All Features Great and Small:

Distinguishing the effects of specific magnetically active features on radial-velocity exoplanet detections

A dissertation presented

by

Timothy W. Milbourne

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

June 2021

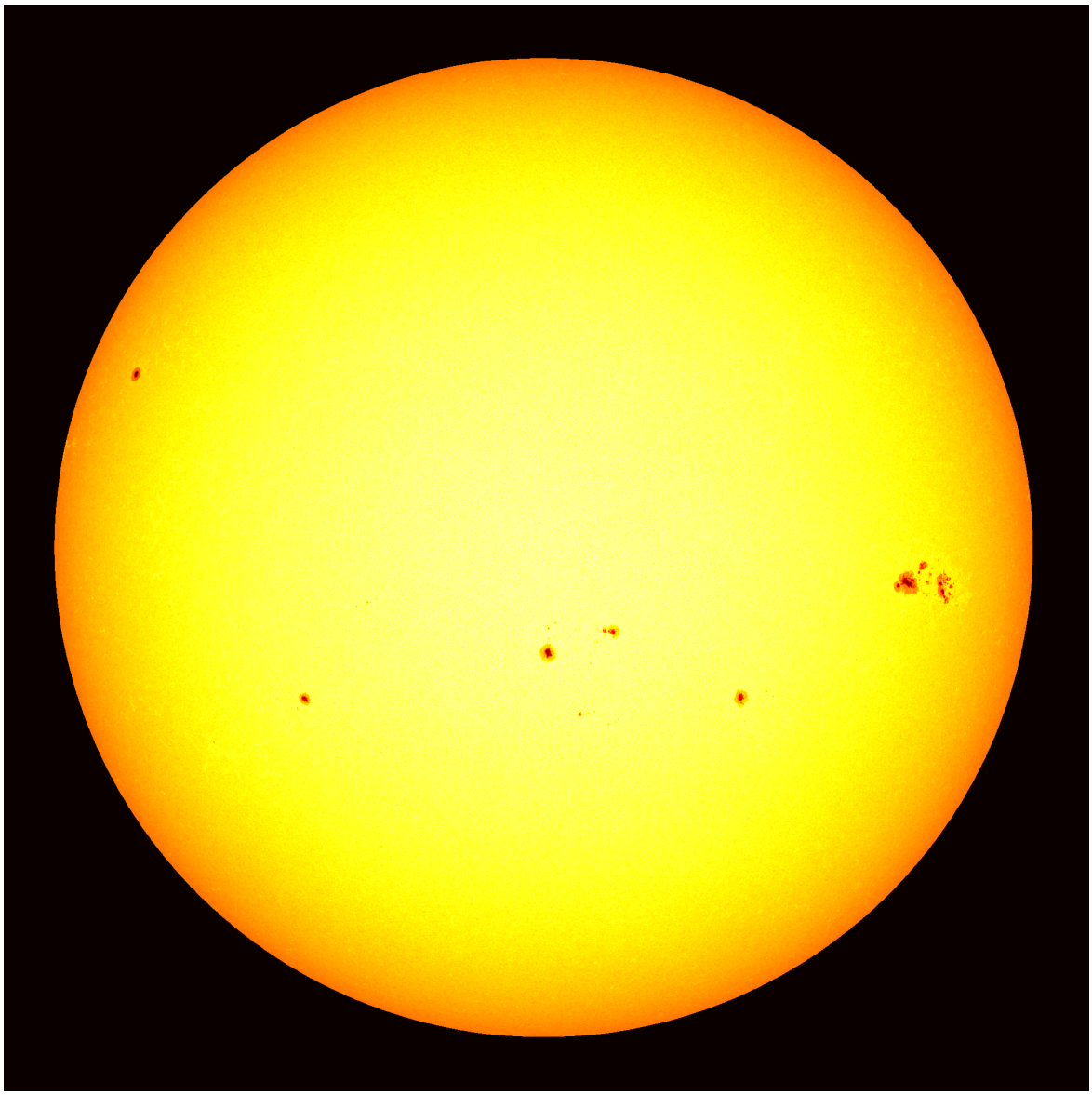

Photo: T. M./JPL Horizons

Dissertation Advisor:
Ronald Walsworth

Author:
Timothy Milbourne

All Features Great and Small:

Distinguishing the effects of specific magnetically active features on radial-velocity exoplanet detections

Abstract

State of the art radial velocity (RV) exoplanet searches are limited by the effects of stellar magnetic activity. Magnetically active features, such as spots, plage, and network regions, each contribute to the observed RV shift through a variety of mechanisms, including the suppression of convective blueshift and by creating rotational imbalance due to brightness inhomogeneities. However, the extent to which these RV contributions depend on the specific properties of individual active regions remains unknown. In this work, we investigate the effects of active region size on activity-driven RV variations, and develop tools for modelling these RV variations on Sun-like stars. We analyze solar observations acquired over Carrington Cycle 24 to test models of stellar magnetic activity and the resulting RV variations of Sun-like stars: we compare direct measurements of solar plage, spots, and network using the Solar Dynamics Observatory (SDO) to measurements of the solar RV and S-index from the solar telescope at the High Accuracy Radial velocity Planet Searcher for the Northern hemisphere (HARPS-N), solar photometry from the Solar Radiation and Climate Experiment (SORCE), and variations in the Sun's acoustic oscillation frequencies from the Birmingham Solar-Oscillations Network (BiSON). By comparing estimates of the contributions of the suppression of convective blueshift and the rotational flux imbalance derived from SDO images to

the HARPS-N solar RVs, we find that that magnetic active regions smaller than 60 Mm^2 do not significantly suppress convective blueshift. Differentiating the relative coverage, or filling factors, of these large plage regions from small network regions is thus necessary to differentiate between activity-driven RV signatures and Doppler shifts due to planetary orbits.

We then investigate several methods for extracting the relative coverage, or filling factors, of spots, plage, and network features. We demonstrate that variations in the solar p-mode frequency are highly sensitive to the presence of large plage regions, and are significantly less sensitive to smaller network regions, making the p-mode frequency a possible — though observsationally intensive — avenue for differentiating the contributions of different classes of active regions. We then develop a technique to estimate feature-specific magnetic filling factors on stellar targets using only spectroscopic and photometric observations. Linear and machine learning implementations of this technique both yield filling factor estimates that are highly correlated with the observed values. Modeling the solar RVs using these filling factors reproduces the expected contributions of the suppression of convective blueshift and rotational imbalance due to brightness inhomogeneities, providing an avenue for estimating these RV contributions on well-observed, Sun-like stars.

Contents

List of Figures

List of Tables

Citations to Previously Published Work

Parts of this dissertation cover results reported in the following articles:

1. G. Frustagli, E. Poretti, T. Milbourne, L. Malavolta, A. Mortier, V. Singh, A. S. Bonomo, L. A. Buchhave, L. Zeng, A. Vanderburg, S. Udry, G. Andreuzzi, A. Collier-Cameron, R. Cosentino, M. Damasso, A. Ghedina, A. Harutyunyan, R. D. Haywood, D. W. Latham, M. López-Morales, V. Lorenzi, A. F. Martinez Fiorenzano, M. Mayor, G. Micela, E. Molinari, F. Pepe, D. Phillips, K. Rice, and A. Sozzetti, "An ultra-short period rocky super-Earth orbiting the G2-star HD 80653," *Astronomy & Astrophysics*, **633**, A133 (2020).

2. T. W. Milbourne, R. D. Haywood, D. F. Phillips, S. H. Saar, H. M. Cegla, A. C. Cameron, J. Costes, X. Dumusque, N. Langellier, D. W. Latham, J. Maldonado, L. Malavolta, A. Mortier, M. L. Palumbo III, S. Thompson, C. A. Watson, F. Bouchy, N. Buchschacher, M. Cecconi, D. Charbonneau, R. Cosentino, A. Ghedina, A. G. Glenday, M. Gonzalez, C-H. Li, M. Lodi, M. López-Morales, C. Lovis, M. Mayor, G. Micela, E. Molinari, F. Pepe, G. Piotto, K. Rice, D. Sasselov, D. Ségransan, A. Sozzetti, A. Szentgyorgyi, S. Udry, and R. L. Walsworth, "HARPS-N solar RVs are dominated by large, bright magnetic regions," *The Astrophysical Journal*, **874**, 107 (2019).

3. W. J. Chaplin, R. Howe, S. Basu, Y. Elsworth, T. W. Milbourne, R. D. Haywood, G. R. Davies, S. J. Hale, A. Miglio, and E. Ross, "Sensitivity of low-degree solar p modes to active and ephemeral regions: frequency shifts back to the Maunder minimum," *Monthly Notices of the Royal Astronomical Society: Letters*, **489**, 1, (2019).

4. T. W. Milbourne, D. F. Phillips, N. Langellier, A. Mortier, R. D. Haywood, S. H. Saar, H. M. Cegla, A. Collier Cameron, X. Dumusque, D. W. Latham, L. Malavolta, J. Maldonado, S. Thompson, A. Vanderburg, C. A. Watson, L. A. Buchhave, M. Cecconi, R. Cosentino, A. Ghedina, M. Gonzalez, M. Lodi, M. López-Morales, A. Sozzetti, and R. L. Walsworth, "Estimating magnetic filling factors from simultaneous spectroscopy and photometry: disentangling spots, plage, and network," (in review: The Astrophysics Journal; *preprint arXiv:2105.09113*, (2021)).

Acknowledgements

This thesis represents the culimination of seven years of work, and it would not exist without the help and support of many people. As such, I wanted to take the opportunity to thank the people who helped and supported me over the course of my graduate school career:

First, I would like to thank my advisor, Ron Walsworth. I started work in Ron's lab the summer before my G1 year, and he immediately welcomed me into the Walsworth Group. I have learned a lot from Ron and all the talented scientists in his group, and I look back fondly on all of our discussions, beach trips, and group lunches we shared at Pho House.

I would also like to thank David Phillips. Since I joined the astrophysics side of the Walsworth group, David has acted as a second scientific advisor, and is always happy to share his wide-ranging scientific knowledge, look over a tricky part of a paper, or just have a cup of coffee and a nice conversation. David's mentorship has made me a far better scientist and communicator, and I am grateful for all his help, advice, and endless patience.

Thank you to Douglas Finkbeiner, Dimitar Sasselov, and John Doyle for serving on my thesis commitee. I am grateful for all their support and ideas over the years.

They say it takes a village to raise a child, and there is no better village than the Harvard University Physics Department: Thank you to Lisa Cacciabaudo, Melissa Franklin, Anna Klales, Greg Kestin, and everyone else who helped make the Physics Department a kind and comfortable home over the past seven years.

Similarly, thank you to David Latham, Steve Saar, and the other members of the HARPS-N Solar Telescope collaboration. The Solar Telescope meetings have been a constant source of scientific knowledge and great ideas. The Sun-as-a-star conference in Göttingen and the ISSI meeting in Bern were some of the most interesting and enjoyable parts of my time in graduate school. In

particular, thanks to Annelies Mortier, who provided the key insight that made all the analysis in Chapter 4 possible and useful — I am grateful for her enthusiasm about my work and her willingness to help. I also wanted to offer a special thanks to Raphaëlle Haywood, who has been a scientific inspiration and an excellent mentor. In many ways Raphaëlle has been my third graduate advisor, and I am grateful for all her help and support.

