

Unsigned magnetic flux proxy from solar optical intensity spectra

F. Lienhard ¹★, A. Mortier ^{1,2,3}, H. M. Cegla,⁴ A. Collier Cameron ⁵, B. Klein ⁶ and C. A. Watson⁷

¹*Astrophysics Group, Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge CB3 0HE, UK*

²*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

³*School of Physics & Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*

⁴*Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK*

⁵*SUPA School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*

⁶*Department of Physics, University of Oxford, OX13RH Oxford, UK*

⁷*Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, BT7 1NN Belfast, UK*

Accepted 2023 April 28. Received 2023 April 28; in original form 2022 October 21

ABSTRACT

The photospheric unsigned magnetic flux has been shown to be highly correlated with radial velocity (RV) variations caused by solar surface activity. This activity indicator is therefore a prime candidate to unlock the potential of RV surveys to discover Earth twins orbiting Sun-like stars. We show for the first time how a precise proxy of the unsigned magnetic flux ($\Delta\alpha B^2$) can be obtained from Sun-as-a-star intensity spectra by harnessing the magnetic information contained in over 4000 absorption lines in the wavelength range from 380 to 690 nm. This novel activity proxy can thus be obtained from the same spectra from which RVs are routinely extracted. We derived $\Delta\alpha B^2$ from 500 randomly selected spectra from the HARPS-N public solar data set, which spans from 2015 to 2018. We compared our estimates with the unsigned magnetic flux values from the Solar Dynamics Observatory (SDO) finding excellent agreement (median absolute deviation: 4.9 per cent). The extracted indicator $\Delta\alpha B^2$ correlates with SDO's unsigned magnetic flux estimates on the solar rotational time-scale (Pearson correlation coefficient 0.67) and on the 3-yr time-scale of our data set (correlation coefficient 0.91). We find correlations of $\Delta\alpha B^2$ with the HARPS-N solar RV variations of 0.49 on the rotational time-scale and 0.78 on the 3-yr time-scale. The Pearson correlation of $\Delta\alpha B^2$ with the RVs is found to be greater than the correlation of the classical activity indicators with the RVs. For solar-type stars, $\Delta\alpha B^2$ therefore represents the best simultaneous activity proxy known to date.

Key words: line: profiles – techniques: radial velocities – techniques: spectroscopic – planets and satellites: detection – stars: magnetic field.

1 INTRODUCTION

A planet causes the radial velocity (RV) of its host star to change periodically over time. Yet, Doppler-like signals caused by the star itself, linked to the interplay between the evolving magnetic field and stellar surface convection, can drown out and mimic planetary signals. These manifestations of stellar magnetic activity represent a major obstacle to detecting planetary-induced RVs below 1 m s^{-1} (see Crass et al. 2021), with only very few measurements below this threshold (e.g. Faria et al. 2022). To date, Earth-like planets orbiting solar-type stars in the habitable zone are out of reach, as they produce RV signals with semi-amplitudes of the order of 10 cm s^{-1} . It is therefore essential to disentangle planetary and stellar RV components to obtain a clean planetary RV curve.

Stellar activity subsumes a range of phenomena including stellar magnetic cycles (Lanza 2010; Costes et al. 2021), star-spots (Saar & Donahue 1997; Desort et al. 2007; Lagrange, Desort & Meunier 2010), faculae and plagues (Saar & Donahue 1997; Saar 2003, 2009; Meunier, Desort & Lagrange 2010a; Meunier, Lagrange & Desort 2010b), meridional flows (Meunier & Lagrange 2020), granulation

(Dravins 1982; Dumusque et al. 2011; Meunier et al. 2015; Cegla et al. 2019), super-granulation (Rieutord et al. 2010; Rincon & Rieutord 2018; Meunier & Lagrange 2019), and p -mode oscillations (Mayor et al. 2003; Medina et al. 2018; Yu et al. 2018; Chaplin et al. 2019). These phenomena act on different time-scales and have different impacts on the RVs. An effect of particular importance is the suppression of convective blueshift (Meunier et al. 2010a). For solar-type stars, the emission emanating from convective upflows dominates over the downflows and leads to a net blueshift of the stellar spectrum. However, this effect is modulated by the magnetic field inhibiting stellar surface convection (e.g. Hanslmeier, Nesis & Mattig 1991). Since the magnetic field is spatially inhomogeneous, regions with suppressed convection rotate in and out of view as the star rotates, leading to a varying Doppler shift and variations in the shape of the absorption lines. In addition, the magnetic field evolves in time, and thus the overall effect also varies in time beyond the rotational time-scale.

The hemispherically averaged unsigned magnetic flux $|\hat{B}_{\text{obs}}|$ has been shown experimentally to be an excellent proxy for variations in solar RV (Haywood et al. 2016, 2022). This finding is supported by analyses of Dopplergrams and magnetograms from the Michelson Doppler Imager (Scherrer et al. 1995) presented in Meunier et al. (2010b). They showed that suppression of convective blueshift

* E-mail: f1386@cam.ac.uk

is pronounced where the magnetic field is strong. In addition, simulations by Meunier et al. (2010a) showed that the attenuation of the convective blueshift is indeed the dominant contributor to star-induced RV variations. The analyses in Meunier et al. (2010a) indicate that the attenuation of the convective blueshift leads to a long-term RV signal with an amplitude of about 8 m s^{-1} and thus impedes the detection of Earth twins orbiting solar-type stars. The photometrically induced RV variations due to bright active regions and dark star-spots rotating in and out of view, on the other hand, partially cancel out and are of lesser concern.

The strength and evolution of stellar magnetic fields are challenging to measure, though. The earliest measurements of the solar magnetic field date back to 1908, with Hale (1908) observing Zeeman splitting (Zeeman 1897) in sunspot spectra. Most of the methods that exist to date are either only applicable to highly active stars, require polarimetric data, measurements at infrared wavelengths, or a combination of these (Saar & Linsky 1985; Valenti, Marcy & Basri 1995; Johns-Krull, Valenti & Koresko 1999; Reiners & Basri 2006). An overview of magnetic field estimation methods is provided in Sections 2.3 and 2.4.

In this study, we show how a proxy of $|\hat{B}_{\text{obs}}|$ can be derived from intensity spectra in the visible wavelength range of weakly active to moderately active stars. These stars have been, and continue to be, prime targets of RV surveys. This work represents an extension of, and builds on, the multi-mask least-squares deconvolution (MM-LSD) method¹ presented in Lienhard et al. (2022). They analysed the performance of least-squares deconvolution (LSD; Donati et al. 1997) as a tool to extract RV information from spectra of FGK-type stars and the dependence of the measured RV on various parameters. For this purpose, Lienhard et al. (2022) developed a pipeline that continuum-normalizes deblazed echelle order spectra, partially corrects for telluric absorption lines, masks problematic wavelength regions, and finally extracts the RV via LSD. The pre-processing steps implemented in this pipeline and the convolution approach are reused in the present analysis.

In Section 2, we explain the basics of Zeeman splitting, describe different magnetic field diagnostics based on this effect, as well as the in this context commonly used LSD method. In Section 3, we describe the data products used in this analysis. The proposed magnetic flux proxy is presented in Section 4. We describe the application on HARPS-N solar spectra in Section 5. Lastly, we show and discuss our results in Section 6, and conclude in Section 7.

2 THEORETICAL BACKGROUND

2.1 Least-squares deconvolution

The LSD method relies on physical information about the absorption lines, such as line depth and wavelength, to model the spectra at hand on the basis of few assumptions. The objective of using a simple model is to describe and model the bulk of the absorption lines, rather than to precisely model single lines. In this way, one can tease out line information from all absorption lines and extract information that is otherwise hidden in the noise. The LSD model for a spectrum is generated by convolving a common profile with a line list consisting of scaled delta functions at the rest wavelengths of the known absorption lines. The delta functions are scaled depending on what needs to be modelled. To model intensity spectra, for instance,

we scale the delta functions by the expected relative depths of the relevant spectral absorption lines (e.g. Lienhard et al. 2022).

By applying least-squares fitting, the best-fitting common profile can be determined. Similarly to the cross-correlation function (CCF; Baranne et al. 1996; Pepe et al. 2002), this common profile represents the average shape of the absorption lines. Analogously to the CCF, one can extract the stellar RV from this common profile. Other applications include modelling Stokes V spectra to estimate the stellar magnetic flux (e.g. Donati et al. 1997).

There are two main assumptions that LSD is based on. One is that absorption lines add up linearly. This assumption is valid for weak absorption lines only (Kochukhov, Makaganik & Piskunov 2010). Secondly, LSD defines the common profile in velocity space. It is thus assumed that the absorption lines have the same shape in velocity space and only scale by a wavelength-specific factor. This standard assumption in LSD (Donati et al. 1997; Kochukhov et al. 2010; Lienhard et al. 2022) is based on the fact that the conditions on the stellar surface are similar, all lines are rotationally broadened, and the dominant atomic absorbers for FGK-type stars, such as Fe, Ni, Cr, and Ti, all have similar atomic masses and hence similar thermal broadening. The width component due to thermal broadening is

$$\Delta\lambda_T = 2 \frac{\lambda_0}{c} \sqrt{\frac{2kT}{m} \ln(2)}, \quad (1)$$

where T is the plasma temperature, c the speed of light, k the Boltzmann constant, m the atomic mass, and λ_0 the rest wavelength of the absorption line (e.g. Bellot Rubio & Orozco Suárez 2019). Since this expression scales linearly with wavelength, the width contribution in the velocity domain remains constant.² Lastly, micro- and macro-turbulent broadening are the same for all species.

The LSD method can therefore be a useful tool if we find the same general shape at the wavelengths of the absorption lines and we can model this shape as a profile fixed in velocity space scaling with factors that depend on the lines' physical properties.

2.2 Zeeman effect

In this Section, we summarize the theoretical background of Zeeman splitting and the relevant equations describing its effect on absorption lines. A magnetic field splits an absorption line involving a magnetically sensitive state into multiple absorption lines at slightly offset wavelengths. This effect is called Zeeman splitting and is due to the external magnetic field splitting an initially degenerate energy level with angular momentum J into $2J + 1$ sublevels. The energy difference between the original degenerate state and the split components is proportional to the magnetic field strength and their respective magnetic quantum number m . Transitions involving these split states are therefore shifted in wavelength. For a review on Zeeman splitting in stellar spectra, see e.g. Reiners (2012), Stenflo (2013), and Bellot Rubio & Orozco Suárez (2019).

For electric dipole transitions, the selection rules allow for $\Delta m = 0$ and $\Delta m = \pm 1$. The former produce the unshifted central π component, while the latter produce the shifted σ components. A simple triplet of lines is created when J of one of the involved states is equal to 0 or if the Landé factors of both states are equal. More complicated patterns can still be treated as a triplet by computing an effective Landé factor g_{eff} for the transition. The polarization of the individual

²The conversion can be made with the standard approximation $\frac{\Delta\lambda}{\lambda} = \frac{\Delta v}{c}$ where $\Delta\lambda$ is the observed Doppler shift due to the source moving with velocity Δv relative to the observer.

¹Available on github: <https://github.com/florian-lienhard/MM-LSD>.

components depends on the direction of the magnetic field relative to the observer. Crucially, the σ components are circularly polarized with opposite orientation if the magnetic field vector points towards or away from the observer. Since the σ components are shifted in wavelength with the shift proportional to the magnetic field strength, this generates a characteristic signal in Stokes V and enables the estimation of the stellar magnetic flux from polarized spectra, as outlined in Section 2.3. For transverse fields, on the other hand, the three components are linearly polarized producing signals in Stokes Q and U (Kochukhov et al. 2011).

The wavelength difference between the π and the σ components is equal to

$$\Delta\lambda = 4.67 \times 10^{-10} \lambda_0^2 g_{\text{eff}} B, \quad (2)$$

which translates to a velocity shift of

$$\Delta v = 1.4 \times 10^{-4} \lambda_0 g_{\text{eff}} B, \quad (3)$$

with the rest-frame wavelength λ_0 in Å, g_{eff} the dimensionless effective Landé factor, the magnetic field strength B in kG, and the velocity shift in km s^{-1} . It follows that lines at longer wavelengths exhibit stronger Zeeman splitting. A field of 1 kG strength typically shifts an absorption line in the visible range by about 1 km s^{-1} . This is orders of magnitudes larger than the typical RV shift due to a planet. However, the Zeeman signal manifests very differently in the absorption lines, as it leads to a varying shift of a varying fraction of the absorption lines rather than a velocity shift affecting the entire spectrum uniformly. Furthermore, only a tiny surface fraction (less than a few per cent for the Sun, as can be derived from Milbourne et al. 2021; Haywood et al. 2022) of weakly active stars is affected by such strong fields. The Zeeman signal is therefore washed out and intermixed with weaker splitting patterns. Since the width of a typical absorption line is much greater than 1 km s^{-1} , Zeeman splitting generally leads to slightly broadened lines in the optical for FGK-type stars, rather than separated Zeeman triplets. For non-saturated lines Zeeman splitting does not alter the equivalent width. However, the splitting expands the saturation region of saturated absorption lines and thereby increases their equivalent widths. This effect is called Zeeman intensification (e.g. Saar, Piskunov & Tuominen 1992; Basri & Marcy 1994; Kochukhov et al. 2020).

In the following two sections, we describe how polarized and unpolarized stellar spectra are affected differently by Zeeman splitting and how this relates to the techniques used to characterize magnetic fields.

