chamberlin P. C., 2012, Sol. Phys., 275, 3

Petit P. et al., 2008, MNRAS, 388, 80

Phillips D. F. et al., 2016, in Navarro R., Burge J. H., eds, Proc. SPIE Conf. Ser. Vol. 9912, Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II. SPIE, Bellingham. p. 99126Z

Piskunov N., Kochukhov O., 2002, A&A, 381, 736

Quanz S. P. et al., 2022, A&A, 664, A21

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

Reiners A., 2012, Liv. Rev. Sol. Phys., 9, 1

Rescigno F., Dixon B., Haywood R. D., 2023, Astrophysics Source Code Library, record ascl:2310.006

Rescigno F. et al., 2024, MNRAS, 527, 5385

Robertson D. R. S., Gallo L. C., Zoghbi A., Fabian A. C., 2015, MNRAS, 453, 3455

Saar S. H., Donahue R. A., 1997, ApJ, 485, 319

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

Scherrer P. H., Wilcox J. M., Howard R., 1972, Sol. Phys., 22, 418

Scherrer P. H. et al., 2012, Sol. Phys., 275, 207

Schou J. et al., 2012, Sol. Phys., 275, 229

Semel M., 1989, A&A, 225, 456

Sergison D. J., Naylor T., Littlefair S. P., Bell C. P. M., Williams C. D. H., 2020, MNRAS, 491, 5035

Serrano L. M., Barros S. C. C., Oshagh M., Santos N. C., Faria J. P., Demangeon O., Sousa S. G., Lendl M., 2018, A&A, 611, A8

Severny A. B., 1971, Q. J. R. Astron. Soc., 12, 363

Sheeley N. R. J., Wang Y. M., 2015, ApJ, 809, 113

Simonetti J. H., Cordes J. M., Heeschen D. S., 1985, ApJ, 296, 46

Strous L. H., Zwaan C., 1999, ApJ, 527, 435

The LUVOIR Team, 2019, preprint (arXiv:1912.06219)

Thompson S. J. et al., 2016, Proc. SPIE, 9908, 99086F

Trippe S., 2014, J. Korean Astron. Soc., 47, 15

van Driel-Gesztelyi L., Green L. M., 2015, Liv. Rev. Sol. Phys., 12, 1

Wang Y. M., Robbrecht E., 2011, ApJ, 736, 136

Xiang N. B., Qu Z. N., 2016, AJ, 151, 76

Xie J. L., Shi X. J., Xu J. C., 2017, AJ, 153, 171

Yeo K. L., Solanki S. K., Krivova N. A., 2013, A&A, 550, A95

Yu L. et al., 2019, MNRAS, 489, 5556

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

APPENDIX A: CORRELATION PLOTS

We include in this appendix the correlation plots between the HARPS-N Sun-as-a-star solar RVs and the activity proxies considered in the paper: the spectra-derived S-index, and the CCF-derived FWHM and BIS. The correlations are plotted in Fig. A1. The data are colour-coded with increasing BJD. In each panel, we also include the computed Spearman correlation coefficient rank. More information can be found in Section 3.1.
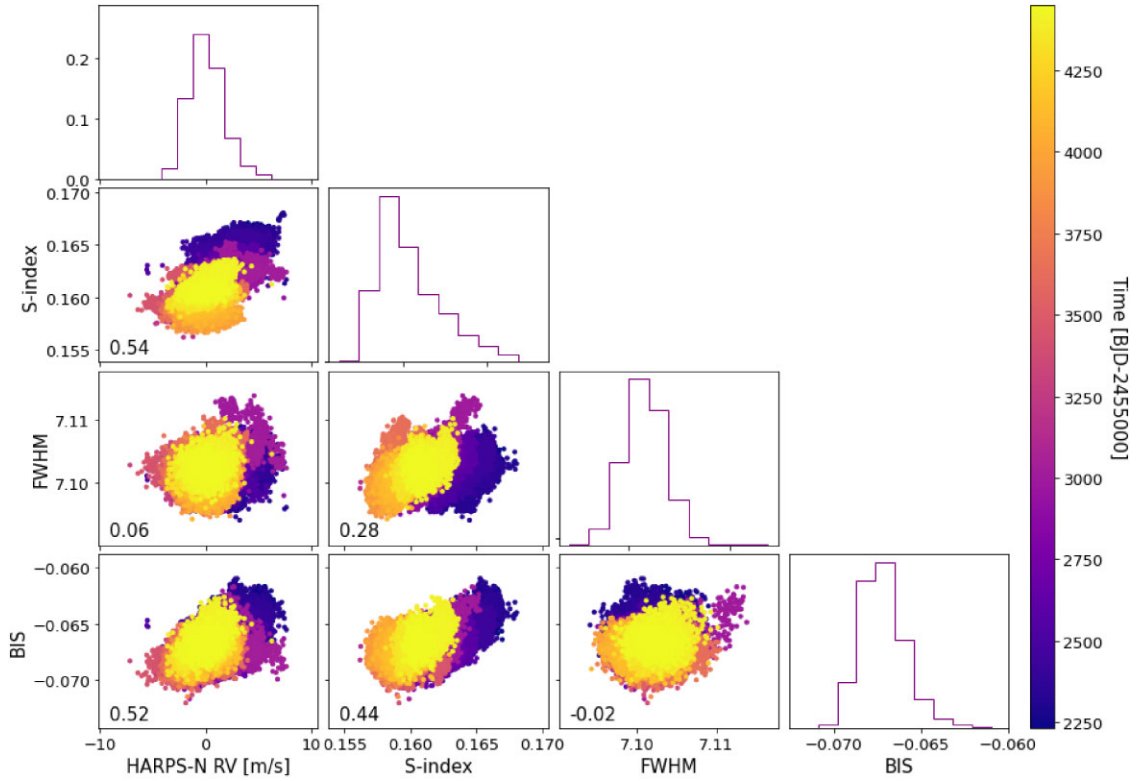
Figure A1. Correlation plot between the HARPS-N RVs and their activity proxies, S-index, FWHM, and BIS. The colour bar indicates the BJD of each data point. The Spearman Rank correlation factor for each set is also included.