I am extremely grateful to have shared an office with Aakash Ravi and Nick Langellier (the "A Team"). Working with them on scientific problems, comiserating about the frustrations of grad school, tinkering with CoffeeBot, blaring the Anthems as we procrasinated with word puzzles, and holding our "board meetings" at José's have been highlights of my graduate student experience, and it has been a pleasure having them as teammates and friends.

I also want to thank Connor Hart for being a great friend and a great housemate — I am very grateful for all the movies, conversations, and waffle-iron cinnamon rolls we shared at Sacramento Street. Despite his questionable stances on Armando's pizza, it is indeed an honor to know Connor.

My parents are a constant source of wisdom and encouragement, and were the first people to teach me the value of math and science: I still remember vividly being quizzed with addition and subtraction flashcards as we waited in check-out lines at the store. They have supported my interest in physics at all points in my life, from taking me to the library as a child, to reassuring me as I stressed about finals in undergrad, to encouraging me now as I finalize this manuscript. Thank you Mom and Dad for your unconditional support during graduate school, my whole of my education, and my entire life. (And, of course, for keeping up with the flashcards in spite of my frustration.) I am also very grateful to my brother Andrew for all his words of encouragement and for keeping me sane and down to Earth during all the highs and lows of research.

Lastly, the biggest thank you to Kathryn Robenhymer, my fiancée and the love of my life. Kat celebrates with me in good times, supports me through bad times, and always believes in me, even when I do not believe in myself. I am grateful to have shared the past two years with her, and excitedly look forward to spending the rest of our lives together. I could not have asked for a better

life partner — thank you so much.

Thank you also to anyone I inevitably forgot to mention here. (A full list of everyone who helped and supported me throughout my PhD would surely be a hundred pages in its own right.) I am endlessly grateful to everyone who joined me on this journey with me, and I hope you will be with me as the next journey begins.

To Kat and my parents

Preface

The radial velocity (RV) method of exoplanet detection has emerged as a powerful tool for exoplanet characterization. In particular, the RV method allows observers to set bounds on planetary masses. However, state-of-the-art RV measurements are limited by the stars themselves: the intrinsic stellar activity process acting on the surface of a star may mask or mimic the Doppler shifts induced by planetary signals. The resulting activity-driven velocities may be in excess of 1 m s^{-1}. The Doppler shift produced by an Earth-like planet orbitting a Sun-like star, in contrast, is 10 cm s^{-1}, an order of magnitude smaller. These activity processes act on distinct timescales ranging from minutes to years, making it difficult to simply average away their effects. In order to successfully disentangle planetary and stellar signals, complete models of stellar activity are necessary.

On timescales of the stellar rotation period, activity driven RVs are dominated by the presence of magnetized regions on the surface of stars. These magnetically active regions induce RV changes into two ways. The presence of a magnetic regions may inhibit the motion of plasma, inhibiting convection and suppressing the overall covnective blueshift. Bright or dark regions also break the symmetry of the star across its rotation axis, resulting in an overall rotational RV signal as the region rotates across the surface of the star. Efforts to disentangle activity-driven signals and planetary signals are further complicated by the fact that different classes of active regions have different contributions to these processes. While they have the same spectroscopic signatures, for Sun-like stars dark sunspots are the dominant contribution to rotationally-driven photometric signals, while large, bright, plage regions primarily contribution to the suppression of convective blueshift. Small, bright network regions provide similar photometric and spectroscopic contributions to those of the larger plage regions, but have a smaller contribution to the suppression of convective blueshift. Successfully modelling activity-driven effects therefore requires knowledge of which types of active regions are on the surface of a stellar target.

While detailed knowledge of these different sources of activity-driven RVs is difficult to obtain for most stellar targets, such knowledge is easily obtained for the Sun. The Sun's proximity means

that its surface may be imaged with high spatial resolutions, and it is the subject of numerous RV, spectroscopic, and photometric observations spanning many years. It also means that we know the precise number and masses of the solar system planets, allowing us to exactly seperate planetary signals from activity-driven variations. This wealth of data means that it is an ideal test case for building more detailed models of stellar activity. We may combine "Sun as a star" observations (that is, disk integrated measurements that treat the Sun as it were any other stellar target) with spatially resolved measurements to determine exactly what information is required to mitigate activity-driven variations, and then identify methods for extracting this information on stellar targets. In this work, we use this framework to determine that knowlege of the size of magnetically active regions is necessary to disentangle activity-driven RVs and planetary signals, and identify two avenues for extracting this size information from stellar targets.

In this work, we discuss the RV method, the activity processes limiting it, and how the solar case allows us to build more complete models of magnetic activity, which may be applied to other stars. In Chapter 1, we provide a demonstration of an RV mass measurement of an ultra-short period planet, and discuss how different activity processes complicate RV planet searches. In Chapter 2, we consider the solar case, demonstrate that large plage regions have a different effect on activity-driven RVs than small network regions, and show that failing to make this distinction can result in spurious, long-term signals. In Chapter 3, we see that these large plage regions and smaller network regions have different contributions to the acoustic modes of the Sun and the evolution of these modes over the solar cycle, demonstrating the difference between these active regions effect observations over a variety of timescales. Finally, in Chapter 4, we present a technique to distinquish between plage, spot, and network regions using only photometric and spectroscopic observations, and discuss the application of these techniques to distant stars. Chapter 5 summarizes the remaining challenges in applying these techniques, and the next steps in the search for exo-Earths.

Co-worker Contributions

Chapter 1

Giuseppe Frustagli (INAF – Osservatorio Astronomico di Brera; Università degli Studi di Milano-Bicocca) and Ennio Poretti (INAF – Osservatorio Astronomico di Brera; Fundación Galileo Galilei-INAF) carried out the primary RV analyis of HD 80653, as well as a detailed analysis of the secondary eclipse present in the photometry. Luca Malavolta (Dipartimento di Fisica e Astronomia "Galileo Galilei"', Università di Padova) and Annelies Mortier (Astrophysics Group, Cavendish Laboratory; Kavli Institute for Cosmology, University of Cambridge) derived the stellar parameters of HD 80653 from the observed spectra, and obtained the stellar mass and radius by fitting stellar isochrones using the resulting stellar parameters. I performed an independent RV analysis of the system to determine an additional estimate of the stellar mass, and therefore the stellar density. That analysis is described in this work.

Chapter 2

I wrote a `MATLAB` implementation of the analysis pipeline for Helioseismic and Magnetic Imager (HMI) images — this pipeline was originally written by Raphaëlle Haywood (University of Exeter) and Andrew Collier Cameron (University of St. Andrews) in `FORTRAN`. I expanded this pipeline to compute the size and RV contribution of individual active regions, and used it to perform all the analysis preset in the resulting manuscript. Nick Langellier and Aakash Ravi provided invaluable technical advice (Harvard University; Harvard-Smithsonian Center for Astrophysics), and David Phillips (CfA) helped edit the manuscript and provided useful scientific advice throughout the project. Steve Saar (CfA) provided theoretical intuition regarding our interpretation of the results. Ron Walsworth (now University of Maryland) supervised the project. The remaining authors provided valuable feedback on the manuscript and helped build and maintain the solar telescope telescope, as well as operating the HARPS-N spectrograph.

Chapter 3

I adapted my HMI analysis pipeline to compute the magnetic flux associated with network and plage filling factors for all of Solar Cycle 24. These flux values were used as an independent check on the Wilcox Solar Observation magnetograms used in this work. William Chaplin (University of Birmingham) and the remaining authors performed the rest of the analysis and were responsible for writing the manuscript.

Chapter 4

I performed all the analysis described in this manuscript. Raphaëlle Haywood wrote an improved, `Python`-based version of my `MATLAB` HMI analysis pipleline. I helped debug and optimize this `Python`-based pipeline, and optimized it to improve its runtime. The HMI filling factors used in this work are the products of that pipeline. Annelies Mortier provided key insights to improve the linear estimations of the spot, network, and plage filling factors. Nick Langellier provided guidance on the proper implementation and training of the machine learning techniques described in this manuscipt. He and David Phillips provided invaluable technical and scientific advice, and helped edit the manuscript. Ron Walsworth supervised the project. The remaining authors provided valuable feedback on the manuscript and helped build and maintain the solar telescope telescope, as well as operating the HARPS-N spectrograph.

1

The Radial Velocity Method and its Limitations

1.1 Introduction

Much research in astronomy and astrophysics has been motivated by the ancient questions "what is our place in the universe" and "are we alone in the universe?" The search for habitable planets beyond our solar system is of foremost importance in answering these questions. However, the large distance scales involved with such searches has made the direct detection of extra-solar planets — or "exoplanets" impossible until very recently. However, over the past few decades, a variety of indirect methods for exoplanet detection have emerged, resulting in an explosion of exoplanet discoveries over the past 25 years.

In order to contextualize our solar system with respect to the galactic population of planetary systems, exoplanet discovery techniques must be able to generate a representative sample of the mass and radius distributions of exoplanetary systems. However, the smallest planets may be masked by physical processes intrinsic to their stellar hosts. These processes thereby bias the measured distributions of exoplanetary masses and radii, hampering our ability to understand our own solar system and to find potentially habitable Earth-like planets.

In this chapter, we introduce the transit and radial velocity (RV) method of exoplanet detection. While it is not the focus of this thesis, we provide a brief introduction and a short discussion of the transit technique for the sake of completeness. We discuss the RV method in further depth, describe the physics underlying this method, and derive the amplitude of the stellar RV induced by an orbiting planet. We provide an example of these methods in action, and use the RV technique to measure the mass of an ultra-short period (USP) planet. Using transit information, we also place measure the planet's radius, allowing us to estimate its density. We discuss the effects of stellar magnetic activity, the physics underlying each activity process, and how these processes limits the detection of low-mass, long-period exoplanets. Finally, we demonstrate how the studies of the solar case may be used to build improved models of these processes, and introduce several solar observation platforms used in this work

1.2 Transit method

While direct imaging of of exoplanets remains difficult (though not impossible, as demonstrated by [8, 9], and others.), various methods have emerged which allow observers to infer the presence of exoplanets. The observed brightness of a star is periodically diminished by the presence of an eclipsing planet. The transit method seeks out exoplanets by measuring integrated flux emitted by exoplanet host stars are as function of time. The presence of periodic ellipses in these brightness versus time measurements ("lightcurves") indiciates the presence of a transiting planets. The depth of these transit features provides constraints on the planetary radius, and the time between subsequent dips in brightness may be used to infer the planet's orbital period. In Fig. 1.1, we show a schematic of a transitting planet, and the characteristic transit pattern that appears in the lightcurve.