2.3 Polarimetric measurements

The 4th component of the Stokes vector, Stokes V , is defined as the difference between right and left-handed circular polarization (for a review see Stenflo 2013). The σ components are oppositely circularly polarized when the magnetic field vector is parallel to the line of sight, as mentioned in Section 2.2. This leads to a characteristic signal in Stokes V thus encoding the strength and orientation of the large-scale magnetic field. In the weak field regime, this information can be extracted from multiple lines simultaneously using LSD, as described in Donati et al. (1997).

The LSD method is used, for instance, to compute surface magnetic maps through Zeeman Doppler Imaging (ZDI; Donati, Semel & Praderie 1989; Semel 1989; Kochukhov & Wade 2016). ZDI capitalises on the fact that Stokes V signatures of active regions are blueshifted as they first appear on the visible stellar hemisphere and then progressively shift towards longer wavelengths as the star rotates. Given a series of observations at different times, surface

magnetic maps can be reconstructed by finding the maximum-entropy magnetic field geometry that produces the observed Stokes V time series (e.g. Skilling & Bryan 1984; Donati et al. 2006; Folsom et al. 2018).

Methods based on Stokes V permit the extraction of magnetic field diagnostics for rapidly rotating stars, but they have some inherent disadvantages. Mainly, Stokes V signals from adjacent stellar surface regions of opposite polarity can cancel out if the respective polarized components are not sufficiently separated in wavelength (e.g. Saar 1988; Donati et al. 1997; Reiners 2012). This leads to an underestimation of the magnetic field strength. Linear polarization signals are much weaker, can be significantly affected by magneto-optical effects (Landolfi & Landi Degl'Innocenti 1982), and are also subject to cancellation effects (Saar 1988; Reiners 2012). For instance, Kochukhov et al. (2011) find the linear polarization signal to be 10 times weaker than the circular polarization signal. Lastly, taking polarimetric measurements requires additional equipment, poses technical challenges, and generally uses up more observation time to collect the same number of photons as compared to Stokes I measurements.

2.4 Stokes I measurements

Extraction of magnetic field information from intensity spectra is fraught with technical complications but in principle has significant advantages over polarimetric methods. It is important to note that, conversely to Stokes V measurements, intensity spectra are not affected by cancellation effects due to regions of opposite polarity. Also, there are many high-resolution spectrographs designed for radial velocity studies on solar-type stars producing extensive time series of Stokes I spectra, but only few that provide polarimetric data. The capability to simultaneously measure the evolution of the instantaneous magnetic flux and the RV from intensity spectra is expected to lead to the discovery of smaller planets and improve mass measurements of known planets in the vast amount of existing and upcoming data.

Zeeman splitting measurements in the visible range are challenging since the line profile changes are very small in Sun-like stars and can be confused with other line broadening effects, such as thermal broadening (see e.g. Reiners 2012; Bellot Rubio & Orozco Suárez 2019). Since Zeeman splitting is proportional to the wavelength squared (see equation 2), many intensity based methods therefore focus on extracting information from one suitable line in the infrared at very high spectral resolution. For instance, a few studies are based on the Mg I line at $12.32 \mu\text{m}$ (e.g. Brault & Noyes 1983; Zirin & Popp 1989; Bruls & Solanki 1995). The infrared domain poses instrumental problems, however, and is riddled with telluric absorption lines that can lead to a higher RV error (Cunha et al. 2014). Water absorption lines are especially problematic as the precipitable water vapour content is spatially inhomogeneous and variable (e.g. Leet, Fischer & Valenti 2019; Cretignier et al. 2021). Furthermore, most of the spectrographs designed for RV studies record stellar intensity spectra in the visible wavelength range. For solar-type stars, this wavelength range is optimal because there is a large number of absorption lines and the SNR is highest as the stellar flux peaks in the visible. For example, HARPS (Mayor et al. 2003) and HARPS-N (Cosentino et al. 2012) measure spectra from 383 to 690 nm, EXPRES from 390 to 780 nm (Jurgenson et al. 2016), and ESPRESSO from 378.2 to 788.7 nm (Pepe et al. 2021). Other instruments, such as CARMENES (520–960 and 960–1710 nm, Quirrenbach et al. 2016) or NEID (380–930 nm, Halverson et al. 2016) also record parts of the near-infrared spectrum. However, these

instruments still do not reach the wavelength regime where cleanly split Zeeman diagnostic lines are found. Hence, there is a need for a magnetic flux proxy for intensity spectra in the visible wavelength range.

There are some studies describing Stokes I magnetic field fitting techniques. Stenflo & Lindegren (1977) fit 402 unblended unpolarized Fe I absorption lines in the visible wavelength range observed at the solar disc centre. They fit the line widths at different depths and were able to estimate an upper limit for the average magnetic flux of 110 G for the Sun. Kochukhov et al. (2020) have extracted magnetic flux estimates for Sun-like stars via Zeeman intensification. However, their error bars are much larger than the typical average magnetic flux variations of less than 1 G that are relevant for stellar activity mitigation. In this context, we present the first method able to produce sufficiently precise magnetic flux time series.

3 DATA

3.1 VALD3

To model the magnetic response of the absorption lines, we require the stellar absorption lines' wavelength, depth, and their effective Landé factor. This information can be retrieved from the Vienna Atomic Line Database (VALD3; Ryabchikova et al. 2015). Since we analyse solar spectra in this work, we set the stellar microturbulence to 1.1 km s^{-1} , the effective temperature (T_{eff}) to 5833 K, and surface gravity ($\log g$) to 4.44, and the chemical composition to solar values. These stellar parameters were estimated as outlined in Lienhard et al. (2022), section 2.3. Our estimate for the solar effective temperature is marginally higher than the recommended value of 5772 K (Prša et al. 2016). We kept the value that we derived from the HARPS-N spectra to keep the analysis consistent. We do not expect this to have any measurable impact on our results. We only included lines with relative depth greater than 0.2 to exclude very weak lines which are often affected by noise. Furthermore, we excluded all molecular absorption lines in the VALD3 list to have a more homogeneous set. This removes about 12 per cent of the lines in our list and very marginally improves our results. About 50 per cent of the remaining lines are due to Fe I.

3.2 HARPS-N

HARPS-N is a pressure and temperature-stabilized cross-dispersed echelle spectrograph operational since 2012. It has a resolving power of $R = 115\,000$ in the visible range from 383 to 690 nm over 69 spectral orders. In addition to the nightly observations, HARPS-N is outfitted with a solar telescope to record disc-integrated spectra of the Sun at 300-s cadence, and has been doing so for several hours on most days since 2015 (Cosentino et al. 2014; Dumusque et al. 2015; Phillips et al. 2016; Collier Cameron et al. 2019; Dumusque et al. 2021).

For this study, we randomly selected one spectrum from each observing day contained in the set of three years of high-quality HARPS-N solar observations presented in Dumusque et al. (2021).³ We note that HARPS-N had a cryostat leak requiring periodic interventions. As a result, the spectra within 5 d from an intervention were excluded from this data set because they can be affected by flux variations and are not representative of HARPS-N's usual performance.

In total, we selected 500 spectra. The exposure time was 300 s for each of them. The first spectrum was recorded on 2015 July 29, and the last spectrum on 2018 May 18. The airmass of the exposures ranges between 1 and 2.9, with the median being around 1.3. The minimum signal-to-noise ratio (SNR) at 550 nm is about 250, the maximum is around 460, and the median is equal to 380. These SNR values are very high compared to nightly HARPS-N observations with a typical SNR between 50 and 200. As we show in Section 6.2, our approach does not require a very high SNR.

For each spectrum, we extracted MM-LSD RVs as well as CCF RVs. Furthermore, the Data Reduction System (DRS) also computes several activity indices. For this study, we used the full width at half maximum (FWHM), contrast and bisector inverse slope (BIS) of the CCF. The RV, FWHM, and contrast values were corrected for effects of Solar system motions as detailed in Collier Cameron et al. (2019). Furthermore, the $\log R'_{HK}$ index was computed directly from the HARPS-N spectra, yielding values between -5.03 and -4.96 . This indicator quantifies the emission in the cores of the Ca II H (3968.47 Å) and K (3933.66 Å) spectral lines which is induced by magnetic activity. First, the S-index is computed standardly by weighing the emission within these bands with a triangular response function with width 1.09 Å and dividing by the reference bands with a width of 20 Å at 3900 and 4000 Å (Gomes da Silva et al. 2011; Dumusque et al. 2021). The emission within the line cores was susceptible to contamination due to effects related to the cryostat leak. The leak led to the build-up of humidity over time, increasing the reflectivity in the detector and producing local flux variations called ghosts. The impact of these ghosts on the extracted S-index is corrected as described in Dumusque et al. (2021). The S-index is then converted to $\log R'_{HK}$ following Noyes et al. (1984).

3.3 SDO

To validate our results, we compare with data from the Helioseismic and Magnetic Imager (HMI) instrument onboard the Solar Dynamics Observatory (SDO; Pesnell, Thompson & Chamberlin 2012; Schou et al. 2012; Couvidat et al. 2016). HMI measures the line-of-sight magnetic flux through the magnetically sensitive Fe I line at 6173.3 Å . We downloaded⁴ the 720 s magnetograms and intensitygrams using a 6-h cadence spanning our full HARPS-N time range. The absolute value of the SDO line-of-sight magnetic fluxes were intensity-weighted and summed over all pixels as outlined in Haywood et al. (2016) to compute $|\hat{B}_{\text{obs}}|$. To estimate the filling factors of active regions, we used the same thresholds as in Haywood et al. (2016) and Milbourne et al. (2019) to distinguish between faculae, sunspots, and quiet photosphere. More specifically, the magnetic field was assumed to be radial. Any pixel with foreshortening-corrected magnetic flux below 24 G was assumed to measure quiet photosphere. Pixels above this threshold were divided into sunspots and faculae using an intensity threshold of 0.89 times the mean pixel intensity corrected for limb-darkening as in Yeo, Solanki & Krivova (2013). The filling factor and unsigned magnetic flux time series can alternatively be obtained using SOLASTER, presented in Ervin et al. (2022).

There is a minor time difference between the HARPS-N observations and the SDO magnetic flux measurements. This time difference is smaller than 6 h (mean absolute difference: 2 h) for all our measurements. Since $|\hat{B}_{\text{obs}}|$ and the RVs only marginally evolve over this time-scale, the time difference only minimally influences our results. By interpolating the SDO data to the timestamps of the

³<https://dace.unige.ch/sun/>

⁴<http://jsoc.stanford.edu/ajax/exportdata.html>

HARPS-N spectra, we achieve correlations about 0.005 higher than those reported in this analysis.

3.4 Tapas

The Transmissions of the AtmosPHERE for AStronomical data database (TAPAS; Bertaux et al. 2014) provides transmission spectra for the Earth’s atmosphere. We use one arbitrarily chosen transmittance spectrum (La Palma Roque de los Muchachos Canarias Spain, 2018/3/30, 01:45:07, airmass 1.03) to identify and exclude spectral regions impacted by deep tellurics. More information is provided in Section 5.1 and in Lienhard et al. (2022).

4 EXTRACTING THE UNSIGNED MAGNETIC FLUX

In this Section, we first describe our model for the difference between a Zeeman-split and an unsplit absorption line and where this model is valid. We subsequently show how the difference between our spectra and a master spectrum can be fit using the LSD approach to extract a proxy for $|\hat{B}_{\text{obs}}|$.

4.1 Residual model

For simplicity, we assumed that the magnetic field strength distribution on the stellar surface can be captured by two components: the quiet surface and the active regions with magnetic field strength roughly three orders of magnitude higher. The exact values do not have to be fixed for the algorithm described below. The intensity profile of a Zeeman-split line can then be modelled as in Title & Tarbell (1975), Robinson (1980), and Marcy (1982):

$$I(\lambda) = \alpha (f(\lambda + \Delta\lambda) + f(\lambda - \Delta\lambda)) + \beta f(\lambda). \quad (4)$$

The parameter α captures the emission from the shifted σ components, β quantifies the emission from unsplit lines and the unshifted π component, $\Delta\lambda$ is the Zeeman shift proportional to the effective Landé factor and the magnetic field strength as in equation (2), and f describes the shape of the line. Both values α and β finally depend on the orientation of the magnetic field relative to the observer, as explained in Section 2.2. The model in equation (4) relies on the assumption that the σ and the π components have the same profile and do not interact. The validity of this assumption is based on the shifted σ components having opposite polarity. This greatly simplifies the radiative-transfer problem, as it allows diagonalizing the transfer matrix leading to non-interaction between the polarization components (e.g. Stenflo et al. 1984). Note also that we assumed that the magnetic field strength does not vary radially within the active regions. We, therefore, assume that all lines are exposed to the same magnetic field independent of their formation height.