APPENDIX B: A JUSTIFICATION FOR THE DATA SET MATCHING METHOD

There are in principle a number of approaches to the problem of combining time series with unequal sampling. The approach we take is to linearly interpolate the SDO/HMI data on to the timestamp of the closest HARPS-N observation, provided that these two observations are separated by no more than 1 h. If there are no HARPS-N observations within 1 h of a given SDO/HMI observation, that data point is omitted. In contrast to interpolating the SDO/HMI data on to every HARPS-N timestamp, each SDO/HMI observation is being considered at most once. This avoids spurious correlations in Section 3 where a single SDO/HMI data point corresponds to many HARPS-N observations.

It may appear counter intuitive to interpolate the sparser data set on to the data set with the denser sampling. To justify this approach, we produced a high-cadence time series from every available set of 720 s-exposure SDO/HMI images from 2017 January 1 to 2017 February 2017. This month was chosen as it represents roughly the median activity level we explore. In Fig. B1, we plot the structure functions of both the HARPS-N RVs and the high-cadence mean longitudinal magnetic field.[4] The steeper slope of the structure function of the mean longitudinal magnetic field demonstrates that there is significantly less fractional variability within the longitudinal
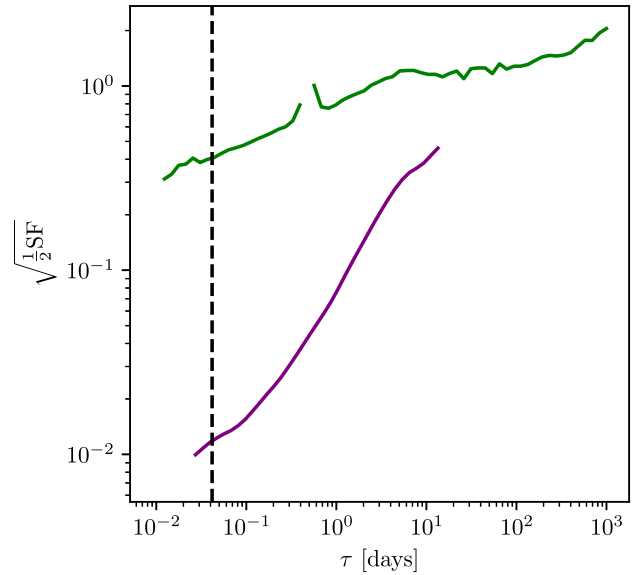
[4]See the main text for more details

Figure B1. Structure functions of the HARPS-N RVs (green) and mean longitudinal magnetic field (purple). Note that the longitudinal magnetic field used here is a higher cadence time series than the one used in the main text. A vertical dashed line indicates 1 h, the maximum interpolation distance in the matching procedure used here. The greater steepness of the purple line indicates that interpolating the SDO/HMI data introduces less spurious signal than interpolating the HARPS-N data.

magnetic field time series at time-scales of less than an hour than there is in the HARPS-N RVs. We are therefore justified in interpolating the sparser time series on to the denser grid, as opposed to the inverse. To further assess the strength of any possibly injected signals via interpolation, we also compute the mean absolute deviation (MAD) of the difference between all consecutive SDO/HMI observations. The derived MADs are 0.074 m s^{-1} for the SDO/HMI ΔRVs and 0.015 G for B_1. These values represent the average dispersion of the separation between each subsequent observations and are also small enough to be negligible for the scope of this work.