The radius of the planet may be found from the light curve by considering the depth of transit. The change in brightness ΔF during transit relative to the star's brightness when the planet is not transiting, F is related to the area of the planet relative to the area of the star. That is:

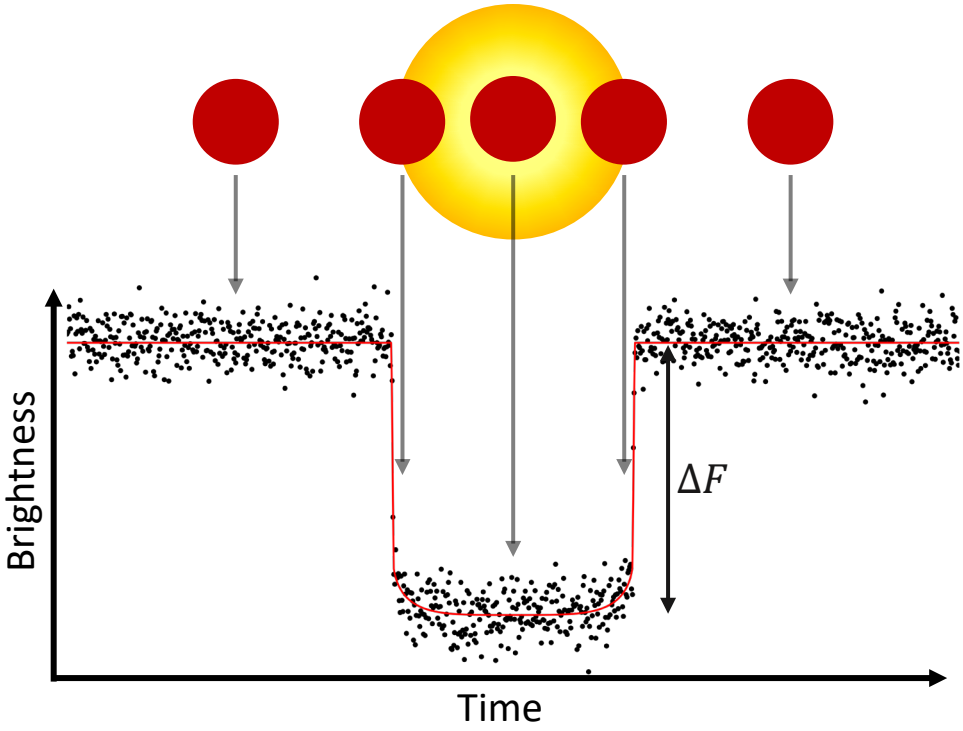

Figure 1.1: An orbiting planet (red disk) transiting its host star (yellow disk). The resulting transit light curve is shown below. The planet's radius relative to the stellar host may be determined from the change in flux observed during the transit, ΔF.

$$\frac{\Delta F}{F} = \left(\frac{R_{planet}}{R_\star}\right)^2 \tag{1.1}$$

Since its first use in 2000 [10, 11], the transit method has resulted in over three thousand confirmed exoplanet discoveries.[1]

The majority of these confirmations are due to the *Kepler* and *K2* missions [12, 13]. The Kepler space telescope was launched in 2009, and surveyed a fixed 115 square degree region of the sky. (0.25 percent of the sky). During its first four years of operation, the Kepler telescope identified over two thousand three hundred confirmed exoplanets. Failure of two of Kepler's reaction wheels resulted in the end of the primary *Kepler* mission. However, Kepler's work continued in the *K2* mission, which used solar pressure to stabilize the spacecraft as it surveyed a wider region of the

[1]For the most up to date value, please see the NASA Exoplanet Archive. Values are publically available at https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html

sky. *K2* has resulted in an additional four hundred confirmed planet discoveries.

In 2018, the NASA launched Kepler's successor, the Transiting Exoplanet Survey Satellite (TESS) [14]. The *TESS* mission surveys the entire sky in search of exoplanets — at the time of this writing, it has identified over a thousand exoplanet candidates, with 126 confirmed planets. In 2018, the European Space Agency launched CHEOPS (CHaracterizing ExOPplanets Satellite) [15], another space-based transit mission.

1.3 Radial velocity method

The radial velocity method provides a useful compliment to the transit method, allowing observers to infer the presence of exoplanets and set precise bounds on their masses. The RV method exploits the fact that the gravitation attraction between an orbiting planet and its host star does not change the center of mass of the star-planet system. The gravitational force exerted by the planet therefore causes the star to orbit the system's center of mass. This "wobble" of the star about the center of mass — also called the star's reflex motion — may be resolved using a sufficiently high-resolution spectrograph, which may be used to observe the periodic Doppler shift induced by the stellar reflection.

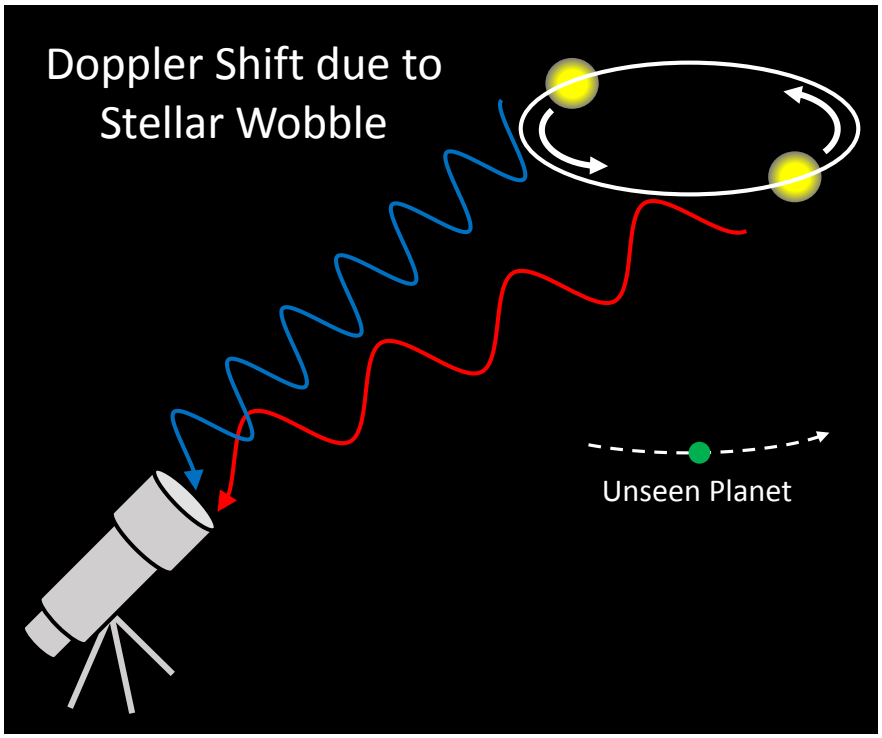

Figure 1.2: The gravitational interaction of a host star and an orbiting exoplanet causes the star to orbit the center of mass of the star-planet system. An observer on Earth may measure this stellar "wobble" using a spectrograph: as the planet moves toward the Earth-based observer, the star moves away from the observer, causing the star's absorption spectrum to be redshifted. Half a cycle later, when the planet moves away from the observer, star moves toward the observe, resulting in a blueshifted absorption spectrum. The amplitude of this periodic Doppler shift is related to the mass of the orbiting planet, as described by Eq. 1.7. *(Figure credit: N. Langellier)*

To gain an intuition for the magnitude of the stellar reflex motion induced by an orbiting planet, we will calculate the RV amplitude associated with a planet in a circular orbit around its host star. Here we follow a simplified version of the derivation presented in [16], as described in [17].

From Kepler's third law, we know that:

$$4\pi^2 a_p^3 / T_p^2 = G(m_s + m_p), \tag{1.2}$$

where G is the gravitational constant, T_p is the planet's period, and a_p is the semimajor axis of the planet's orbit. Assuming that the planet's mass is negligible compared to the stellar mass

($m_s \gg m_p$) and solving for a_p, we find

$$a_p = (Gm_s T_p^2/4\pi^2)^{1/3}. \tag{1.3}$$

From Newton's Third Law, the gravitational pull of the planet causes the star to orbit the center of mass of the star-planet system. Solving this two-body problem, we see that the semimajor axis of the star's orbit is related to the planetary orbit by

$$a_s = (m_p/m_s)a_p = (m_p/m_s) \times (Gm_s T_p^2/4\pi^2)^{1/3}. \tag{1.4}$$

For a circular orbit, the angular position of the planet in its orbit is given by $\theta = 2\pi t/T_p$. The position on the star along the line-of-sight axis, x, is thus given by

$$x = a_s \sin(2\pi t/T_p). \tag{1.5}$$

The velocity along the line-of-sight axis may be found by taking the derivative with respect to time:

$$v = \frac{\mathrm{d}}{\mathrm{d}t}\left(a_s \sin(2\pi t/T_p)\right) = 2\pi a_s/T_p \cos(2\pi t/T_p). \tag{1.6}$$

The amplitude of this sinusoid gives the amplitude of the star's reflex motion. Substituting our expression for $a_s$into the amplitude of the above equation, we find that

$$\mathrm{RV} = 2\pi a_s/T_p = \frac{2\pi}{T_p}\frac{m_p}{m_s}\left(\frac{Gm_s T_p^2}{4\pi^2}\right)^{1/3}. \tag{1.7}$$

Notice that the RV amplitude scales linearly with the mass of the orbiting planet. Plugging in values for Jupiters orbit around the Sun, we see that Jupiter induces a 12.5 m/sDoppler shift. The Earth, in contrast, induces a 9 cm/s shift as it orbits the Sun — this is the same velocity as turtle's crawl.

1.4 Demonstration of the RV Method: HD 80653

As discussed above, the RV method allows for the determination of planetary masses. By combining RV observations with transit measurements, which provide estimates of planetary radii, observers may estimate the densities of exoplanets.

The success of this technique is well demonstrated on HD 80653 b, an ultra-short period (USP) planet in orbit around a G2 star, HD 80653. HD 80653 is known to have a stellar mass $M_\star = 1.179 \pm 0.046 M_{sun}$ and radius $R_\star = 1.221 \pm 0.014 R_{sun}$[18–20]. This planet was observed by the spaced-based K2 transit-hunting mission for about 80 days between December 9, 2017 and February 25, 2018. RV observations were taken by the HARPS-N spectrograph (R = 115,000) installed on the 3.6-m Telescopio Nazionale Galileo (TNG), located at the Observatorio del Roque de los Muchachos in La Palma, Spain [21]. HARPS-N observed 115 spectra from November 2018 to May 2019.

An initial estimate of the planet candidate's period was provided by [22]. We may verify this period estimate by flattening the K2 lightcurve (that is, by fitting a low-order polynomial to the lightcurve and subtracting the best fit) to remove any long-term instrumental drift, and computing a Boxcar Lomb-Scragle (BLS) periodogram [23] to the flattened lightcurve. The BLS lightcurve fits the lightcurve to a periodic boxcar function over a grid of different orbital periods and durations, and returns the likelihood associated with each combination of parameters. The K2 lightcurve and HARPS-N observations of this system are shown in Fig. 1.3.