Assuming that the lines’ equivalent widths remain the same, equation (4) can be simplified:

$$I(\lambda) = \alpha (f(\lambda + \Delta\lambda) + f(\lambda - \Delta\lambda)) + (1 - 2\alpha)f(\lambda). \quad (5)$$

In the following, we assume that the ratio of π to σ components for the active regions remains about the same. In this case, the parameter α in equation (5) is directly proportional to the filling factor of active regions on the visible stellar hemisphere.⁵ We note that the strength

⁵Assume the proportion of magnetic area: r_{active} , and non-magnetic area $1 - r_{\text{active}}$ and ratio r_s of σ components. Then: $I(\lambda) = r_{\text{active}}(1 - 2r_s)f(\lambda) +$

of the π to σ components for a given active region depends on the angle between the line of sight and the magnetic field vector (e.g. Seares 1913; Marcy 1982; Skumanich & López Ariste 2002). By the Seares’ relation, the intensity of one σ component at a given position on the stellar surface is equal to

$$I_{\sigma} = I_0 \frac{1 - \cos^2 \theta}{4}, \quad (6)$$

where I_0 is the total intensity of all three components and θ is the angle between the line-of-sight and the magnetic field vector. The strength of the π component is then:

$$I_{\pi} = I_0 \frac{\sin^2 \theta}{2}. \quad (7)$$

For disc-integrated spectra, Marcy (1982) assumed a radial field and estimated the average θ to 34° . Since we are mainly interested in the evolution of the magnetic flux, we do not need to estimate an average field-line to line-of-sight angle. However, we assume that the active regions are homogeneously distributed such that in the disc-integrated spectra the ratio remains about constant. This means that the factor capturing the disc-averaged value of $0.25(1 - \cos^2 \theta)$ is absorbed in α .

Lastly, we assume that the line profiles are all Gaussian. With few exceptions, this is a valid assumption for optical absorption lines of main-sequence FGK-type stars given our resolution and precision. The intensity profile of an absorption line with line depth d , central wavelength μ , and width σ_{λ} is then

$$I_{\text{non-magnetic}}(\lambda) = de^{-\frac{(\lambda-\mu)^2}{2\sigma_{\lambda}^2}} \quad (8)$$

in the absence of magnetic flux. The line profile emerging from the active region is equivalently:

$$I_{\text{magnetic}}(\lambda) = d\alpha e^{-\frac{(\lambda\xi_{\text{R}}-\mu)^2}{2\sigma_{\lambda}^2}} + d\alpha e^{-\frac{(\lambda\xi_{\text{L}}-\mu)^2}{2\sigma_{\lambda}^2}} + d(1 - 2\alpha)e^{-\frac{(\lambda-\mu)^2}{2\sigma_{\lambda}^2}}, \quad (9)$$

where ξ_{L} is equal to $1 + \epsilon$ and ξ_{R} is equal to $1 - \epsilon$ with ϵ representing the Zeeman-induced velocity shift (see equation 3) divided by the speed of light c :

$$\epsilon = \frac{1.4 \times 10^{-4} \lambda_0 g_{\text{eff}} B}{c}. \quad (10)$$

Subtracting the magnetic line from the non-magnetic line, we get

$$I_{\text{diff}}(\lambda) = d\alpha \left(2e^{-\frac{(\lambda-\mu)^2}{2\sigma_{\lambda}^2}} - e^{-\frac{(\lambda\xi_{\text{R}}-\mu)^2}{2\sigma_{\lambda}^2}} - e^{-\frac{(\lambda\xi_{\text{L}}-\mu)^2}{2\sigma_{\lambda}^2}} \right) \quad (11)$$

$$= d\alpha (g(\lambda; \xi_{\text{L}}) + g(\lambda; \xi_{\text{R}})). \quad (12)$$

with

$$g(\lambda; \xi) = e^{-\frac{(\lambda-\mu)^2}{2\sigma_{\lambda}^2}} - e^{-\frac{(\xi\lambda-\mu)^2}{2\sigma_{\lambda}^2}} \quad (13)$$

The expression in equation (13) is equal to the difference between two Gaussian absorption lines that are shifted relative to each other. It can be decomposed in terms of the Hermite–Gaussian polynomials as in Holzer et al. (2020):

$$g(\lambda; \xi) = \sum_{n=0}^{\infty} C_n(\xi) \psi_n(\lambda; \mu, \sigma_{\lambda}), \quad (14)$$

$r_{\text{active}} r_s f(\lambda + \Delta\lambda) + r_{\text{active}} r_s f(\lambda - \Delta\lambda) + (1 - r_{\text{active}}) f(\lambda)$. This simplifies to $I(\lambda) = (1 - 2r_{\text{active}} r_s) f(\lambda) + r_{\text{active}} r_s f(\lambda + \Delta\lambda) + r_{\text{active}} r_s f(\lambda - \Delta\lambda)$, which is equivalent to our expression. α is thus equal to the filling factor of magnetic regions times a factor related to the distribution of the orientation of the magnetic field vectors.

where ψ_n are the Hermite–Gaussian functions defined as in equation (A5) and C_n are coefficients (equations A3 and A4) that depend on the shift of the two Gaussians relative to each other. For more details, we refer to Appendix A or Holzer et al. (2020). By Taylor-expanding the relevant coefficients in ϵ , a simple expression emerges, as the terms that are odd in ϵ cancel out due to the symmetry of Zeeman splitting. The odd components $C_1\psi_1$ and $C_3\psi_3$ can be neglected for the same reason. Thus, we get the following convenient expression:

$$I_{\text{diff}}(\lambda) = 2d\alpha\epsilon^2 \left(\frac{\mu^2}{2\sigma_\lambda^2} + \frac{3}{8} - \left(1 + \frac{\mu^2}{2\sigma_\lambda^2} \right) \left(\frac{\lambda - \mu}{\sigma_\lambda} \right)^2 \right) e^{-\frac{(\lambda - \mu)^2}{2\sigma_\lambda^2}}. \quad (15)$$

Note that σ_λ denotes the standard deviation of the Gaussian absorption line profile in wavelength space. We can translate this to the velocity space and assume that the width of all lines in this space is about the same, as in LSD. $\frac{\mu}{\sigma_\lambda}$ then translates to $\frac{c}{\sigma_v}$ and $\frac{\lambda - \mu}{\sigma_\lambda}$ to $\frac{v - v_0}{\sigma_v}$ with v_0 the absolute radial velocity of the star:

$$I_{\text{diff}}(v) = 2d\alpha\epsilon^2 \left(\frac{c^2}{2\sigma_v^2} + \frac{3}{8} - \left(1 + \frac{c^2}{2\sigma_v^2} \right) \left(\frac{v - v_0}{\sigma_v} \right)^2 \right) e^{-\frac{(v - v_0)^2}{2\sigma_v^2}}. \quad (16)$$

4.2 Range of validity

4.2.1 Residual approximation

We included only the dominant quadratic terms of the Taylor expansion in our estimation of the residual shape in equation (16). Quartic and later terms only become relevant if $\frac{\mu}{\sigma_\lambda} > \frac{1}{\epsilon}$, i.e. if $\frac{\sigma_v}{c} < \epsilon$. Our approximation is valid if the width (σ_v) of the absorption lines in velocity space is larger than the Zeeman shift $1.4 \times 10^{-4} \lambda_0 g_{\text{eff}} B$. While the condition above resembles the weak-field condition, it only refers to the validity of the Taylor expansion.

Note that the expression in equation (16) is equally valid for local as well as for disc-integrated spectra, given our assumptions. The reason for this lies in the fact that we model absorption lines as Gaussian functions and rotational and macroturbulent broadening are well described by a convolution with a Gaussian kernel (Takeda & UeNo 2017; Sheminova 2019). A Gaussian $N(\mu, \sigma^2)$ convolved with a Gaussian kernel $N(0, \sigma_b^2)$ results in $N(\mu, \sigma^2 + \sigma_b^2)$ and therefore remains Gaussian in shape. Our residual shape is expressed as a sum of Gaussian functions. To translate this residual expression from the local to the disc-integrated spectra, it is convolved with a Gaussian kernel. Since the convolution is distributive, the Gaussian functions in the sum can be individually convolved with the broadening kernel. Therefore, the residual shape is equally a sum of Gaussian functions for the local and disc-integrated spectra. The only difference is the line width and line depth. Since we assumed constant σ_v for the line profiles, the line widths are equally constant for the local profiles. By convolving with a Gaussian kernel, the line depths change by a multiplicative factor. Therefore, the extracted residual shape amplitudes scale by a factor that depends on the broadening kernels. A multiplicative factor is of no concern for the purpose of RV detrending, however, and can remain unaccounted for. For our purposes, we investigate disc-integrated residual spectra, and therefore we need to ensure the Taylor expansion is a good fit to those spectra. Therefore, we set σ_v to the line width measured from these spectra.

Our approximation is generally valid for absorption lines of weakly active FGK-type stars in the optical wavelength range. For the Sun, σ is about 3 km s^{-1} and a 1 kG magnetic field produces a typical

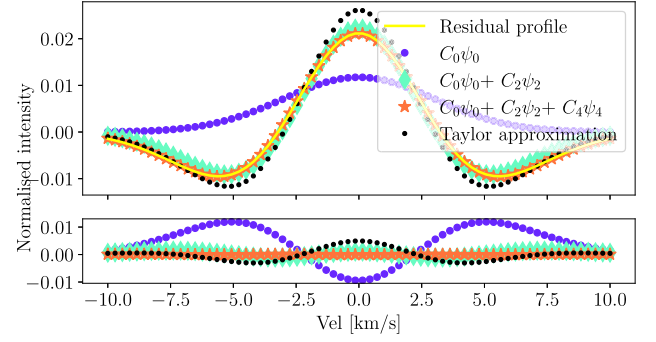


Figure 1. Upper panel: Residual profile emerging from subtracting a Zeeman-split absorption line exposed to a 2 kG field from a non-split line (yellow) and approximations thereof. Lower panel: Difference between the residual profile and the approximations.

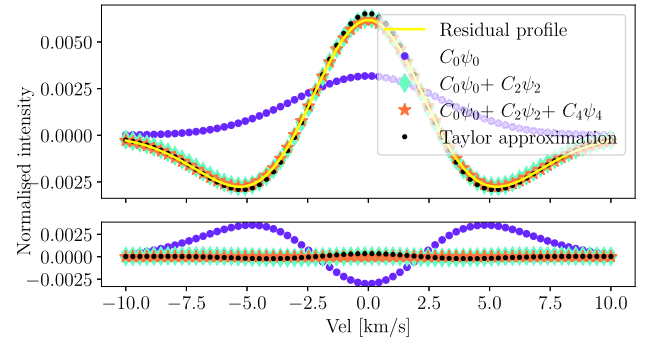


Figure 2. Upper panel: Difference between a Zeeman-split absorption line exposed to a 1 kG field and a non-split line (yellow) and approximations thereof, as in Fig. 1. Lower panel: Difference between the residual profile and the approximations.

Zeeman shift of about 1 km s^{-1} , which therefore comfortably lies within our range of validity. The typical magnetic field strength of plage regions on the Sun, however, is about 1.5 kG with some areas reaching up to 2 kG (e.g. Ruedi et al. 1992; Martínez Pillet, Lites & Skumanich 1997; Buehler et al. 2015). Similarly, magnetic field concentration in the solar network reach the same magnetic flux strengths (Buehler et al. 2019). In the following, we investigate the behaviour of the residual profile for a magnetically sensitive line for a magnetic field strength of 1 and 2 kG.

In Fig. 1, we show the residual profile computed from the original Gaussian expression in equation (11) in yellow, as well as the dominant Hermite–Gaussian components and our approximation as in equation (16). For this example, we chose an absorption line at 5000 \AA , with σ equal to 3 km s^{-1} , relative depth 0.3, g_{eff} of 2, α set to 0.1 and the magnetic field strength in the active region to 2 kG. This results in a Zeeman shift of 2.8 km s^{-1} which is just within our validity range of 3 km s^{-1} in this case. For comparison, we display the approximations for a 1 kG field keeping the other line parameters the same in Fig. 2. Since the residual profiles scale with B^2 , the amplitude is reduced by a factor of 4 in Fig. 2. Note that only small fractions of the solar surface are affected by such strong fields (e.g. Haywood et al. 2016) and only about 3 per cent of the included lines have an effective Landé factor greater than 2.

For both cases, $C_0\psi_0 + C_2\psi_2$ is already very close to the numerical solution. The difference between our Taylor approximation as in equation (16) and the Hermite–Gaussian approximations $C_0\psi_0 + C_2\psi_2$ and $C_0\psi_0 + C_2\psi_2 + C_4\psi_4$ lies mainly in the quartic

components. Given that the example shown in Fig. 1 treats a case that is at the very edge of the validity range and given the noise present in the spectra, including quartic components is not warranted. Also, the quartic components break the linearity of the problem that allows us to use the LSD approach in Section 4.3.

Apart from the question of when our Taylor expression in equation (16) is valid, we also need to consider when the underlying parametrization in equation (9) is valid on the local level where the lines are narrower. The parametrization in equation (9) is based on the weak-field approximation which breaks down for absorption lines with very high Landé factors in active regions with high magnetic field strengths (Jefferies, Lites & Skumanich 1989; Lehmann et al. 2015). The violation of this condition may have a larger impact on the extraction of $\Delta\alpha B^2$ for more active stars but overall holds for solar-type stars. For now, we thus recommend applying our model to the spectra of solar-type stars.