APPENDIX C: AUTOCORRELATION FUNCTION ANALYSIS OF THE STELLAR-LIKE OBSERVATIONS

We include in this appendix the ACF plots of all the considered time series in Section 4 for the three selected levels of activity: the matched longitudinal magnetic field, HARPS-N RVs, S-index, FWHM and BIS. All the ACFs are plotted in Figs C1, C2 and C3 for the high-, medium-, and low-activity data sections, respectively . In each panel, we also include the solar rotation period and its second pulse at $\sim$27 and $\sim$54 d, respectively (the Carrington period and twice that value), as grey vertical dashed lines. The only ACF that can be used to reliably inform us about the solar rotation period is the ACF of the high-activity time series of B_1, but is all other cases the peaks due to the rotation are either too wide or not present at all.
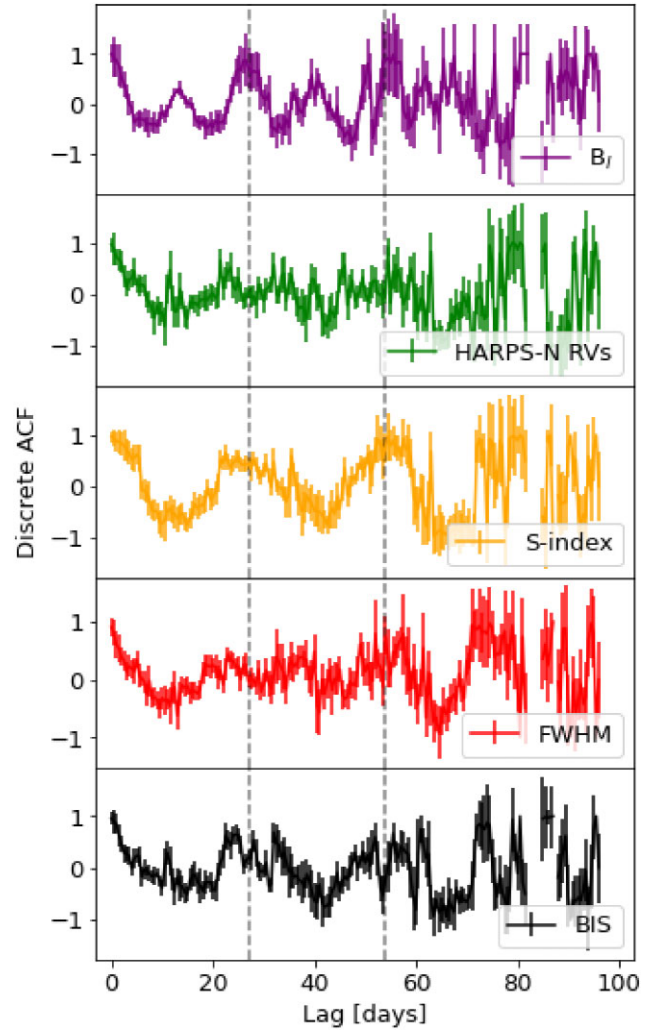