From the BLS periodogram, we find the planet's orbital period $P = 0.7196 \pm 0.0007$ days - the uncertainty on the period is derived by fitting the peak to a Cauchy distribution, as shown in Fig. 1.4. Since the orbital period of the planet is much lower than the rotation period of the star, $P_\star \sim 20$ days, we can easily separate the planetary signal by considering the RV measurements relative to constant offsets applied every two rotation periods. These offsets mitiage the effects of instrumental systematics or stellar process which occur on timescales longer than the planet's orbit, allowing us to isolate the planetary signal from these additional processes. The values of these offsets and the

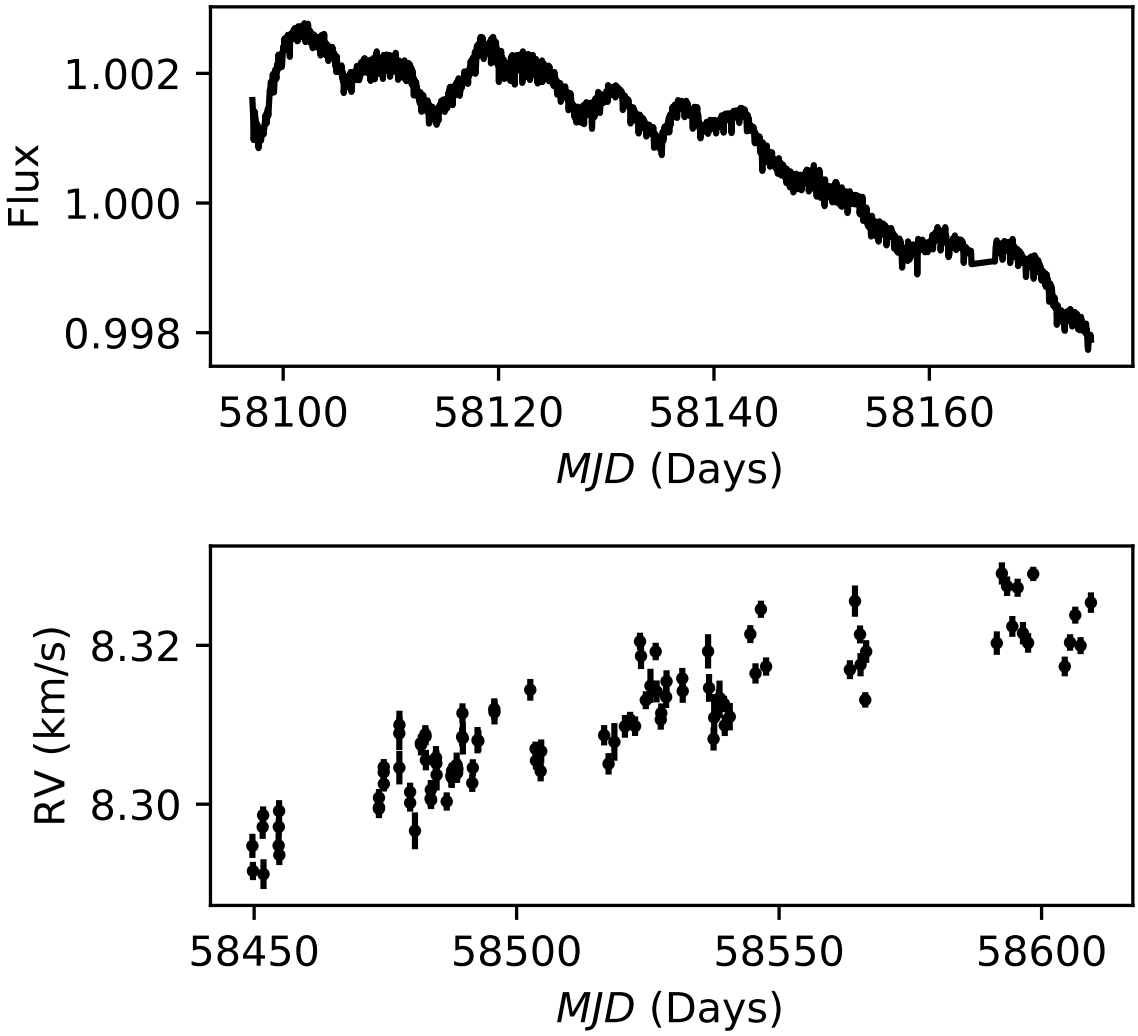

Figure 1.3: *Top:* HD 80653 lightcurve, as measured by K2. *Bottom:* HARPS-N measurements of the HD 80653 RVs.

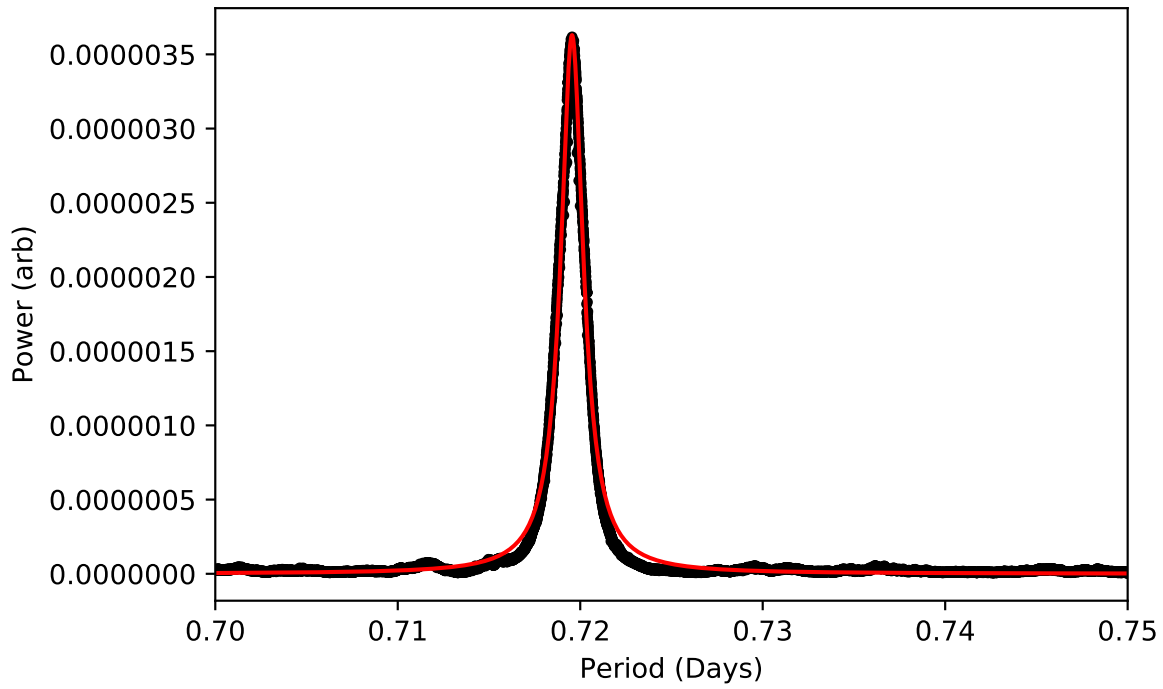

Figure 1.4: Boxcar Lomb-Scargle periodogram power as a function of period (black). By fitting this curve to a Cauchy distribution (red), we can extract the error on the period.

amplitude of planetary reflex motion may be found by fitting the measured RVs to the following expression:

$$\mathrm{RV} = K \sin\left(\frac{2\pi}{P} x(t)\right) + \sum_{i=0}^{N} B_i \Theta(t - t_0 - 2iP) \tag{1.8}$$

where K is the RV semiamplitude of the planet-induced reflex motion, B_i is the i-th RV offset, $\Theta(t)$ is the Heaviside step function, t_0 is the time of the first RV observation, N is the number of RV offsets, and $x(t)$ is the phase-folded exposure time, as given by the following expression:

$$x(t) = [(t - t_0 + P/2) \mod P] - P/2 \tag{1.9}$$

The planetary signals are easily visible in the phase-folded lightcurve and RV measurements (ac-

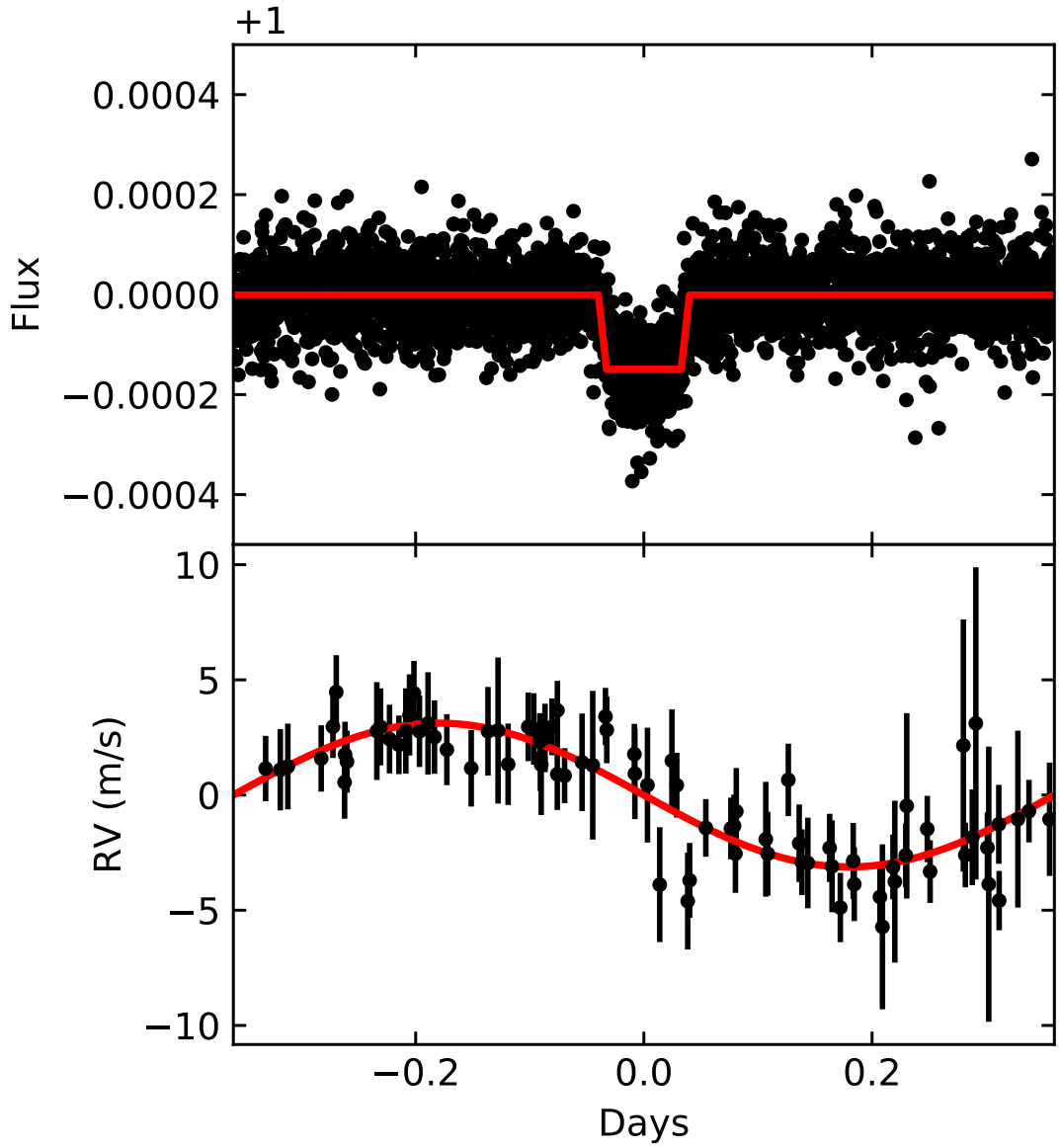

Figure 1.5: *Top:* Phase-folded K2 lightcurve. Data is in black, best-fit model in red. *Bottom*: Phase-folded HARPS-N RVs (minus nightly offsets). Data is in black, best-fit model in red.

counting for the offsets of Eq. 1.8), as shown in Fig. 1.5.