4.2.2 Gaussian absorption lines

In the derivation, we assumed a line is well-characterized by a Gaussian profile. Deviations from Gaussian profiles can arise due to pressure broadening leading to more prominent line wings. These absorption lines are generally better characterized by Voigt profiles. The residual profile emerging from these lines is still symmetric and broadly follows our approximation such that Zeeman-induced variability can still be captured. Another factor leading to deviations from the Gaussian line shape is stellar surface convection (see e.g. Gray 2005). The latter produces slightly asymmetric absorption lines in disc-integrated spectra (Gray 2005; Cegla et al. 2019). More specifically, the asymmetry is due to granulation. As we observe a star, we record both the blueshifted light emitted from upwards-flowing hot matter in the granules and the redshifted light emitted from the downwards-flowing cooler matter in the surrounding intergranular lanes. Averaged over the stellar hemisphere, this leads to C-shaped bisectors. A bisector is defined as the line connecting the mid-points of each horizontal segment of an absorption line (see e.g. Gray 2005, p. 297). For our purposes, the variation of the line shapes in time is more important than the line shapes themselves. To investigate the residual shapes caused by bisector variations, we analysed the CCF variations of HD 166435⁶ recorded with the ELODIE spectrograph (Baranne et al. 1996). This star is known to have large bisector variations that mimic the presence of a planet (Queloz et al. 2001). The CCF variations are unsurprisingly not symmetric relative to the line centre but antisymmetric instead. This result is expected because the line cores shift RV-like leading to shapes that are well-characterized by the odd first Hermite–Gaussian function but are not expected to interfere when fitting an even function that also naturally places most of the statistical weight on the line centres. We leave the detailed analysis of strong line shape variations on our residual approach to future work.

In our derivation, we assumed that the equivalent width is conserved. This is not the case for saturated lines displaying Zeeman intensification. Saturation is an issue when absorption lines are to be modelled. However, as we model residual spectra and the Zeeman-induced line variations are very small, the impact of saturation is reduced. Nevertheless, we expect improved results if saturation can be incorporated into the residual model.

⁶The spectra are available on the ELODIE archive: <http://atlas.obs-hp.fr/elodie/>.

Blended lines are also known to pose problems to magnetic field extraction techniques. For this reason, unblended lines are selected in most works (e.g. Stenflo & Lindegren 1977; Giampapa, Golub & Worden 1983; Saar 1988). As we include a large number of absorption lines, we expect the blend effects to be strongly diluted. We did, however, remove heavily blended regions and heavily saturated lines, as outlined in Section 5.1.

4.2.3 Zeeman triplets

We modelled the residual profiles as triplets, despite most absorption lines not being simple triplets. As explained in Section 2.2, an energy level with angular momentum J splits into $2J + 1$ sublevels, which is of course the case for both states involved in a transition. If both states have non-zero angular momentum and different Landé factors, we see multiple σ components at different wavelength shifts. The effective Landé factor is defined as the factor that captures the shift of the centre of gravity of the redshifted σ components (Landi Degl'Innocenti 1982).

Since the σ components are defined to originate from the transitions with $\Delta m = \pm 1$, they come in pairs of a redshifted and a blueshifted component, as long as there is a state to populate, with the π component in between. However, triplets can also be shifted relative to the non-magnetic case. This shift is proportional to the difference between the Landé factors of the upper and the lower state and produces a distribution of π components that is symmetric relative to the non-magnetic transition wavelength.

As long as the Landé factors of both involved states are very similar, we can treat them as a superposition of unshifted triplets with different strengths. Therefore, these non-triplets produce a superposition of residual profiles with the same shape but different amplitudes. To first order, the effective Landé factor squared captures the amplitude of this residual profile. Thus, we can also model non-triplets with the triplet model by using the effective Landé factor, as done in this study.

We investigated the dependence of our results on the non-triplet transitions. We could remove up to 60 per cent of the absorption lines without noticeably deteriorating the results. This cut corresponds to excluding all absorption lines for which the Landé factors of the upper and the lower state differ by more than 0.1. This shows that, on the one hand, non-triplets do not interfere with our extraction. On the other hand, it shows that the magnetic information in the non-triplet lines is not yet fully harnessed.

4.3 Convolution model

In this Section, we show how the line model can be applied on multiple lines simultaneously to boost the residual signal. For this, we assume that line residuals add up linearly, which allows us to model the residual spectrum as a convolution using the residual profile

$$R(v) = \left(\frac{c^2}{2\sigma_v^2} + \frac{3}{8} - \left(1 + \frac{c^2}{2\sigma_v^2} \right) \left(\frac{v - v_0}{\sigma_v} \right)^2 \right) e^{-\frac{(v-v_0)^2}{2\sigma_v^2}}. \quad (17)$$

scaled by the expected amplitude of the signal, $2d\alpha\epsilon^2$. Note that the amplitude of the signal can be split into a line-specific component

$$2d \left(\frac{1.4 \times 10^{-4} \lambda g_{\text{eff}}}{c} \right)^2 \quad (18)$$

multiplied by a general part αB^2 for all absorption lines. We now have a profile that is constant in velocity space and only scales with

a line-specific amplitude and a multiplicative factor. Therefore, we can use the LSD approach and model the convolution via matrix multiplication. For this, we need the following definitions:

- (i) λ_i : wavelength of pixel i .
- (ii) λ_l : central rest frame wavelength of absorption line l .
- (iii) d_l : depth of absorption line l .
- (iv) w_l : line-specific amplitude $2d_l(1.4 \times 10^{-4} \lambda_l g_{\text{eff}}/c)^2$.
- (v) v_j : velocity grid point j of residual profile R .
- (vi) $v_{il} = c \frac{\lambda_i - \lambda_l}{\lambda_l}$: radial velocity v_{il} which shifts λ_l to λ_i .

With these definitions, the convolution can be expressed as the matrix multiplication

$$I_{\text{diff,model}} = \alpha B^2 \mathbf{M} \mathbf{R} \quad (19)$$

using

$$\mathbf{M}_{ij} = \sum_l w_l \Lambda \left(\frac{v_j - v_{il}}{\Delta v} \right), \quad (20)$$

where Δv is the velocity increment, and Λ is defined as in Lienhard et al. (2022):

$$\Lambda(x) = \begin{cases} 0 & |x| \geq 1 \\ 1 - |x| & |x| < 1. \end{cases} \quad (21)$$

To find the best-fitting model to the data, we compute the value of αB^2 that minimizes

$$\chi^2 = (I_{\text{diff,measured}} - \alpha B^2 \mathbf{M} \mathbf{R})^\top \mathbf{S} (I_{\text{diff,measured}} - \alpha B^2 \mathbf{M} \mathbf{R}), \quad (22)$$

where the matrix \mathbf{S} contains the inverse squared flux uncertainties on the diagonal.

4.4 Non-magnetic profile

Above we estimated the residual profile arising from subtracting a non-magnetic line from a line affected by Zeeman splitting. However, we do not have an exact model for the unsplit line because the magnetic field is non-zero even in the quiet regions of the solar photosphere and we measure disc-integrated spectra. Robinson (1980) chose lines with low Landé factor as a model for the unsplit lines. However, for the line comparison to be meaningful, the formation heights of the split and the unsplit lines must be similar and the equivalent width must be comparable (Robinson 1980). Alternatively, the spectra of another magnetically quiet star of the same spectral type may be used as a comparison spectrum as in Giampapa et al. (1983). However, this approach comes with some complications, such as accounting for residual differences between the stars and estimating the magnetic flux of the quiet star (Saar 1988).

Another option consists in comparing the absorption lines to a reference spectrum of the same star (Saar 1988; Thompson et al. 2017, 2020). This means that each absorption line is compared to a line with the same formation height and practically identical equivalent width. Instead of choosing one spectrum as the reference, we compare each spectrum to the average spectrum computed from a selection of spectra. This reduces the extent of spurious variations due to telluric absorption lines, the continuum correction, and photon noise in the master profile leading to a cleaner comparison profile. We found this averaging procedure to be crucial as it significantly suppresses the photon noise in the template and therefore also reduces the scatter in the extracted values of αB^2 .

To generate the master spectrum, we stack 100 normalized solar spectra in the barycentric reference frame and fit a univariate spline.

The exact selection of solar spectra is negligible for this process. The residual profile evolves smoothly across the full velocity grid spanning 20 km s^{-1} with the two minima being about 10 km s^{-1} apart. This is orders of magnitudes larger than the expected RV shift due to planets. The extracted amplitude of the residual shape is thus not affected by planet-induced RV variations and v_0 in equation (17) can be set constant. Furthermore, Doppler-shift-induced residuals have a very different shape as compared to broadening-induced residuals. The extracted αB^2 is therefore largely independent of the presence of planets. Nevertheless, we shift the master profile to the RV of the individual spectra to match the derivation of the residual profiles.

Since we compare the absorption lines to their individual average profiles, we extract the change in magnetic flux strength $\Delta \alpha B^2$, rather than the absolute B or αB^2 :

$$I_{\text{diff,measured}} = I_{\text{master}} - I_i \quad (23)$$

$$= (I_{\text{non-magnetic}} - I_i) - (I_{\text{non-magnetic}} - I_{\text{master}}) \quad (24)$$

$$= I_{\text{diff}}(\alpha_i, \epsilon_i) - I_{\text{diff}}(\alpha_{\text{master}}, \epsilon_{\text{master}}) \quad (25)$$

$$= (\alpha_i B_i^2 - \alpha_{\text{master}} B_{\text{master}}^2) M \quad (26)$$

$$\propto \Delta \alpha B^2 \quad (27)$$

This expression is valid under the assumption that what actually evolves is the filling factor of active regions on the visible stellar hemisphere rather than the unsigned magnetic flux within the magnetically active regions. To verify this, consider a triplet as in equation (9). Keeping all parameters the same but varying only α , we see that the mean of any number of triplets can also be modelled as a triplet with well-defined parameters α and ϵ . Within our toy model, the difference between the master profile and a non-magnetic line can therefore be modelled with equation (12), as we did in equation (25).

However, our method does not rely critically on this assumption. Injection-recovery tests show that we can vary B within the defined validity range (cf. Section 4.2) and keep α constant and still recover $\Delta \alpha B^2$ with a mere constant offset which does not interfere with the linear correlations. Note that we measure $\Delta \alpha B^2$ relative to the master profile which means that the difference in $\Delta \alpha B^2$ between two spectra is equal to the difference between their respective αB^2 and thus an overall offset is not worrisome.

Measuring the absolute evolution of the magnetic flux requires two measurements to calibrate αB^2 and determine the offset. These measurements can be carried out with the established magnetic field estimation approaches. The technique presented in this study consequently also provides a gateway to get precise and cost-effective absolute unsigned magnetic flux time series.

Lastly, we assumed in equation (9) that the magnetic field strength is radially constant within the active and quiet regions. This assumption is not critical for our application because we are only interested in how the residual profiles evolve. In fact, this assumption is of no concern if the magnetic flux within both regions is constant and only the filling factor of active regions evolves in time.

4.4.1 Direct αB^2 extraction from the spectra

We investigated applying an extended version of LSD modelling the absorption lines as triplets to the spectra themselves rather than the residual spectra. However, the combined problem consisting of the problematic LSD-intrinsic line addition assumptions and imperfect absorption line depths represented a Gordian Knot that we could only

cut by applying the weak-field splitting approach on the residual spectra, at the expense of getting the variation in αB^2 rather than αB^2 itself. As mentioned, this does not impact RV mitigation and can be overcome by doing two calibration measurements.

4.5 Relation to hemispherically averaged unsigned magnetic flux $|\hat{B}_{\text{obs}}|$

As shown in the preceding sections, we extract $\Delta\alpha B^2$ from our spectra. SDO data shows that the changes in $|\hat{B}_{\text{obs}}|$ are mainly driven by the variation of the filling factor of active regions rather than the magnetic flux within those regions. In fact, in our SDO test data set containing one SDO observation at 6-h cadence from 2015 July to 2021 September, we find the overall filling factor of active regions to correlate almost perfectly with $|\hat{B}_{\text{obs}}|$ (Pearson correlation coefficient: 0.99). Similarly high correlations between the filling factor and $|\hat{B}_{\text{obs}}|$ were found for the data set analysed in Ervin et al. (2022). For the Sun, this correlation is mainly driven by faculae (Pearson correlation coefficient: 0.98) as the Sun is a faculae-dominated star. The correlation of the filling factor with the emission reversal in the Ca II H and K lines, as measured by $\log R'_{HK}$, and the correlation between $\log R'_{HK}$ and the average magnetic flux was shown in Meunier (2018) and is also discussed in Haywood et al. (2022).

Assuming no magnetic field in the quiet regions and constant magnetic field strength in the active regions, the only factor that evolves in time is α . This factor includes variations in the ratio of transverse to longitudinal field components, which we assume to be negligible in the disc-averaged spectra. In this case, α is directly proportional to the filling factor of active regions. Since we assumed a constant magnetic field strength in the active regions, αB^2 is directly proportional to the filling factor of active regions multiplied by their magnetic field strength. The latter is equal to $|\hat{B}_{\text{obs}}|$ in this idealised two-component model.

A two-component model is supported by evidence that plages exhibit fairly tight distributions of the magnetic field strength (e.g. Rueedi et al. 1992; Martínez Pillet et al. 1997; Buehler et al. 2015). This is due to the efficient concentration of the small-scale magnetic fields in flux tubes through the convective collapse mechanism (Parker 1978; Spruit 1979). For sunspots, there is a wider distribution of magnetic field strengths between spots of different sizes and within the individual spots themselves. The peak magnetic field strength is found in the umbra reaching 2000–3700 G and decreases towards the periphery of the spot to 700–1000 G (Solanki 2003). Since all of these components contribute to the amplitude of the residual signal described in this study, we expect the spread of the magnetic field strengths to impact the scaling of $\Delta\alpha B^2$ with $|\hat{B}_{\text{obs}}|$ for spot-dominated stars. Therefore, we expect the direct proportionality between $\Delta\alpha B^2$ and $|\hat{B}_{\text{obs}}|$ to hold for less active stars, as they are expected to be plage-dominated (e.g. Radick et al. 2018).