Figure C1. ACF of the matched time series of the high-activity section of from top to bottom panels: longitudinal magnetic field in purple, the HARPS-N RVs in green, the HARPS-N S-index in orange, the HARPS-N FWHM in red and the HARPS-N BIS in black. Uncertainties are included in the form of error bars. The vertical dashed lines in grey represent (from the left to right) the approximated solar rotation period ($\sim$27 d) and twice that value ($\sim$54 d).
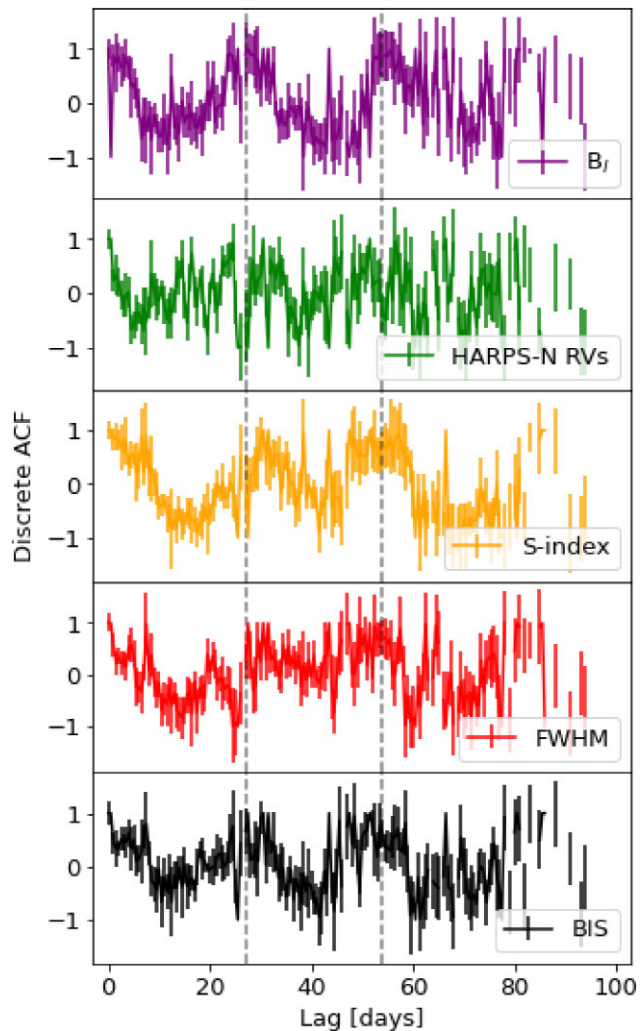
Figure C2. ACF of the matched time series of the medium-activity section of from the top to bottom panels: longitudinal magnetic field in purple, the HARPS-N RVs in green, the HARPS-N S-index in orange, the HARPS-N FWHM in red, and the HARPS-N BIS in black. Uncertainties are included in the form of error bars. The vertical dashed lines in grey represent (from the left to right) the approximated solar rotation period ($\sim$27 d) and twice that value ($\sim$54 d).
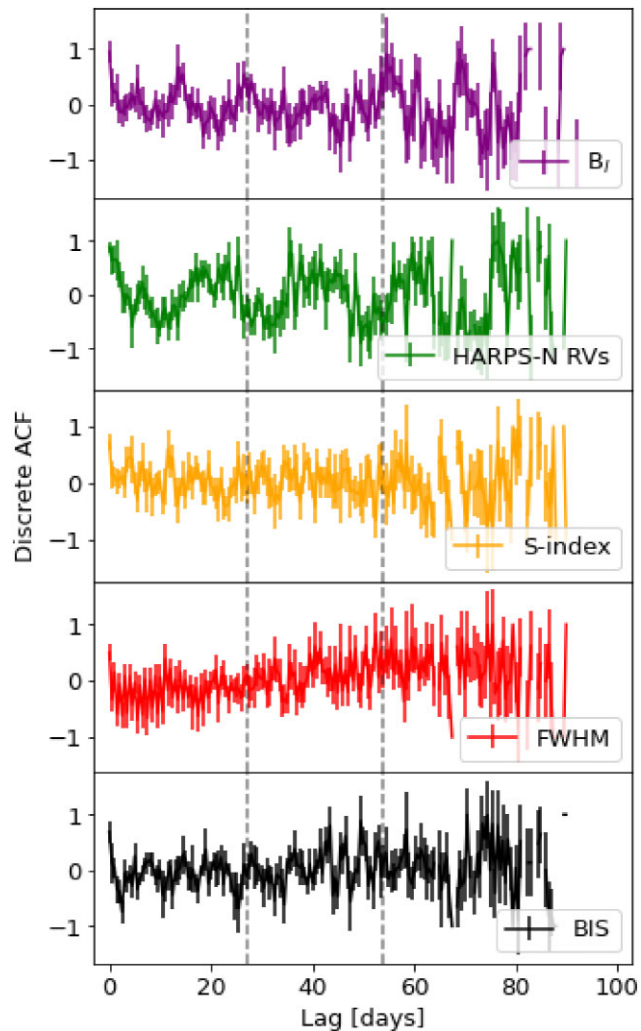
Figure C3. ACF of the matched time series of the low-activity section of from the top to bottom panels: longitudinal magnetic field in purple, the HARPS-N RVs in green, the HARPS-N S-index in orange, the HARPS-N FWHM in red, and the HARPS-N BIS in black. Uncertainties are included in the form of error bars. The vertical dashed lines in grey represent (from the left to right) the approximated solar rotation period ($\sim$27 d) and twice that value ($\sim$54 d).

This paper has been typeset from a TEX/LATEX file prepared by the author.