Fitting the RV observations to the above equation yields an RV semiamplitude of $K = 3.46 \pm 0.27\,\mathrm{ms}^{-1}$. Using Eq. 1.7, we find that this corresponds to a planetary mass of $M = 5.5 \pm 0.5 M_{earth}$.

Plugging the observed transit depth into Eq. 1.1, we find that HD 80653b has a radius of $R = 1.62 \pm 0.03 R_{earth}$. This results in a density $\sigma_b = 7.1 \pm 0.8\,\mathrm{gcm}^{-3}$. HD 80653b therefore has bulk density consistent with an Earth-like rocky planet (32.5% Fe/Ni-metal+ 67.5% Mg-silicates-rock).

Note that the relative ease with which we determined the mass and radius of HD 80653b was largely due to its ultra-short period. K2 was able to record nearly 114 transits of HD 80653b over its 80 days of observation, allowing us to determine its period and transit depth with a high degree of precision. Similarly, HARPS-N recorded over 140 orbits of HD 80653 around its host star. More importantly, however, the planet's orbital period is much shorter than the star's rotational period, allowing us to easily differentiate planetary signals from signals produced by the star itself. For planets with longer orbital periods, it is much harder to distinquish planets from these stellar signals. In the next section, we discuss the origins of these stellar signals, and the limits they place on planet detection.

1.5 Limits to the RV Method: Stellar Magnetic Activity

While it is relatively simple to identify the ultra-short period planet orbiting HD 80653, most exoplanet detections are not so straightforward. An Earth-like planet orbiting a Sun-like star, for instnace, induces a 10 cms^{-1} RV shift. For comparision, this is the speed of a tortoise's crawl. While previous RV measurements were limited by long-term instrumental stability, previous work by previous members of the Walsworth group, including Chih-Hao Li, Alex Glenday, Aakash Ravi, David Phillips, and others have produced laser frequency combs to calibrate astronomical spectrographs [24–27]. These "astro-combs" are capable of providing the long-term instrumental stability needed to detect a $10\mathrm{cms}^{-1}$shift over the several-year observational baseline needed to detect the orbit of an Earth-like planet.

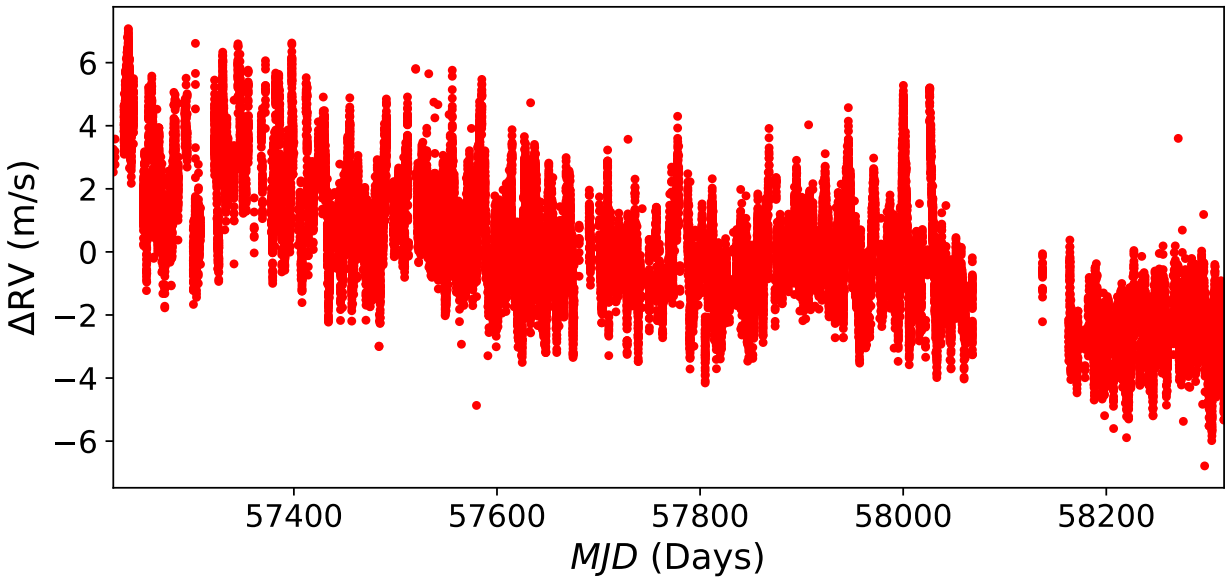

Figure 1.6: The radial velocity of the Sun in the heliocentric frame is shown *(red points)* — if the Sun were completely inactive, these RVs would be zero. However, the presence of activity processes on the solar surface results in variations with an RMS amplitude of 2 m s^{-1}. Understanding the physical origins of these variations are crucial for the succesful detections of low-mass, long-period exoplanets: The RV signal of an Earth-mass planet in a circular orbit in the habital zone (that is, with an amplitude of 10 m s^{-1}, a period of 365 days, and zero eccentricity) is shown for comparision *(black line)*.

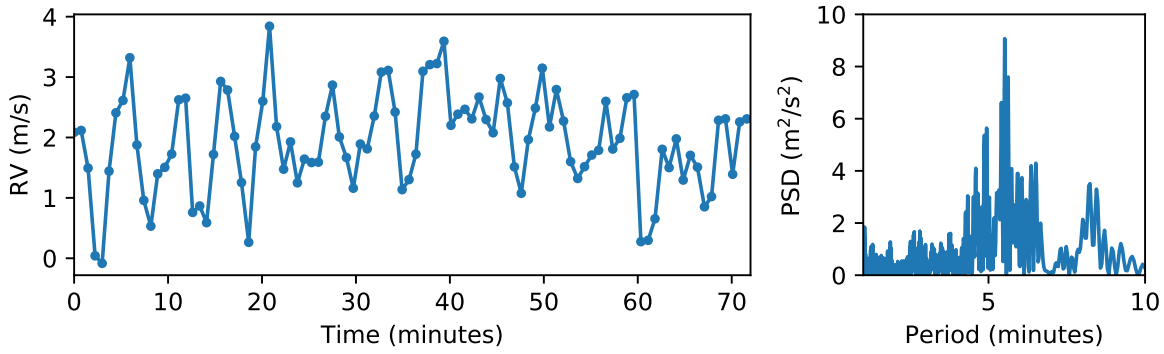

Figure 1.7: *Left:* HARPS-N observations of p-mode RV oscillations. Observations taken on March 18th, 2016 starting at 10:20:20. *Right*: Lomb-scargle periodogram estimate of the p-mode power spectrum.

State of the art RV surveys are now limited by the intrinsic variability of stars themselves. These variations are the result of various physical processes, each operating on a unique timescale: different tools must therefore be applied to mitigate each process. Below, we summarize the physical origins of each process, its contribution to the apparent RV, and strategies currently used to mitigate the effects of each process on planetary signals.

1.5.1 Acoustic oscillations

Acoustic oscillations propogating throughout stars result in "breathing modes", which cause the apparent solar radius to oscillate with time. These p-mode oscillations have an RV amplitude of several ms^{-1}, and a characteristic timescale of several minutes — on the Sun, a periodogram shows a forest of peaks centered around 5 minutes[28].

Helioseismology, the study of these p-modes, is itself a rich field of astrophysics, and can provide a valuable insights about the properties of stars. However, for the purposes of exoplanet detection, to is necessary to remove the RV signatures of p-modes from observations. Many telescopes will take five to ten minute exposure to average over these p-modes. However, a variety of more sophisticated techniques for mitigating the effects of p-modes exist. For example, W. Chaplin and H. Cegla have developed a set of tools to determine the optimum exposure time to reduce the RV amplitude of p-modes [29].

1.5.2 Magnetoconvection

In addition to the accoustic modes described above, energy is carried through stars through several other processes. Near the core of Sun-like stars, radiation provides the most efficient means of moving energy from the star's core to its outer layers. However, above a certain radius, the opacity of the stellar plasma inhibits radiative transport, and convection becomes the most efficient means of transporting energy. Plasma near the bottom of this convective zone is heated by the trapped radiative energy, and becomes less dense than the surrounding plasma. As a result, it rises toward the surface of the star, driven by the bouyant force. As the plasma reaches the surface of the star, it cools, increasing in density, and falls back into the star. Sun-like stars are therefore covered with a patchwork of these convective cells. Each cell spans several megameters (about the size of Texas). Hot, bright plasma in the center of cell rises toward the star's surface, while cool, dark plasma on the edge of each granule falls back toward the center of the star. Taking a brightness-weighted average over each granule, we see that there is more, brighter rising (blueshifted) plasma, versus the darker, sinking (redshifting) plasma, leading to an overall convective blueshift associated with each convective cell. These granules will merge and split, causing the convective blueshift to evolve on timescales between several hours and several days [30–34]. This results in RV variations on the order of a few ms^{-1}. To mitigate the effects of these variations, many telescopes will perform repeated observations of the same target at several points throughout the evening in an effort to average out these convective variations, though other, more sophisticated schemes to mitigate the effects of convection have been proposed [35].

1.5.3 Suppression of Convective blueshift

Stellar dynamos can produce extremely strong magnetic fields. These strong fields can interact with the moving, charged plasma, resulting in a variety of physical effects. In particular, the presence of strong stellar magnetic fields can halt the motion of convecting plasma — this reduces the temperature of the magnetized region, resulting in a dark starspot. These dark spots are often surrounded by bright plage regions, in which narrow magnetic flux tubes pierce otherwise inactive

regions of plasma. These flux tubes increase the optical depth of the region, allowing further visiblity into the hotter stellar core, and making them appear brighter than the surrounding quiet surface [36]. (Note that the magnetohydrodynamics governing the interactions between stellar magnetic fields and the charged plasma, as well as the formation of these magnetically active regions is a rich, complex, and active field of research in its own right. However, it is beyond the scope of this work — here we consider these active regions in the context of their effects on RV meaxurements.)

As mentioned above, these strong magnetic fields interact with the charged plasma, inhibiting the motion of the convective cells which cover the stellar surface. This results in a suppression of the convective blueshift, which changes the measured stellar RV. Since starspots and plage have lifetimes of several rotation periods, we therefore expect to see quasi-periododic RV variations on timescales of the rotation period, with an amplitude varying on timescales of several rotation periods [1, 2, 7]. In the solar case, these variations can range from 1 to 10 m s^{-1}. Traditionally, these processes are modelled using activity indicators derived from one or more features in the stellar spectrum. (The calcium S-index, for example, is discussed in Sec. 1.6.1). However, several groups have created more sophisticated models of these variations using a Gaussian process with a quasi-periodic kernel function [37–39].