4.6 Comparison of methodology to other techniques

Skumanich & López Ariste (2002) applied principal component analysis (PCA) on intensity spectra. They found the first component to be equal to the first derivative of the profiles and their score to correlate with the RVs. This finding is in perfect agreement with the results from Holzer et al. (2020) who find that the difference between two shifted Gaussians is well-modelled by one profile with the shape of the first Hermite–Gaussian polynomial scaled with the stellar RV. Skumanich & López Ariste (2002) also find the second derivative to scale with the filling factor multiplied by the magnetic field strength

squared. This agrees with our findings that the Zeeman signature can also be modelled by one profile that scales with α and the squared unsigned magnetic field strength of the active regions. This result is also apparent from equation (10) in Stenflo (2013).

For the range of validity defined in 4.2, the difference between a line and a line broadened by a factor $1 + r$ while preserving the equivalent width can be expressed as

$$de^{-\frac{(v-v_0)^2}{2\sigma_v^2}} - \frac{d}{1+r} e^{-\frac{(v-v_0)^2}{2(\sigma_v(1+r))^2}}. \quad (28)$$

By Taylor expanding this expression for small width variations, i.e. for r near 0, it can be seen that this residual profile is practically identical to the expression in equation (16) for r equal $\alpha\epsilon^2\frac{c^2}{\sigma_v^2}$. Since the amplitude of the residual signal in equation (16) scales linearly with αB^2 , the residual profile in equation (28) scales linearly with r for small width variations. The derivation is shown in detail in Appendix B. The emergent residual profile is therefore not unique to Zeeman splitting. This also means we can model other broadening mechanisms by adding the same residual profile scaled by the appropriate factor. Also, if Zeeman splitting is modelled as a line-broadening effect, we expect to get the same scaling behaviour. Indeed, Stenflo & Lindgren (1977) fit the absorption line width of 402 unblended Fe I absorption lines as a polynomial expression with the magnetic factor scaling with B^2 .

Lehmann et al. (2015) applied a PCA approach on the moderately active star ϵ Eridani. They find one eigenprofile in good agreement with the second derivative of the line profiles and extract B^2 via the respective principal component score. We do not extract the magnetic flux profiles from the spectra themselves. Instead, we derive the shape and scaling behaviour from our weak-field triplet splitting model and thus assume the residual profiles to be known *a priori*. As also pointed out in Lehmann et al. (2015) and following directly from the Hermite–Gaussian expansion, the purely RV-induced residual variations are orthogonal to the Zeeman splitting induced shape variations. These components, therefore, interfere negligibly which makes the modelling of the RV effect in the residuals unnecessary. Lehmann et al. (2015) choose 30 spectral lines with Landé factor greater than 1.59. They generate a calibration mapping from the extracted B^2 values to the average magnetic flux by comparison with synthetic line profiles. Such a mapping is challenging for our technique because we include over 4000 absorption lines. The magnetic flux values of Lehmann et al. (2015) vary by a few tens of Gauss from spectrum to spectrum, with the average magnetic flux being 186 G. This error ratio may inhibit stellar activity mitigation.

There is direct observational evidence for the existence of the features that we derive from Zeeman splitting in the present analysis. Thompson et al. (2017) compared HARPS spectra of α Cen B to investigate the impact of stellar activity on absorption lines. Their data set spans a range of $\log R'_{HK}$ values from -5 to -4.82 and covers a sizeable fraction of α Cen B's activity cycle of about 8.1 yr (Ayres 2014). This range of $\log R'_{HK}$ is larger than for the solar data used in this work implying a stronger variation of the magnetic flux αB (Schrijver et al. 1989). They generated a low-activity stellar template by stacking the spectra recorded during a night in 2008 when α Cen B was most inactive within their data set. Thompson et al. (2017) then divided the nightly stacked spectra from 2010, when α Cen B was more active, by this template to investigate the differences. By visual inspection, they found features that closely resemble the Zeeman-induced residual shapes derived in the present analysis. The fact that Thompson et al. (2017) used division while we used subtraction does not significantly affect the morphological

similarity. However, the Hermite–Gaussian approach we use only applies to residual spectra produced by subtraction.

By simulating specific absorption lines, Thompson et al. (2017) deduced that magnetically active regions can produce the observed residual shapes. They furthermore found that the strength of the central component of the residual shape correlates with $\log R'_{HK}$. The residual shapes were also seen in a follow-up study in which Thompson et al. (2020) stacked daily HARPS-N solar spectra to discover a number of the same features as in α Cen B. They also found that the strength of those features correlated well with $\log R'_{HK}$ and with the facular filling factor which agrees with our findings. We do not visually see the Zeeman residual shapes, however, since we analyse individual spectra of the Sun. We can only deduce the magnetic flux proxy by combining the information from thousands of lines. This approach does make it possible to use this proxy as a simultaneous indicator for RV variations. As Thompson et al. (2017) see a strong correlation for single lines, we expect K-dwarfs such as α Cen B to be exquisite targets for our multiline approach too, as long as the extent of Zeeman splitting is within our range of validity and they are still plage-dominated.

5 APPLICATION TO HARPS-N SOLAR SPECTRA

In this Section, we describe how we applied the residual model to our data set of 500 HARPS-N solar spectra to extract $\Delta\alpha B^2$. We subsequently outline the results from injection-recovery tests using the solar spectra and describe a combination of the magnetic field modelling with RV modelling.

5.1 Preprocess spectra

For this analysis, we used the deblazed 2D echelle order spectra, their associated uncertainties, and barycentric wavelengths contained in the spectral files that we selected as described in Section 3.2. We continuum normalised the spectra with RASSINE Cretignier et al. (2020), corrected for residual cryostat leak effects, and divided by a simple telluric model. Details can be found in Lienhard et al. (2022), section 4.

We excluded any wavelength range in the barycentric frame that is affected by a telluric line deeper than 1 percent in any of the spectra. Such a strict threshold is warranted since we measure very small signals that could easily be distorted by telluric absorption lines (Cunha et al. 2014; Ulmer-Moll et al. 2019).

In Lienhard et al. (2022), the spectra were modelled by convolving the best-fitting common profile, representing the average line profile, with a line list containing the wavelength and depth of the absorption lines. For this, the velocity grid on which to evaluate the common profile had to be defined. We adopt the same velocity grid centred at the stellar RV of the first spectrum. The width of the velocity grid was set to 3 times the FWHM of the first common profile and the velocity increment to 0.82 km s^{-1} . The latter is equal to the average velocity increment per physical pixel on the HARPS-N CCD. This results in a grid width of about 20 km s^{-1} . Moreover, we excluded regions of the spectrum containing fluxes deviating by more than 0.5 in relative depth from the convolution model. This essentially removes lines that are heavily blended or absorption lines with inaccurately estimated depths in the VALD3 list. Furthermore, we included only atomic lines with relative depth greater than 0.2 as per the VALD3 list. We excluded all lines deeper than 0.9 and spectral regions near such absorption lines to avoid including heavily saturated absorption lines which are not well modelled by our residual model. This leaves us

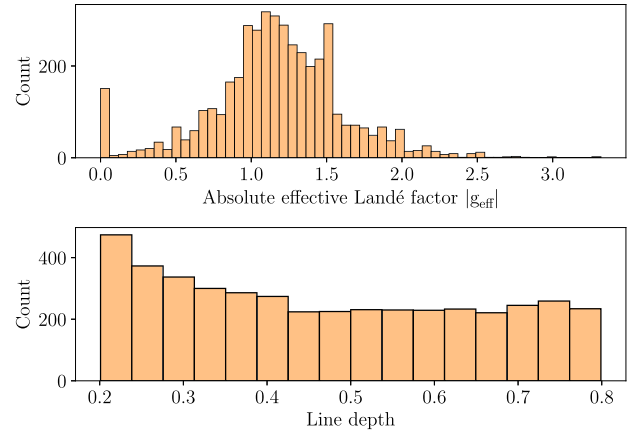


Figure 3. *Top panel:* Histogram of the effective Landé factors of the included absorption lines. *Bottom panel:* Histogram of the absorption lines depths from VALD3.

with 4636 absorption lines with mean effective Landé factor of 1.17. The distribution of the absolute effective Landé factors and the line depths is shown in the histograms in Fig. 3. We show the absolute value of the Landé factors because Zeeman splitting in Stokes I does not depend on the sign.

The correlations found in this study are negligibly dependent on the exact choice of the velocity grid width, the model-spectrum deviation, or the minimal and maximal depth of the included lines. Note that, as in the MM-LSD technique, we ensure that the same stellar absorption lines are included for all spectra.

5.2 Extraction from S2D spectra

We ran the residual fitting technique described in Section 4 on all echelle order spectra individually and combined the extracted $\Delta\alpha B^2$ values by computing the weighted mean of the extracted values of each order. The weight of each order was set to the sum of the inverse squared uncertainties of all included fluxes. The same weights were used in Lienhard et al. (2022) to combine the common profiles of the individual orders. We also tested running the extraction code on all orders simultaneously. This approach yielded marginally lower correlations with $|\hat{B}_{\text{obs}}|$.

Overall we find a good correlation of the extracted $\Delta\alpha B^2$ with $|\hat{B}_{\text{obs}}|$ from SDO for each order except order 65 (around 6650 \AA) where there are only very few included absorption lines. The Pearson correlation coefficients of $\Delta\alpha B^2$ with $|\hat{B}_{\text{obs}}|$ for each order are displayed in Fig. 4. The central orders between 5000 and 5500 \AA , where we find a high number of isolated absorption lines at high SNR, correlate best with $|\hat{B}_{\text{obs}}|$ and also get the highest statistical weight. For lower orders, the absorption lines are more often blended, affected by noise, and the continuum is expected to be less precise. At wavelengths beyond 5500 \AA , the number of included absorption lines can reach low values leading to high scatter in the extracted order $\Delta\alpha B^2$. Since these orders have a very low statistical weight, their impact on the final $\Delta\alpha B^2$ is minor.

We found it necessary to correct for an observational effect unique to the Sun. As the Earth revolves around the Sun on a slightly eccentric orbit that is inclined relative to the Sun's axis of rotation, the observed rotational line broadening evolves over the course of a year. The width variation is present in the HARPS-N solar spectra and can be modelled as a double-sinusoid (Collier Cameron et al. 2019). This signal results in a strong 182-d peak in the periodogram

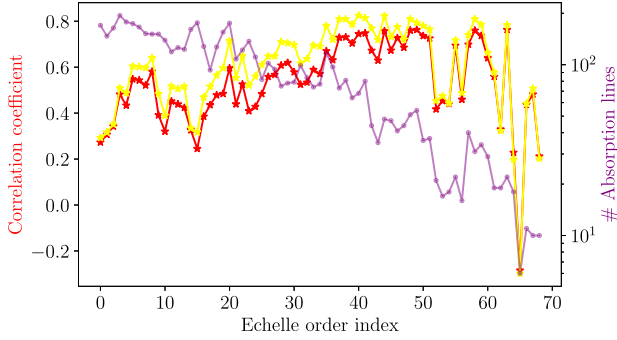


Figure 4. The red (yellow) stars show the Pearson correlation of the order of $\Delta\alpha B^2$ values with $|\hat{B}_{\text{obs}}|$ before (after) correcting for the Sun-Earth viewing angle effect. The purple dots indicate the number of included absorption lines per order.

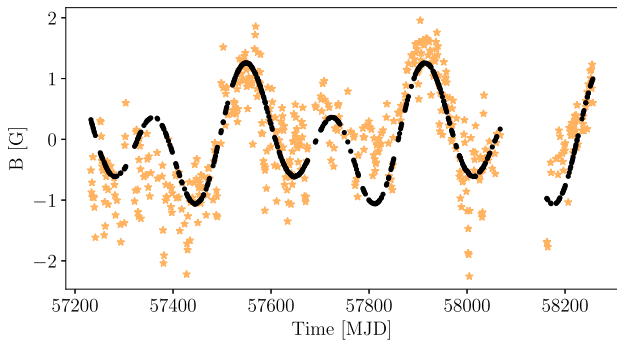


Figure 5. Difference between the scaled $\Delta\alpha B^2$ and $|\hat{B}_{\text{obs}}|$ (orange stars) and the model for the HARPS-N line width variations (black dots) caused by the varying viewing angle on the Sun.

of the width measurements. Our method, being susceptible to line width variations, is susceptible to this viewing angle effect too. However, spectra of other stars are not affected and therefore we remove this signal to get a more realistic estimate of the expected proxy performance for other stars. To remove the double-sinusoid signal, we find the scaling factor that eliminates the 182-d signal from the time series when we subtract the expected line width variation times this scaling factor from the $\Delta\alpha B^2$ time series. This is achieved by minimizing the power of the respective peak in the Bayesian generalized Lomb–Scargle periodogram (Mortier et al. 2015). We show the model and the difference between $\Delta\alpha B^2$ and $|\hat{B}_{\text{obs}}|$ in Fig. 5. Note that we do not use $|\hat{B}_{\text{obs}}|$ to compute this correction factor, but the additional signal is most obvious once the magnetic signal has been removed. The correlation coefficients of the individual orders increase by on average 0.06 after applying this correction. The overall correlation of $\Delta\alpha B^2$ with $|\hat{B}_{\text{obs}}|$ increases from 0.8 to 0.914 after the removal of this spurious signal. The CCF contrast and FWHM that we use in Section 6 have also been corrected for the viewing angle effect following Collier Cameron et al. (2019).

5.3 Injection-recovery test

We generated mock spectra to validate our approach given our assumptions. Each of these mock spectra is based on our real spectra. This means that the wavelength solution, the uncertainties and the SNR of the i th real spectrum correspond to those of the i th mock spectrum. To generate the mock fluxes, we produced a new line list by splitting each VALD3 line into a π and two σ components and

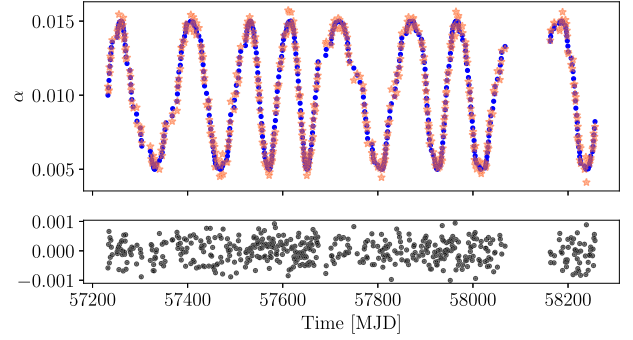


Figure 6. Upper panel: Scaled extracted $\Delta\alpha B^2$ (orange stars) and comparison to injected values of α (blue dots) assuming a two-component model with the magnetic field strength in the active region being 1.5 kG. Lower panel: Difference between scaled $\Delta\alpha B^2$ and injected values.

convolved this list with a normalized Gaussian profile scaled by the depth of the respective component as in equation (9). We set the field strength in the active regions to 1.5 kG and assumed a factor α between 0.005 and 0.015. We made the same assumptions as in Section 4. This means that we assume a homogeneous distribution of active regions making the average line-of-sight angle constant and absorbing this factor into α . Given the angle-dependent factor in equation (6) and assuming a characteristic angle (θ) of 34° as in Marcy (1982), the values of α chosen here correspond to about a filling factor of active regions of about 2 per cent which is within the plage filling factors presented in Milbourne et al. (2021).

We added Gaussian noise, with the standard deviation of each flux equal to its associated uncertainty estimate, to the spectra. As shown in Fig. 6, the LSD extraction of $\Delta\alpha B^2$ based on the Taylor-expanded Hermite–Gaussian expression successfully retrieves the injected values with no systematic differences and minimal scatter.

5.4 RV extraction

It is tempting to simultaneously model the RV and the magnetic effect within the residual profiles to obtain an RV estimate that is less affected by (1) stellar RV effects and (2) effects specific to the RV extraction.

The RV impact of stellar activity is expected to be reduced if we allow the spectral component emerging from the active regions to vary in relative strength, to be Zeeman broadened, and to be shifted in RV. RV extraction effects, on the other hand, are conceivable to originate from the Zeeman-induced width variations of individual lines leading to a varying degree of line blending, for example. This can cause shape variations of the LSD common profile or the CCF because neither of these methods perfectly models line blends. Zeeman intensification can furthermore influence the weight of magnetically sensitive absorption lines in the computation of the common profile. This makes the contribution of these lines to the common profile dependent on the magnetic field and thus produces an activity-dependent contribution to the extracted RV. Correcting for these RV extraction effects cannot remove the magnetically driven convective RV signals or RV variations caused by the inhomogeneous brightness distribution on the stellar surface. Equally, modelling the photometric effect requires an additional step even if we allow the emission from active regions to vary in strength and Doppler shift within our model.

The difference between two absorption lines that are Doppler shifted relative to each other is well described by the first Hermite–

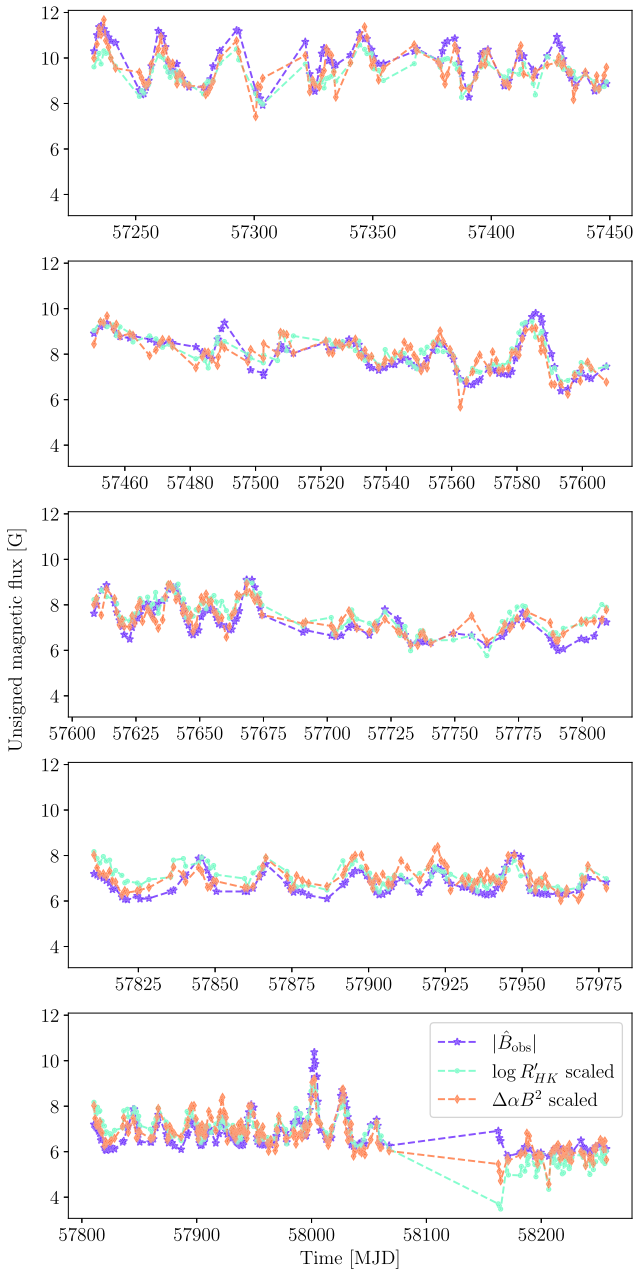


Figure 7. Evolution of $\Delta\alpha B^2$ and $\log R'_{HK}$, extracted from the HARPS-N solar spectra, and $|\hat{B}_{\text{obs}}|$ from HMI/SDO from 2015 July to 2018 May.

Gaussian polynomial and scales linearly with the RV, as shown in Holzer et al. (2020). Using the weak-field triplet splitting model, it can be shown that the RV extracted using the first Hermite–Gaussian polynomial is largely unaffected by Zeeman splitting. This may partially explain the lower RV scatter obtained in Holzer et al. (2020). With our line selection including partially blended lines and imperfect line depth estimates, we were able to extract RVs using this method. However, the RV semi-amplitudes were reduced which prohibited meaningful scatter analyses without discarding more data and degrading $\Delta\alpha B^2$. The reason for the suppressed RV amplitude can be deduced directly from equation (18) in Holzer et al. (2020). This expression assumes line additivity which is problematic for the partially blended lines which we included in our analysis.

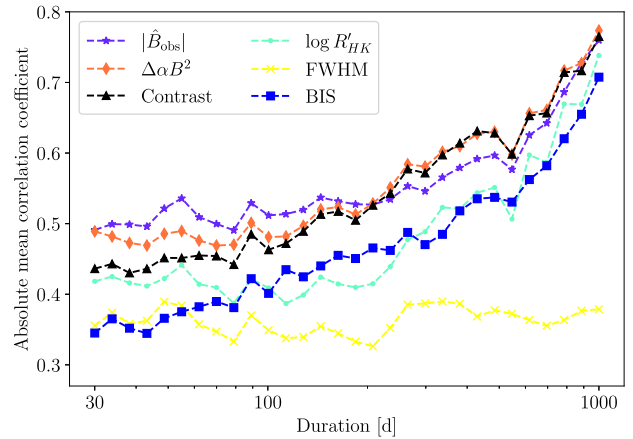


Figure 8. Absolute value of the mean Pearson correlation coefficient of the six activity indices with the heliocentric CCF RVs for data chunks covering a fixed time-span.

6 DISCUSSION

In this section, we present the magnetic flux proxy $\Delta\alpha B^2$ as extracted from the HARPS-N solar spectra and compare it with other available activity indicators and RVs.

6.1 Comparisons of activity indicators

In Fig. 7, we show the complete time series of $\Delta\alpha B^2$ and $\log R'_{HK}$ extracted from the HARPS-N S2D spectra and $|\hat{B}_{\text{obs}}|$ from SDO. The overall evolution of the magnetic flux and the quasi-periodic variations on the solar rotational time-scale are well traced by all three indicators.

We further evaluated how well our new indicator, $\Delta\alpha B^2$, traces the HARPS-N RV variations extracted using the CCF technique and compare it to the SDO $|\hat{B}_{\text{obs}}|$ and other standard activity indicators ($\log R'_{HK}$, BIS, FWHM and contrast). For this, we split the data into chunks of a given duration and computed the Pearson correlation coefficient between the RVs and the indicators within these chunks.

To divide the data with timestamps t_1, t_2, \dots, t_n into intervals of d days, we proceeded as follows. As the first chunk of data, we select all measurements taken between t_1 and $t_1 + d$. We then computed the Pearson correlation coefficient if this interval contained at least $\frac{d}{3}$ measurements and if these covered at least 90 per cent of the duration d . Next, we selected all measurements between $t_1 + d/3$ and $t_1 + d/3 + d$, proceeding the same way as above. We thus get a list of Pearson correlation coefficients for data chunks of duration d . In the following Figs 8, 9, and 10, we display the absolute value of the mean of the Pearson correlation coefficients for each time-scale d . We included all chunks of data irrespective of whether stellar activity impacted the data significantly. Excluding chunks with RV RMS below 1 m s^{-1} did not significantly alter our results. The correlation analyses were also computed including only chunks of data with RV RMS greater than 1.5 m s^{-1} . For these chunks of data, stellar activity is expected to be the dominant contributor to the RV variability with the photon noise contribution being around 0.2 m s^{-1} and the long-term instrumental stability of HARPS-N being near 1 m s^{-1} . The respective results are shown in Appendix C.

In Fig. 8, we show the absolute value of the mean of the correlation coefficients computed for these chunks of fixed duration. We computed the absolute value because the contrast is anticorrelated with $|\hat{B}_{\text{obs}}|$. This has also been noted in Costes et al. (2021).

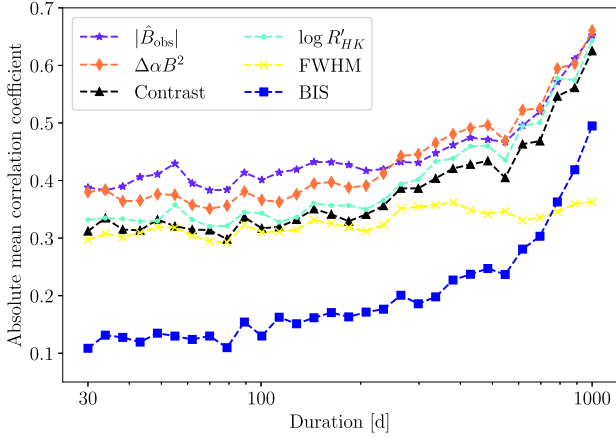


Figure 9. Absolute value of the mean Pearson correlation coefficient of the six activity indices with the heliocentric MM-LSD RVs for data chunks covering a fixed time-span.

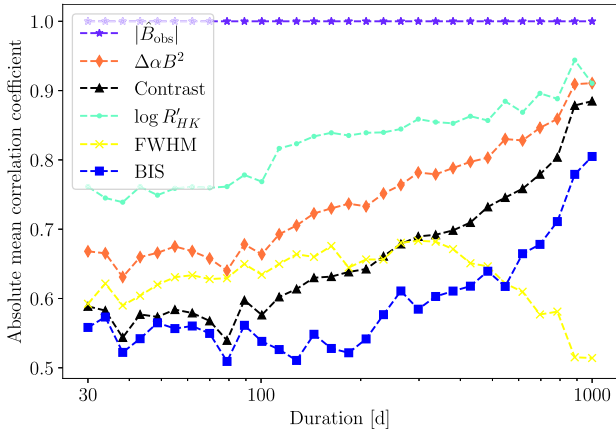


Figure 10. Absolute mean Pearson correlation coefficient of the five activity indices derived from HARPS-N spectra with $|\hat{B}_{\text{obs}}|$ for data chunks covering a fixed time-span.

The lower correlation coefficients towards shorter time-scales can partially be explained by phase offsets between the RV and the indicators, mentioned e.g. in Collier Cameron et al. (2019). These offsets are believed to originate from geometric effects and are therefore expected to be a function of the stellar rotation period. Higher correlations can be achieved by accounting for the time delay between the RVs and the activity indicators.

Unsurprisingly, $|\hat{B}_{\text{obs}}|$ traces the RV variations best as it originates from a dedicated magnetic field measurement with extremely high SNR resulting from combining the information from all relevant pixels, and is not affected by the Earth’s atmosphere. $\Delta\alpha B^2$ shows a consistently higher correlation with the RV variations as compared to the classical activity indicators. The contrast shows high correlations with the RVs too, but lower correlations with $|\hat{B}_{\text{obs}}|$ as compared to $\Delta\alpha B^2$. It is not surprising that the contrast and $\Delta\alpha B^2$ show some similarities because Zeeman broadening alters the depth of absorption lines where the signal-to-noise ratio is highest. Altogether, this makes $\Delta\alpha B^2$ the best-known activity indicator that can be extracted from Stokes I spectra simultaneously with the RV for solar-type stars.