1.5.4 Photometric asymmetries

In addition to the suppression of convective blueshift, the bright plage and dark spots described above change the rotational profile of the star. On a rotating star, light emitted by the side rotating toward the observer is blueshifted, and the light emotted by the side rotation away from the observer is redshifted. On a perfectly quiet star, the blueshift associated with the half of the star rotating towards the observer perfectly cancels the redshift associated with the half of the star rotating away from the observer, resulting in no observable change in the stellar RV.

However, the presence of bright plage and dark spots break the symmetry of the rotating star, resulting in a net blueshift as a bright region rotates toward the observer (or as a dark region

rotates away from the observer), and a net redshift as a bright region rotates away from the observe (or as a dark region rotates toward the observer). This results in RV pertubations on timescales of the star's rotation period, resulting in RV variations of several meters per second. These variations may be estimated from a stellar lightcurve [40]. Like the suppression of convective blueshift, these variations are also well modeled using a Gaussion process using a quasi-periodic kernel.

1.5.5 Stellar magnetic cycle

On timescales of several years, the overall activity level of stars is modulated by their long-term magnetic cycle. In the solar case, this cycle has a period of 11 years. The amplitude of the convective and photometric RV variations are therefore modulated by this long-term activity cycle. These long-term variations can mask low-amplitude Doppler shifts, and failing to account for these signals can resulting in aliasing between the long-term signal and the window function associated with the observation schedule, leading to false positives [41, 42].

1.6 The Sun is also a star! Considering the solar case

The wide range of timescales associated with stellar activity processes makes it difficult to mitigate by simply averaging observations. In order to fully disentangle activity-driven signals from real planetary signals, it is necessary to understand and model these activity-driven RV variations. While it is difficult to differentiate these two sources of RV variation on stellar targets, the Sun provides a convenient test case for developing models of activity-driven variations. The solar system planets are well characterized, with known positions, masses and radii. This allows their contributions to be removed from the observed solar RV, allowing for direct measurements of activity-driven RVs.

Using a specially-designed instrumentation, the Sun may be observed as a star. That is, spatial information is removed by averaging over the solar disk, allowing us to treat the Sun as a point source, like any other star. In addition, the wealth of additional space and ground based observato-

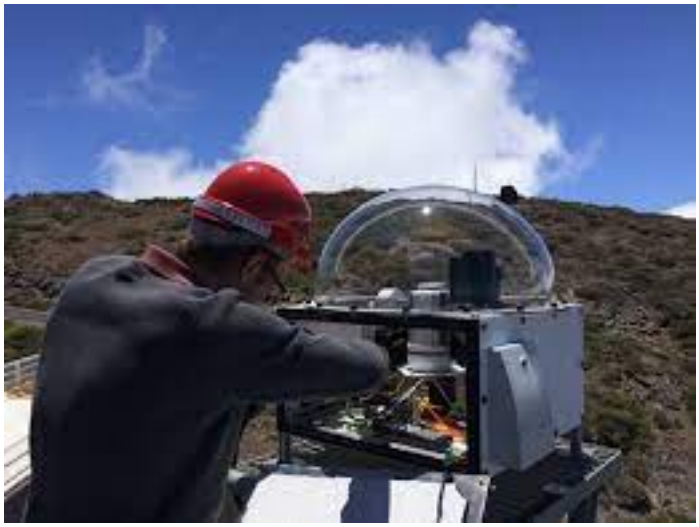

Figure 1.8: The solar telescope at HARPS-N *(Photo credit: David Phillips)*

ries allow us to incorporate additional information about the activity on the solar surface into our models. Here, we discuss three instruments that have proven most valuable in our investigation of solar magnetic processes:

1.6.1 Solar Telescope at HARPS-N

The HARPS-N spectrograph at the Galileo National Telescope (TNG) in La Palma is a cross-dispersed echelle spectrograph spanning the visible range [21]. The primary mission of HARPS-N is to provide RV follow-up to *Kepler* and *K2* candides, as well as TESS candidates visible from the northern hemisphere. However, during the day a custom-built solar telescope connected to HARPS-N provides a near-continuous stream of disk-integrated solar spectra [26], [25]. This instrument, in operation since 2015, works in combination with the HARPS-N spectrograph to observe the Sun as a star, giving unprecedented temporal coverage (about one exposure every five minutes, with a typical daily coverage of 6 hours) of the solar spectrum with resolving power R=115,000 and optical bandwidth spanning 383 nm - 690 nm. The solar telescope has a 3" lens that feeds an integrating sphere, which scrambles all angular information and converts solar images into the equivalent of a point source. Systematic laboratory and on-sky tests show the solar telescope captures the full disk of the Sun with RV precision below 10 cm $s^{-}1$ as compared to independent SDO/HMI images, well

below the 40 cm s^{-}1 per exposure precision of HARPS-N itself [43], [25].

The solar telescope takes five minute solar exposures. This exposure time is chosen to coincide with the dominant timescale associated with the solar p-modes — each exposure therefore integrates out the acoustic oscillations, mitigating their effects on our observations. The resulting RV precision is approximately 40 cm s^{-}1. This temporal coverage allows us to investigate solar activity on timescales between minutes and years, as demonstrated in [25]. To reduce the effects of solar oscillations, granulation, and other processes with variability timescales less than 24 hours, we take daily averages of the solar RVs. The changes in these daily-averaged RVs are therefore dominated by stellar activity effects [26], [7].

Solar RVs are derived from the measured spectra using the HARPS-N Data Reduction System (DRS) [44], [45]. The contributions of planetary reflex motion to the solar RVs are removed using JPL HORIZONS ephemerides [46]. The HORIZONS system provides high-precision tracking of the center-of-mass positions of the Earth, Sun, and other solar system bodies.[2]HORIZONS therefore allows us to track the Sun's position relative to the solar telescope, and thus allows us to remove the determine the Sun-TNG relative velocity from the observed radial velocities. If the Sun were a totally inactive star, the resulting velocities would be consistent with zero; the residual RV signal is the result of activity processes acting on the solar surface.

The effects of differential atmospheric extinction are removed from the RVs by calculating the intensity-weighted mean rotational velocity across the solar disk, accounting for the extinction gradient across the disk and the inclination of the solar rotation axis to the local vertical, as described in [47]. Exposures contaminated by clouds are identified using the HARPS-N exposure meter, and are removed from the final data set. If any of the 1-second sampled exposure meter measurements are below a certain threshold, the corresponding exposure is rejected. The remaining "RV residuals" are predominantly the result of solar variability: i.e., if the Sun were a uniform, homogeneous disk, they would be consistent with zero and limited to statistical noise and residual spectrograph

[2]The JPL HORIZONS On-Line Solar System Data and Ephemeris Computation Service is publically available at https://ssd.jpl.nasa.gov/horizons.cgi

systematic variations. These residuals have an RMS amplitude of 1.6 m $s^{-}1$, comparable to those observed on stars of similar activity levels [48].

Additionally, we extract the calcium S-index, a known correlate of magnetic activity and the derivative R'_{HK} [49], [50] from the Ca II H&K lines in the solar spectra. The presence of chromospheric calcium leads to strong absorption by the Fraunhofer H and K lines. However, the presence of magnetically active regions in the photosphere causes magnetic heating of the chromosphere above that active region. This magnetic heating excites the chromospheric calcium, resulting in chromospheric re-emission of photos in the core of the H and K lines. The S-index quantitatively represents this activity-driven chromospheric re-emission in the Calcium II H and K lines — the presence of spots, plage, and network all increase the S-index. Each measurement of the S-index has a precision of 2.51×10^{-4}, or a fractional uncertainty of 0.0016.

1.6.2 SORCE/TIM

The Solar Radiation and Climate Experiment (SORCE) is a stellite-based mission managed by the Laboratory for Atmospheric and Space Physics at the University of Colorado, Boulder. It was launched in 2003, and provides state-of-the-art measurements of x-ray, UV, visible, near-IR, and total solar radiation.

The Total Irradiance Monitor (TMI) onboard SORCE monitors the total solar irradiance (TSI) incident at the top of the Earth's atmosphere. These measurements are used to continue the solar climate date record, and are used to determine the sensitivity of Earth's climate to the natural effects of the changing solar irradiance. [5, 6, 51]

The total solar irradiance (TSI), however, is analogous to light curves obtained by *Kepler*, K2, TESS, and CHEOPS [12–14]. We may therefore use TIM to measure photometry for the whole solar cycle. While the solar telescope at HARPS-N is equipped with an exposure meter, variable atmospheric transparency, the aging of telescope components, and the lack of a reference source

makes ground-based photometry impractical. The space-based SORCE/TIM is therefore a valuable tool in the study of solar activity, allowing us to compare our disk-integrated and disk-resolved data products with simultaneous photometry.

On timescales between days and months, changes in the TSI are related to the movement of bright and dark active regions across the solar surface — the TSI therefore functions as a purely photometric measure of the solar activity. In particular, the observed solar irradiance increases with the presence of plage and network and decreases with the presence of spots. This modulation makes the TSI, and stellar light curves in general, good tools for isolating the effects of stellar magnetic activity [52]. The TIM level 3 data products are averaged over 6 hours, with a precision of 0.005 W/m^2.

1.6.3 SDO/HMI

The Solar Dynamics Observatory (SDO) was launched in 2010 as part of NASA's Living With a Star Program to understand the causes of solar variability and it's impacts on Earth. SDO is specifically designed to stuy the solar atmosphere on short spatial and time scales, and over many wavelengths.

The Helioseismic and Magnetic Imager (HMI) onboard SDO captures full disk images of the Sun with near single-granule resolution [53], [54]. HMI determines the continuum intensity, line depth, line width, doppler velocity, and magnetic flux at each point along the solar disk by measuring six wavelengths around the 6173.3 Å neutral iron (Fe I) line in two polarization states [55]. (Other instruments onboard SDO are used to image the Sun over different wavelengths.) Although SDO and HMI are designed to study how the solar surfaces changes over short timescales (HMI takes images every 45 seconds, and also reports integrated 12 minute exposures), these images may be use to study longer-term variations in solar activity and the solar RVs.

2

HARPS-N Solar RVs Are Dominated By Large, Bright Magnetic Regions

2.1 Introduction

In 2019, T. Milbourne et al., published "HARPS-N solar RVs are dominated by large, bright magnetic regions," (*The Astrophysical Journal*, **874**, 107 2019), demonstrating the importance of the size of active regions in understanding RV shifts. In this chapter, we reproduce their findings. Their results are presented in the following chapter.