One particular advantage of $\Delta\alpha B^2$ is that it can easily be extended to include other absorption lines and is not restricted to particular

wavelength ranges of detectors. $\log R'_{HK}$, in contrast, depends on collecting enough photons where the line reversal emerges around 4000 Å. Furthermore, $\log R'_{HK}$ weighs the contributions of surface areas differently to the RV extraction because the used wavelength regime, and therefore the limb-darkening, differ. $\Delta\alpha B^2$, on the other hand, weighs the individual echelle orders equivalently to the RV extraction. Another difference between $\Delta\alpha B^2$ and $\log R'_{HK}$ is that the flux reversal in the Ca II H and K lines forms above the photosphere, whereas the Zeeman splitting signature traces the magnetic fields in the photosphere where the absorption lines form. This also leads to $\log R'_{HK}$ essentially seeing a different field distribution since the magnetic field expands with height (Wiegelmann, Thalmann & Solanki 2014). Lastly, $\log R'_{HK}$ shows a non-linear correlation with the magnetic flux (e.g. Schrijver et al. 1989; Chatzistergos et al. 2019), whereas we expect $\Delta\alpha B^2$ to scale linearly within our model assumptions.

The same comparison for the MM-LSD RVs in Fig. 9 shows generally lower correlation coefficients. Strikingly, the BIS is significantly less correlated with the MM-LSD RVs than with the CCF RVs. This can be partially due to the BIS being calculated from the CCFs themselves rather than the MM-LSD common profiles. However, this would indicate that the BIS is method-specific and therefore less useful as a general diagnostic for the stellar surface conditions. Alternatively, this may indicate that the line selection for the CCF method is more susceptible to bisector variations. We performed a simple linear detrending of the RVs with $\Delta\alpha B^2$ to estimate the minimal RV RMS improvement as linear detrending is a crude method (for an overview on stellar activity mitigation, see Zhao et al. 2022). The RMS of the CCF RVs decreased from 2.01 to 1.25 m s⁻¹. The RMS of the MM-LSD RVs reduced from 1.65 to 1.22 m s⁻¹. The very similar final RMS values indicate that MM-LSD already picks up less stellar activity as compared to the CCF method. This is in agreement with evidence presented in Lienhard et al. (2022).

In Fig. 10, the correlation of the classical activity indicators and $\Delta\alpha B^2$ with $|\hat{B}_{\text{obs}}|$ is displayed. $\log R'_{HK}$ traces the $|\hat{B}_{\text{obs}}|$ variations better than $\Delta\alpha B^2$. Both indicators show a high correlation especially for time-scales longer than 200 d and trace the variations of $|\hat{B}_{\text{obs}}|$ better than the other classical activity indicators. This indicates that $\log R'_{HK}$, despite originating in principle in the chromosphere, and $\Delta\alpha B^2$ are good indicators for photospheric magnetic flux variations.

6.2 Signal-to-noise ratio dependence

The extracted $\Delta\alpha B^2$ does not depend crucially on the SNR of the spectra. Neglecting all but photon noise, we can add normally distributed noise to the spectra to simulate spectra at lower SNR. For each flux value, we randomly sampled from a Gaussian distribution with a mean of 0 and the variance equal to the squared uncertainty multiplied by a factor. To double the noise and simulate spectra with half of the original SNR, we set the mentioned factor to 1. The uncertainty estimates were adjusted accordingly.

By doubling the noise, we get a median SNR of 190, which is at the lower end of the expected SNR for very bright targets with visual magnitude of about 6. For these spectra, we still get a Pearson correlation coefficient of 0.91 for $\Delta\alpha B^2$ with $|\hat{B}_{\text{obs}}|$ overall. This is virtually identical to the Pearson correlation coefficient of 0.914 that we computed for the original spectra with an SNR of about 360. On the rotational time-scale of 30 d, the mean correlation coefficient with $|\hat{B}_{\text{obs}}|$ reduces negligibly from 0.67 to 0.66 as we double the

noise, whereas the mean correlation coefficient with the CCF RVs reduces from 0.49 to 0.48.

For simulated spectra with median SNR equal to 60, we get a very good overall correlation with $|\hat{B}_{\text{obs}}|$ of 0.89. The mean correlation with the $|\hat{B}_{\text{obs}}|$ on the rotational time-scale is slightly reduced from 0.67 to 0.56. The mean correlation with the CCF RVs reduces from 0.49 to 0.42. It follows that the described approach is negligibly affected by white noise within the typical SNR range achieved for RV studies.

7 CONCLUSIONS

We showed that the difference between a Zeeman-split and an unsplit absorption line is well modelled by Hermite Gaussian polynomials with Taylor-expanded coefficients. The expansion provided a residual model scaling with line-specific factors that allowed us to use the LSD framework to condense the information contained in over 4000 absorption lines and extract $\Delta\alpha B^2$. This indicator represents the hemispherically averaged unsigned magnetic flux variations. We found a correlation of $\Delta\alpha B^2$ with $|\hat{B}_{\text{obs}}|$ from SDO of 0.91 overall and about 0.67 on the rotational time-scale. Importantly, we find a minimal dependence of $\Delta\alpha B^2$ on the SNR of the spectra, which makes it a prime activity index even for faint stars. Also, we show in Section 6 that $\Delta\alpha B^2$ correlates better with activity-induced RV variations than the classical activity indicators, on the rotational time-scale and on the overall time-scale of 3 yr. For solar-type stars, $\Delta\alpha B^2$ is thus the best activity tracer known to date that can be simultaneously extracted with the RVs.

We expect this method to perform better if longer wavelength regions are included, provided that the telluric lines are properly corrected and there is a sufficient number of absorption lines. This is due to the Zeeman effect being more pronounced at longer wavelengths. The proxy for $|\hat{B}_{\text{obs}}|$ presented in this study provides many potential avenues for extensions and additional applications. For instance, it can provide a magnetic field estimator for stars for which the Calcium H and K lines are too weak, contaminated, or affected by instrumental effects. Also, this provides a magnetic flux estimator for instruments that do not cover the wavelength range required for extracting $\log R'_{HK}$. Furthermore, it provides an independent estimate of the evolution of the magnetic flux. The computation of this indicator can be done on readily available spectra that were recorded for RV purposes. No additional equipment or dedicated observation strategies are required.

Several potential modifications of the present approach are conceivable. For instance, additional line-broadening signals could be modelled simultaneously if temperature and magnetic relations are known. Regarding the Zeeman signal, the Doppler shift of the various components could be modelled as the magnetic and velocity fields are not independent (e.g. Cegla et al. 2013). Active regions can thus be Doppler shifted relative to the quiet photospheric regions. On the technical side, instead of comparing the stellar spectra to an averaged stellar spectrum of the same star, simulated spectra could be used.

Lastly, there are many applications and extensions for the presented indicator. The magnetic flux at different heights within the photosphere may be probed by making line selections. For our next steps, we intend to generalize this method to other stars and assess the performance of the indicator for stellar activity modelling.

ACKNOWLEDGEMENTS

The authors would like to thank the referee Steven Saar for his detailed and insightful comments. FL gratefully acknowledges a

scholarship from the Fondation Zdenek et Michaela Bakala. FL would like to thank Prof. J.O. Stenflo for useful discussions. AM acknowledges support from the senior Kavli Institute Fellowships. ACC acknowledges support from STFC consolidated grant numbers ST/R000824/1 and ST/V000861/1, and UKSA grant number ST/R003203/1. BK acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No 865624, GPRV).

DATA AVAILABILITY

The solar spectra are available on the Data & Analysis Center for Exoplanets (DACE) web-platform operated by the University of Geneva (<https://dace.unige.ch/sun>)

REFERENCES

- Ayres T. R., 2014, *AJ*, 147, 59
 Baranne A. et al., 1996, *A&AS*, 119, 373
 Basri G., Marcy G. W., 1994, *ApJ*, 431, 844
 Bellot Rubio L., Orozco Suárez D., 2019, *Living Rev. Solar Phys.*, 16, 1
 Bertaux J.-L., Lallement R., Ferron S., Boonne C., Bodichon R., 2014, *A&A*, 564, A46
 Brault J., Noyes R., 1983, *ApJ*, 269, L61
 Bruls J. H. M. J., Solanki S. K., 1995, *A&A*, 293, 240
 Buehler D., Lagg A., Solanki S., van Noort M., 2015, *A&A*, 576, A27
 Buehler D., Lagg A., van Noort M., Solanki S. K., 2019, *A&A*, 630, A86
 Cegla H. M., Shelyag S., Watson C. A., Mathioudakis M., 2013, *ApJ*, 763, 95
 Cegla H. M., Watson C. A., Shelyag S., Mathioudakis M., Moutari S., 2019, *ApJ*, 879, 55
 Chaplin W. J., Cegla H. M., Watson C. A., Davies G. R., Ball W. H., 2019, *AJ*, 157, 163
 Chatzistergos T., Ermolli I., Solanki S. K., Krivova N. A., Giorgi F., Yeo K. L., 2019, *A&A*, 626, A114
 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](https://arxiv.org/abs/2107.14291))
 Cretignier M., Francfort J., Dumusque X., Allart R., Pepe F., 2020, *A&A*, 640, A42
 Cretignier M., Dumusque X., Hara N. C., Pepe F., 2021, *A&A*, 653, A43
 Cunha D., Santos N. C., Figueira P., Santerne A., Bertaux J. L., Lovis C., 2014, *A&A*, 568, A35
 Desert M., Lagrange A. M., Galland F., Udry S., Mayor M., 2007, *A&A*, 473, 983
 Donati J. F., Semel M., Praderie F., 1989, *A&A*, 225, 467
 Donati J. F., Semel M., Carter B. D., Rees D. E., Collier Cameron A., 1997, *MNRAS*, 291, 658
 Donati J. F. et al., 2006, *MNRAS*, 370, 629
 Dravins D., 1982, *ARA&A*, 20, 61
 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
 Ervin T. et al., 2022, *AJ*, 163, 272
 Faria J. P. et al., 2022, *A&A*, 658, A115
 Folsom C. P. et al., 2018, *MNRAS*, 474, 4956
 Giampapa M. S., Golub L., Worden S. P., 1983, *ApJ*, 268, L121

- Gomes da Silva J., Santos N. C., Bonfils X., Delfosse X., Forveille T., Udry S., 2011, *A&A*, 534, A30
- Gray D. F., 2005, *The Observation and Analysis of Stellar Photospheres*. 3rd Edition, Cambridge University Press, Cambridge, UK
- Hale G. E., 1908, *ApJ*, 28, 315
- Halverson S. et al., 2016, in Evans C. J., Simard L., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 99086P
- Hanslmeier A., Nesis A., Mattig W., 1991, *A&A*, 251, 307
- Haywood R. D. et al., 2016, *MNRAS*, 457, 3637
- Haywood R. D. et al., 2022, *ApJ*, 935, 6
- Holzer P., Cisewski-Kehe J., Fischer D., Zhao L., 2021, *The Annals of Applied Statistics*, 15, 527
- Jefferies J., Lites B. W., Skumanich A., 1989, *ApJ*, 343, 920
- Johns-Krull C. M., Valenti J. A., Koresko C., 1999, *ApJ*, 516, 900
- Jurgens C., Fischer D., McCracken T., Sawyer D., Szymkowiak A., Davis A., Muller G., Santoro F., 2016, in Evans C. J., Simard L., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 99086T
- Kochukhov O., Wade G. A., 2016, *A&A*, 586, A30
- Kochukhov O., Makaganiuk V., Piskunov N., 2010, *A&A*, 524, A5
- Kochukhov O. et al., 2011, *ApJ*, 732, L19
- Kochukhov O., Hackman T., Lehtinen J. J., Wehrhahn A., 2020, *A&A*, 635, A142
- Lagrange A. M., Desort M., Meunier N., 2010, *A&A*, 512, A38
- Landi Degl'Innocenti E., 1982, *Sol. Phys.*, 77, 285
- Landolfi M., Landi Degl'Innocenti E., 1982, *Sol. Phys.*, 78, 355
- Lanza A. F., 2010, in Kosovichev A. G., Andrei A. H., Rozelot J.-P., eds, Proc. IAU Vol. 264, Solar and Stellar Variability: Impact on Earth and Planets. Kluwer, Dordrecht, p. 120
- Leet C., Fischer D. A., Valenti J. A., 2019, *AJ*, 157, 187
- Lehmann L. T., Künstler A., Carroll T. A., Strassmeier K. G., 2015, *Astron. Nachr.*, 336, 258
- Lienhard F., Mortier A., Buchhave L., Collier Cameron A., López-Morales M., Sozzetti A., Watson C. A., Cosentino R., 2022, *MNRAS*, 513, 5328
- Marcy G. W., 1982, *PASP*, 94, 989
- Martínez Pillet V., Lites B. W., Skumanich A., 1997, *ApJ*, 474, 810
- Mayor M. et al., 2003, *The Messenger*, 114, 20
- Medina A. A., Johnson J. A., Eastman J. D., Cargile P. A., 2018, *ApJ*, 867, 32
- Meunier N., 2018, *A&A*, 615, A87
- Meunier N., Lagrange A. M., 2019, *A&A*, 625, L6
- Meunier N., Lagrange A. M., 2020, *A&A*, 638, A54
- Meunier N., Desort M., Lagrange A. M., 2010a, *A&A*, 512, A39
- Meunier N., Lagrange A. M., Desort M., 2010b, *A&A*, 519, A66
- Meunier N., Lagrange A. M., Borgniet S., Rieutord M., 2015, *A&A*, 583, A118
- Milbourne T. W. et al., 2019, *ApJ*, 874, 107
- Milbourne T. W. et al., 2021, *ApJ*, 920, 21
- Mortier A., Faria J. P., Correia C. M., Santerne A., Santos N. C., 2015, *A&A*, 573, A101
- Noyes R. W., Hartmann L. W., Baliunas S. L., Duncan D. K., Vaughan A. H., 1984, *ApJ*, 279, 763
- Parker E. N., 1978, *ApJ*, 221, 368
- Pepe F., Mayor M., Galland F., Naef D., Queloz D., Santos N. C., Udry S., Burnet M., 2002, *A&A*, 388, 632
- Pepe F. et al., 2021, *A&A*, 645, A96
- Pesnell W. D., Thompson B. J., Chamberlin P. C., 2012, *Sol. Phys.*, 275, 3
- 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
- Prša A. et al., 2016, *AJ*, 152, 41
- Queloz D. et al., 2001, *A&A*, 379, 279
- Quirrenbach A. et al., 2016, in Evans C. J., Simard L., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 990812
- Radick R. R., Lockwood G. W., Henry G. W., Hall J. C., Pevtsov A. A., 2018, *ApJ*, 855, 75
- Reiners A., 2012, *Living Rev. Solar Phys.*, 9, 1
- Reiners A., Basri G., 2006, *ApJ*, 644, 497
- Rieutord M., Roudier T., Rincon F., Malherbe J. M., Meunier N., Berger T., Frank Z., 2010, *A&A*, 512, A4
- Rincon F., Rieutord M., 2018, *Living Rev. Solar Phys.*, 15, 6
- Robinson R. D. J., 1980, *ApJ*, 239, 961
- Rueddi I., Solanki S. K., Livingston W., Stenflo J. O., 1992, *A&A*, 263, 323
- Ryabchikova T., Piskunov N., Kurucz R. L., Stempels H. C., Heiter U., Pakhomov Y., Barklem P. S., 2015, *Phys. Scr.*, 90, 054005
- Saar S. H., 1988, *ApJ*, 324, 441
- Saar S. H., 2003, in Deming D., Seager S., eds, ASP Conf. Ser. Vol. 294, Scientific Frontiers in Research on Extrasolar Planets. Astron. Soc. Pac., San Francisco, p. 65
- Saar S. H., 2009, in Stempels E., ed., AIP Conf. Proc. Vol. 1094, 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun. Am. Inst. Phys., New York, p. 152
- Saar S. H., Donahue R. A., 1997, *ApJ*, 485, 319
- Saar S. H., Linsky J. L., 1985, *ApJ*, 299, L47
- Saar S. H., Piskunov N. E., Tuominen I., 1992, in Giampapa M. S., Bookbinder J. A., eds, ASP Conf. Ser. Vol. 26, Cool Stars, Stellar Systems, and the Sun. Astron. Soc. Pac., San Francisco, p. 255
- Scherrer P. H. et al., 1995, *Sol. Phys.*, 162, 129
- Schou J. et al., 2012, *Sol. Phys.*, 275, 229
- Schrijver C., Cote J., Zwaan C., Saar S., 1989, *ApJ*, 337, 964
- Seares F. H., 1913, *ApJ*, 38, 99
- Semel M., 1989, *A&A*, 225, 456
- Sheminova V. A., 2019, *Kinematics Phys. Celest. Bodies*, 35, 129
- Skilling J., Bryan R. K., 1984, *MNRAS*, 211, 111
- Skumanich A., López Ariste A., 2002, *ApJ*, 570, 379
- Solanki S. K., 2003, *A&AR*, 11, 153
- Spruit H. C., 1979, *Sol. Phys.*, 61, 363
- Stenflo J. O., 2013, *A&AR*, 21, 66
- Stenflo J. O., Lindgren L., 1977, *A&A*, 59, 367
- Stenflo J. O., Solanki S., Harvey J. W., Brault J. W., 1984, *A&A*, 131, 333
- Takeda Y., UeNo S., 2017, *PASJ*, 69, 46
- Thompson A. P. G., Watson C. A., de Mooij E. J. W., Jess D. B., 2017, *MNRAS*, 468, L16
- Thompson A. P. G. et al., 2020, *MNRAS*, 494, 4279
- Title A. M., Tarbell T. D., 1975, *Sol. Phys.*, 41, 255
- Ulmer-Moll S., Figueira P., Neal J. J., Santos N. C., Bonnefoy M., 2019, *A&A*, 621, A79
- Valenti J. A., Marcy G. W., Basri G., 1995, *ApJ*, 439, 939
- Wiegelmann T., Thalmann J. K., Solanki S. K., 2014, *A&AR*, 22, 78
- Yeo K. L., Solanki S. K., Krivova N. A., 2013, *A&A*, 550, A95
- Yu J., Huber D., Bedding T. R., Stello D., 2018, *MNRAS*, 480, L48
- Zeeman P., 1897, *Nature*
- Zhao L. L. et al., 2022, *AJ*, 163, 171
- Zirin H., Popp B., 1989, *ApJ*, 340, 571

APPENDIX A: EXPANSION IN HERMITE-GAUSSIAN POLYNOMIALS

$$g(\lambda; \xi) = e^{-\frac{(\lambda-\mu)^2}{2\sigma_\lambda^2}} - e^{-\frac{(\epsilon\lambda-\mu)^2}{2\sigma_\lambda^2}} \quad (\text{A1})$$

can be expanded as

$$\sum_{n=0}^{\infty} c_n(\xi) \psi_n(\lambda; \mu, \sigma_\lambda) \quad (\text{A2})$$

with $\epsilon = \xi - 1$.

For $n = 0$:

$$c_0(\epsilon) = \sqrt{\sigma_\lambda \sqrt{\pi}} - \frac{1}{\sqrt{\sigma_\lambda \sqrt{\pi}}} I_0 \times \left(\frac{1 + \epsilon + \frac{\epsilon^2}{2}}{\sigma_\lambda^2}, -\frac{2\mu + \epsilon\mu}{\sigma_\lambda^2}, \left(\frac{\mu}{\sigma_\lambda}\right)^2 \right) \quad (\text{A3})$$

for $n \geq 1$:

$$c_n(\epsilon) = -\sqrt{\frac{\sigma_\lambda n! 2^n}{\sqrt{\pi}}} \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m}{4^m m! (n-2m)!} \times I_{n-2m} \left(1 + \epsilon + \frac{\epsilon^2}{2}, \frac{\epsilon \mu}{\sigma_\lambda} (1 + \epsilon), \frac{1}{2} \left(\frac{\epsilon \mu}{\sigma_\lambda} \right)^2 \right) \quad (\text{A4})$$

with

$$\psi_n(\lambda; \mu, \sigma_\lambda) = \frac{1}{\sqrt{\sigma_\lambda 2^n n! \sqrt{\pi}}} H_n \left(\frac{\lambda - \mu}{\sigma_\lambda} \right) e^{-\frac{(\lambda - \mu)^2}{2\sigma_\lambda^2}} \quad (\text{A5})$$

H_n is the n th degree physicist's Hermite polynomial:

$$H_0(s) = 1$$

$$H_1(s) = 2s$$

$$H_2(s) = 4s^2 - 2$$

$$H_3(s) = 8s^3 - 12s$$

I_n is defined as:

$$I_0(a, b, c) = \sqrt{\frac{\pi}{a}} e^{\left(\frac{b^2}{4a} - c\right)} \quad (\text{A6})$$

$$I_1(a, b, c) = -\frac{\sqrt{\pi} b}{2a^{3/2}} e^{\left(\frac{b^2}{4a} - c\right)} \quad (\text{A7})$$

$$I_n(a, b, c) = -\frac{b}{2a} I_{n-1}(a, b, c) + \frac{n-1}{2a} I_{n-2}(a, b, c) \quad (\text{A8})$$

for all $n \geq 2$.

APPENDIX B: RESIDUAL PROFILE CAUSED BY SIMPLE BROADENING

Taylor expand equation (28) for r near 0:

$$d e^{-\frac{(v-v_0)^2}{2\sigma_v^2}} - d e^{-\frac{(v-v_0)^2}{2\sigma_v^2}} + d e^{-\frac{(v-v_0)^2}{2\sigma_v^2}} r - d \frac{(v-v_0)^2}{\sigma_v^2} e^{-\frac{(v-v_0)^2}{2\sigma_v^2}} r \quad (\text{B1})$$

which is equal to

$$d \left(r - r \frac{(v-v_0)^2}{\sigma_v^2} \right) e^{-\frac{(v-v_0)^2}{2\sigma_v^2}} \quad (\text{B2})$$

Equation (14) on the other hand was:

$$I_{\text{diff}}(v) = 2d\alpha\epsilon^2 \left(\frac{c^2}{2\sigma_v^2} + \frac{3}{8} - \left(1 + \frac{c^2}{2\sigma_v^2} \right) \left(\frac{v-v_0}{\sigma_v} \right)^2 \right) e^{-\frac{(v-v_0)^2}{2\sigma_v^2}}, \quad (\text{B3})$$

which is approximately equal to (keeping only the dominant terms, $\frac{c}{\sigma_v} \gg 1$):

$$I_{\text{diff}}(v) = d\alpha\epsilon^2 \left(\frac{c^2}{\sigma_v^2} - \frac{c^2}{\sigma_v^2} \left(\frac{v-v_0}{\sigma_v} \right)^2 \right) e^{-\frac{(v-v_0)^2}{2\sigma_v^2}} \quad (\text{B4})$$

Equation (B2) and (B4) are equal for

$$r = \alpha\epsilon^2 \frac{c^2}{\sigma_v^2} \quad (\text{B5})$$

Therefore, the residual profile from the broadened Gaussians matches the residual profile computed using the π and σ components for r set to the value in equation (B5). It follows that for small variations of the magnetic field, it does not matter whether we model the Zeeman effect as simple broadening or splitting, as the residual shapes are the same and $\Delta\alpha B^2$ can be extracted from either residual profile.

APPENDIX C: CORRELATIONS WITH HIGH ACTIVITY DATA CHUNKS

The Figs C1, C2, and C3 show the same correlation analyses as those shown in Figs 8, 9, and 10, but with the threshold for included data chunks set to 1.5 m s^{-1} .

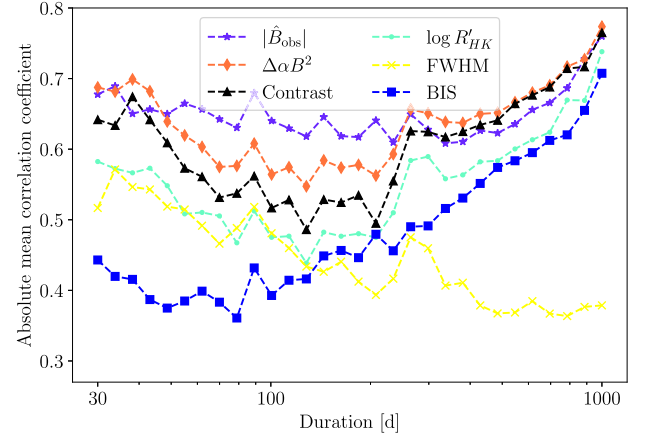


Figure C1. Absolute mean Pearson correlation coefficient of the six activity indices with the heliocentric CCF RVs for data chunks covering a fixed time-span.

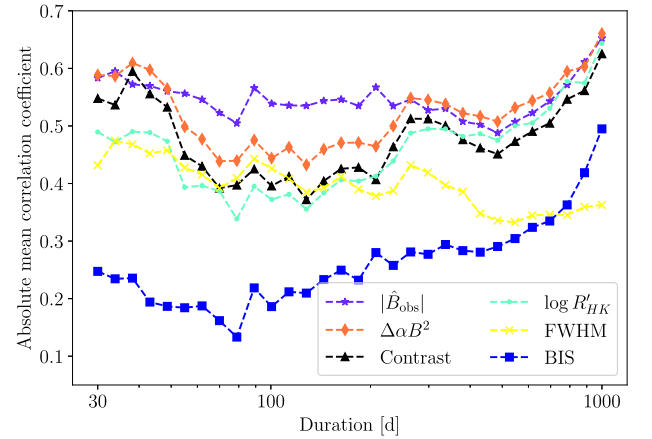


Figure C2. Absolute mean Pearson correlation coefficient of the six activity indices with the heliocentric MM-LSD RVs for data chunks covering a fixed time-span.

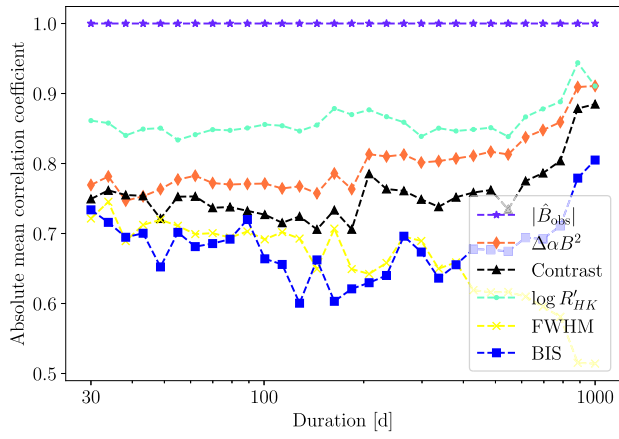


Figure C3. Absolute mean Pearson correlation coefficient of the five activity indices derived from HARPS-N spectra with $|\hat{B}_{\text{obs}}|$ for data chunks covering a fixed time-span.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.