The radial velocity (RV) method is a powerful tool for exoplanet detection and mass estimation ([56], [57]). When used in conjunction with transit measurements, determined using observations from CoRoT, *Kepler*, *K2*, and TESS ([58], [12], [13], [14]), the RV method allows for determinations of planetary densities, relevant to studies of the internal structure of detected planets [59]. The reflex RV amplitude induced by an Earth-mass planet in the habitable zone of a Sun-like star is about 10 cm s^{-1}, the target sensitivity of next-generation spectrographs [60]. However, RV measurements are currently dominated by the effects of stellar activity. In particular, acoustic oscillations, granulation and supergranulation due to surface magneto-convection, and other stellar activity processes contribute to RV signals exceeding 1 m s^{-1} as discussed by [61], [62], [7], [63], and

[64]. These activity processes must be understood to successfully interpret current observations.

Stellar activity processes act on distinct timescales. Over periods of a few minutes, stellar p-modes (that is, the propagation of acoustic vibrations) are dominant [28]. The upward and downward motion of convecting plasma also contributes to the overall RV signal; these granulation processes (including supergranulation and mesogranulation) are dominant over periods between hours and a few days ([34], [30], [32]). Contributions from magnetic features, such as dark sunspots, bright photospheric plage (i.e., the magnetically laced photosphere under chromospheric plage) and photospheric network, dominate on timescales longer than a rotation period ([7], [2]). Since exoplanet surveys often target low-activity stars, the behavior of stars near activity minimum must be considered to ensure accurate RV detections of low-mass exoplanets.

The close proximity of the Sun makes it an ideal test case for studying different stellar signals and correlating with RV measurements. Numerous ground and space-based instruments, such as Global Oscillations Network Group (GONG, [65]), the SOlar Radiation and Climate Experiment (SORCE, [51]), and the Solar Dynamics Observatory (SDO, [54]), perform detailed observations of the Sun's surface. In parallel with these instruments, a custom-built solar telescope installed at the Telescopio Nazionale Galileo (TNG) on La Palma makes disk-integrated, spectroscopic measurements of the Sun as a star using the state-of-the-art High Radial velocity Planet Searcher for the Northern Hemisphere (HARPS-N) spectrograph ([26], [25]). This allows us to observe the Sun as we would any other star in high-precision RV exoplanet surveys. By comparing this rich data set with solar photometry and disk-resolved images, we investigate the contributions of different activity processes to RV measurements, and how these contributions vary over different timescales.

Previously, [1] reconstructed the solar RV using images from the Michelson Doppler Imager (MDI) onboard the Solar and Heliospheric Observatory (SOHO). They also investigated how the convective RV shift scales with active region area. In another work, the authors investigate the relative contributions of large versus small active regions to the solar RV [7]. [2] used data from the Helioseismic and Magnetic Imager (HMI, [53]) onboard SDO to reconstruct the magnetically-driven solar RVs, and compared these to HARPS RVs derived from sunlight reflected off the asteroid Vesta. Both

these works, however, suffer from practical limitations: [1] were limited by the spatial resolution of MDI, and were unable to measure the impact of small active regions on the activity-driven RV. While [2] used higher-resolution HMI images in their work, their observations of Vesta only spanned 70 days, approximately 2.5 synodic solar rotation periods. In order to fully characterize the effects of magnetic activity on the solar RVs, we need to understand the contributions of large and small active regions, and how these contributions evolve over the course of many rotation periods.

In this work, we use contemporaneous disk-averaged solar telescope spectra, HMI solar images, and SORCE Total Irradiance Monitor (TIM) measurements of the Total Solar Irradiance (TSI) ([5], [6]) taken between July 2015 and September 2017, near the end of Solar Cycle 24, to estimate how the RV contributions from convective and photometric solar magnetic activity vary over several solar rotation periods, approaching solar minimum (late 2018/early 2019). We also investigate how these contributions vary with the size of the active regions producing these RV perturbations.

2.2 Observations

2.2.1 Solar Telescope at HARPS-N

The HARPS-N spectrograph at the TNG is a cross-dispersed echelle spectrograph spanning the visible range [21]. During the day, a custom-built solar telescope connected to HARPS-N provides a near-continuous stream of disk-integrated solar spectra ([26], [25]). This instrument, in operation since 2015, works in combination with the HARPS-N spectrograph to observe the Sun as a star, giving unprecedented temporal coverage (about one exposure every five minutes, with a typical daily coverage of 6 hours) of the solar spectrum with resolving power $R = 115,000$ and optical bandwidth spanning 383 nm - 690 nm. Systematic laboratory and on-sky tests show the solar telescope captures the full disk of the Sun with RV precision below 10 cm s^{-1} as compared to independent SDO/HMI images, well below the 40 cm s^{-1} per exposure precision of HARPS-N itself [43], [25]

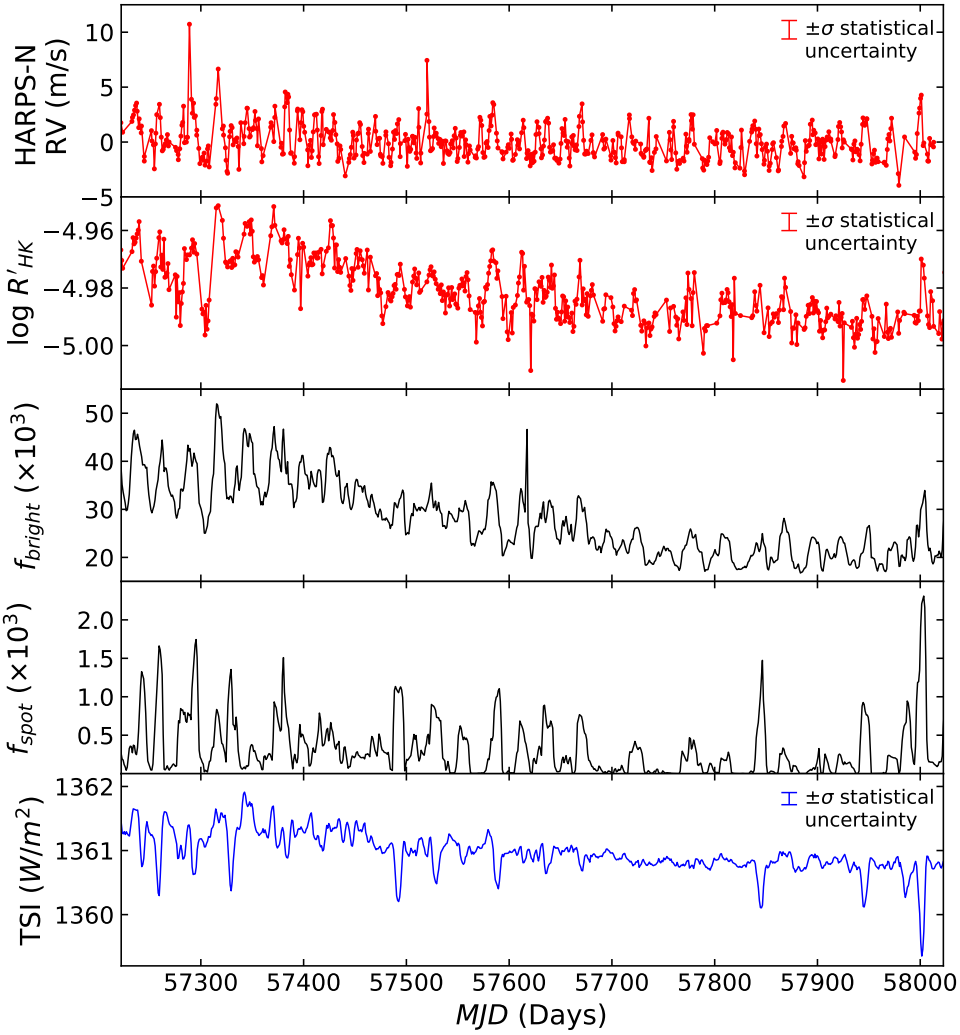

Figure 2.1: Solar measures used in this work. From top to bottom: solar telescope/HARPS-N RV after subtracting the effects due to all planets using the JPL *Horizons* System (measured relative to the averaged HARPS-N solar RV) and $R'_{\rm HK}$ (red), SDO/HMI bright (plage and network) and dark (spot) filling factors (black), and SORCE/TIM TSI (blue). A noticeable decrease in solar activity beginning around Day 200 is visible in all of the displayed activity indicators but not in the HARPS-N RVs. However, we do note an apparent decrease in the RV scatter at this time. Dips in the TSI are coincident with peaks in the spot filling factor. Observations are taken between July 2015 through September 2017, with solar minimum expected in late 2018/early 2019. For the solar telescope/HARPS-N and SORCE/TIM derived quantities, we plot a representative $\pm\sigma$ statistical error bar. Since the SDO/HMI-derived quantities are determined by averaging over $\sim 10^6$ CCD pixels, the associated statistical errors are vanishingly small. We therefore omit error bars for those quantities.

We acquire five minute solar exposures to average over solar acoustic oscillations (p-modes), and achieve RV precision of approximately 40 cm s^{-1}. This temporal coverage allows us to investigate solar activity on timescales between minutes and years, as demonstrated in [25]. To reduce the effects of solar oscillations, granulation, and other processes with variability timescales less than 24 hours, we take daily averages of the solar RVs. The changes in these daily-averaged RVs are therefore dominated by stellar activity effects ([26], [7]).

Solar RVs are derived from the measured spectra using the HARPS-N Data Reduction System (DRS) ([44], [45]). The contributions of planetary reflex motion to the solar RVs are removed using JPL Horizons ephemerides to determine the Sun-TNG relative velocity [46], and the effects of differential atmospheric extinction are removed from the RVs using the techniques described in [47]. Exposures contaminated by clouds are identified using the HARPS-N exposure meter, and are removed from the final data set. If any of the 1-second sampled exposure meter measurements are below a certain threshold, the corresponding exposure is rejected. The remaining "RV residuals" are predominantly the result of solar variability: i.e., if the Sun were a uniform, homogeneous disk, they would be consistent with zero and limited to statistical noise and residual spectrograph systematic variations. These residuals have an RMS amplitude of 1.6 m s^{-1}, comparable to those observed on stars of similar activity levels [48]. Additionally, we extract the calcium S-index, a known correlate of magnetic activity and the derivative $R'_{\rm HK}$ ([49], [50]), from the Ca II H&K lines in the solar spectra. The resulting values of the HARPS-N RVs and $\log R'_{\rm HK}$ are shown in the top two panels of Fig. 2.1.

2.2.2 SDO/HMI

HMI onboard SDO captures full disk images of the Sun with near single-granule resolution ([53], [54]). HMI determines the continuum intensity, line depth, line width, doppler velocity, and magnetic flux at each point along the solar disk by measuring six wavelengths around the 6173.3 Å neutral iron (Fe I) line in two polarization states [55].

Using thresholding algorithms pioneered by [66] and subsequently used for solar RV modelling by [1] and [2], we identify active regions along the solar disk and calculate the magnetic filling factor, f_{total}. (See Fig.2.1), the percentage of the solar disk covered by magnetic activity. We use the same intensity thresholds determined by [4] and employed by [2] to distinguish between dark regions (sunspots) and bright regions (plage and network), allowing us to calculate filling factors for each type of magnetic feature (f_{bright} and f_{spot} respectively). By combining the intensity and magnetic flux information with the Doppler velocities, we estimate the contributions of magnetic activity to solar RVs.

In this work, we consider the 720 second exposure line-of-sight measurements of the continuum intensity, magnetic field, and Doppler velocity[1]. We use six images each day, sampled evenly over the 2.5 year operational period of the solar telescope at HARPS-N. Note that all HMI observables have a strong 24-hour modulation related to an imperfect removal of the SDO spacecraft's orbit [55]; to mitigate the effects of these and other systematics ([67], [68], [69]) we therefore reference all derived RVs to the quiet-sun velocity and take daily averages of the derived filling factors and activity-driven RVs ([1], [2]). (See Sec. 2.4 and Appendix A for further discussion of these calculations.)

2.2.3 SORCE/TIM

TIM onboard SORCE measures the TSI using a set of four Electrical Substitution Radiometers, providing a near-continuous stream of solar photometry analogous to *Kepler* data [6] for the Sun[2]. On timescales between days and months, changes in the TSI are related to the movement of bright and dark active regions across the solar surface (as shown in Fig. 2.1). The TSI therefore functions as a purely photometric measure of the solar activity. While the solar telescope at HARPS-N is equipped with an exposure meter, variable atmospheric transparency, the aging of telescope components, and the lack of a reference source makes ground-based photometry impractical. The space-based SORCE/TIM is therefore a valuable tool in the study of solar activity, allowing us to compare our disk-integrated and disk-resolved data products with simultaneous photometry.

[1]Publicly available at http://jsoc.stanford.edu/
[2]Publicly available at http://lasp.colorado.edu/home/sorce/data/tsi-data/

2.3 Comparing Measurements of Solar Magnetic Activity

The solar telescope, SDO/HMI, and SORCE/TIM each provide a unique lens for analyzing solar activity. Using the broadband, spectroscopic information derived from the solar telescope/HARPS-N, we extract the solar RVs, Mt. Wilson S-index and the derivative index $\log R'_{HK}$ (see below for a further discussion of these activity indicators). SDO/HMI directly images active regions on the solar disk, allowing us to identify them as sunspots or plage and network. SORCE/TIM measurements of the TSI provide a photometric measurement of solar activity.

Comparing the time series of each activity indicator, shown in Fig. 2.1, we see qualitative agreement between the data products of each instrument. The HARPS-N-derived $\log R'_{\rm HK}$, the SDO/HMI bright region filling factor (plage and network), the peak amplitudes of the SDO/HMI spot filling factor, and the SORCE/TIM TSI all show the same downward trend as the Sun approaches solar minimum. Furthermore, peaks in the spot filling factor are coincident with sharp dips in the SORCE/TIM TSI. In this section, we make quantitative comparisons between these independent measurements of solar activity and demonstrate the instruments provide a consistent picture of solar magnetic processes.

2.3.1 Comparison of SDO/HMI with Solar Telescope/HARPS-N

Magnetic heating of the solar chromosphere results in enhanced emission reversals in the cores of the Ca II H&K lines [70]. The observed correlation between these emission reversals and sunspot number, as described in [71], led to the development of the Mt. Wilson S-index and the color-corrected $\log R'_{HK}$, as defined in [49]. Given the correlations between chromospheric plage and the photospheric spots and faculae, we expect a high degree of correlation between $\log R'_{HK}$ and the magnetic filling factors as well [72].

The SDO/HMI-derived magnetic filling factor and spectroscopic measurements of $\log R'_{HK}$ are highly correlated, with a Spearman correlation coefficient of 0.8836. (Fig. 2.2). Dividing the to-

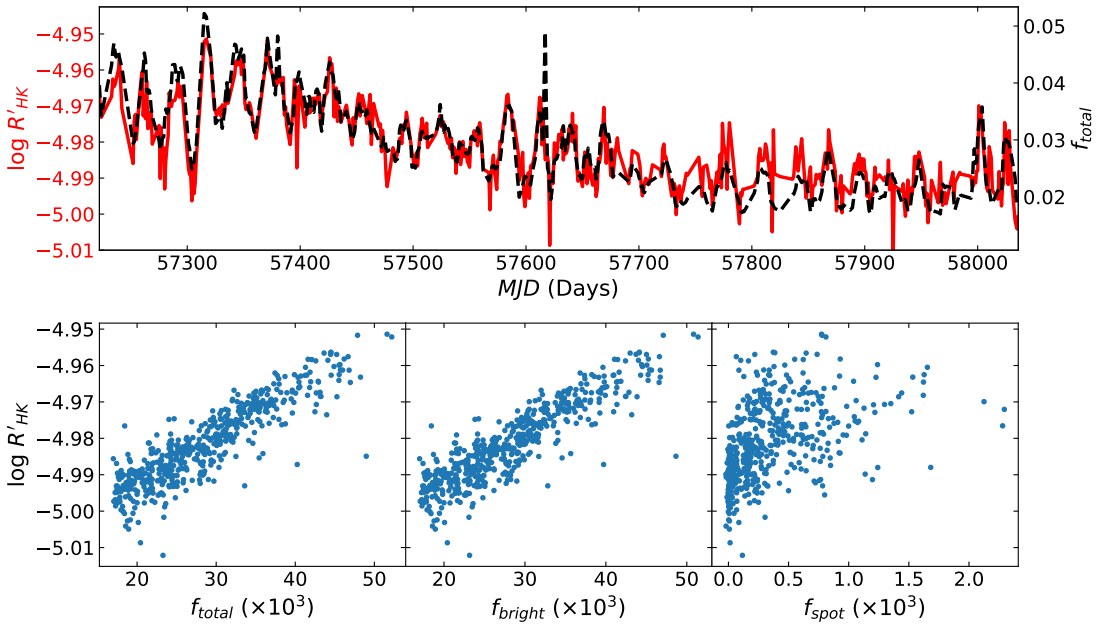

Figure 2.2: *Top:* Spectrally derived $\log R'_{HK}$ (black dotted line) and SDO/HMI-calculated total magnetic filling factor (red solid line), plotted as a function of time. A strong correlation between the two quantities is clearly visible in the time series. Both indicators demonstrate oscillations at the synodic solar rotation period (28 days). *Bottom:* Correlation plots between $\log R'_{HK}$ and the total filling factor (left), the network and plage filling factor (center), and the spot filling factor (right). We see that the correlation between between the filling factor and $\log R'_{HK}$ is driven by the bright regions: the Sun is a plage-dominated star entering activity minimum, resulting in a factor of $\sim 10^2$ fewer sunspots, and a much weaker correlation with the spot filling factor. This relationship is captured by the Spearman correlation coefficients for each filling factor and $\log R'_{HK}$: the correlation coefficient between the overall filling factor and $\log R'_{HK}$ is 0.8836, the correlation coefficient between the plage/network filling factor and $\log R'_{HK}$ is 0.8833, and the correlation coefficient between the spot filling factor and $\log R'_{HK}$ is 0.590.

tal magnetic filling factor into a bright (plage and network) and dark (spot) contributions also shows the expected behavior; the plage and network, which dominate the total activity, are strongly correlated with $\log R'_{HK}$. The spots, however, cover a much smaller portion of the solar surface ($f_{bright}/f_{spot} > 80$) and exhibit a much weaker correlation with $\log R'_{HK}$.

2.3.2 Comparison of SDO/HMI with SORCE/TIM TSI

The presence of bright (plage and network) and dark (sunspots) features on the solar surface causes the TSI to fluctuate in response to magnetic activity. This is readily apparent in Fig. 2.1: spikes in the spot filling factor derived from SDO/HMI are accompanied by corresponding decreases in the SORCE/TIM TSI and the long-term decrease in the plage and network filling factor are correlated with the long-term decrease of the TSI.

The different brightnesses of these features arise because bright plage and network regions are hotter than the quiet Sun, and that spots are colder. We therefore define ΔT_{bright} and ΔT_{spot}, the brightness temperature contrasts of these two features. Following [7], we use the SDO/HMI derived plage and spot filling factors to reproduce the measured TSI:

$$TSI = \mathcal{A}\sigma[(1 - a_{spot}f_{spot} - a_{bright}f_{bright})T_{quiet}^4 + a_{spot}f_{spot}(T_{quiet} + \Delta T_{spot})^4 + a_{bright}f_{bright}(T_{quiet} + \Delta T_{bright})^4] \quad (2.1)$$

where σ is the Stefan-Boltzmann constant, $\mathcal{A} = (R_\odot/1AU)^2$ is a geometrical constant relating the energy emitted at the solar surface to the energy received at Earth, T_{quiet} is the quiet Sun temperature, and f_{spot} and f_{bright} are the HMI spot and plage/network filling factors.

Additionally, a_{bright} and a_{spot} are scaling factors, used to account for systematic differences in the calculation of filling factors. The values of the bright and spot filling factors depend strongly the choice of magnetic flux and intensity thresholds used to differentiate spots, plage, and quiet sun, as well as the wavelength(s) used to observe these features. Variations in these parameters mean that established sunspot datasets may differ by over 50% [7]. Including these scaling factors in our model allows us to account for these definition-dependent factors and to compare the brightness

temperature contrasts of each feature to literature values. Since these scaling factors are constant, multiplicative values, they do not affect correlations between the filling factors and the other activity measurements.

A wide range of spot and plage/network temperature contrasts are given in the literature. From [7], we infer that -649 K $< \Delta \mathrm{T}_{\mathrm{spot}} < -450$ K and 38 K $< \Delta \mathrm{T}_{\mathrm{bright}} < 55$ K. Note that the apparent temperature of plage varies with position on the solar disk: Since the disk-averaged SDO/HMI plage filling factor contains no spatial information, we take ΔT_{bright} to be the average brightness temperature contrast of solar plage. In our analysis, we assume $\Delta T_{spot} = -550$K and $\Delta T_{bright} = 46.5$K, corresponding to the midpoints of the above ranges.

Assuming the above values of ΔT_{spot} and ΔT_{bright} and using the SDO/HMI-derived time series of f_{spot} and f_{bright}, we fit Eq. 2.1 to the SORCE/TIM TSI as shown in Fig. 2.3. From this fit, we extract a quiet-Sun temperature ($T_{quiet} = 5770.080 \pm 0.007$K) and scaling coefficients ($a_{bright} = 0.959 \pm 0.007$ and $a_{spot} = 1.54 \pm 0.02$). These coefficients are consistent with the 20% and 50% definitional variations in feature area described in [7].

The three-way agreement between the solar telescope $\log R'_{\mathrm{HK}}$, SDO/HMI filling factors, and SORCE/TIM TSI indicates that our activity models provide a consistent picture of solar magnetic processes. However, the solar telescope/HARPS-N solar RVs are not in full agreement with these activity measurements. In particular, Fig. 2.1 shows that $\log R'_{\mathrm{HK}}$, f_{bright}, and the TSI all display a downward trend over the 800 day observation period. The solar RVs do not display this trend. To quantify this disagreement, we compute the Spearman correlation coefficient between the solar RVs and $\log R'_{\mathrm{HK}}$, f_{bright}, and TSI, yielding values of 0.42, 0.40, and 0.04 respectively. To understand this discrepancy, we now use SDO/HMI-derived RVs to reproduce solar telescope/HARPS-N measurements.

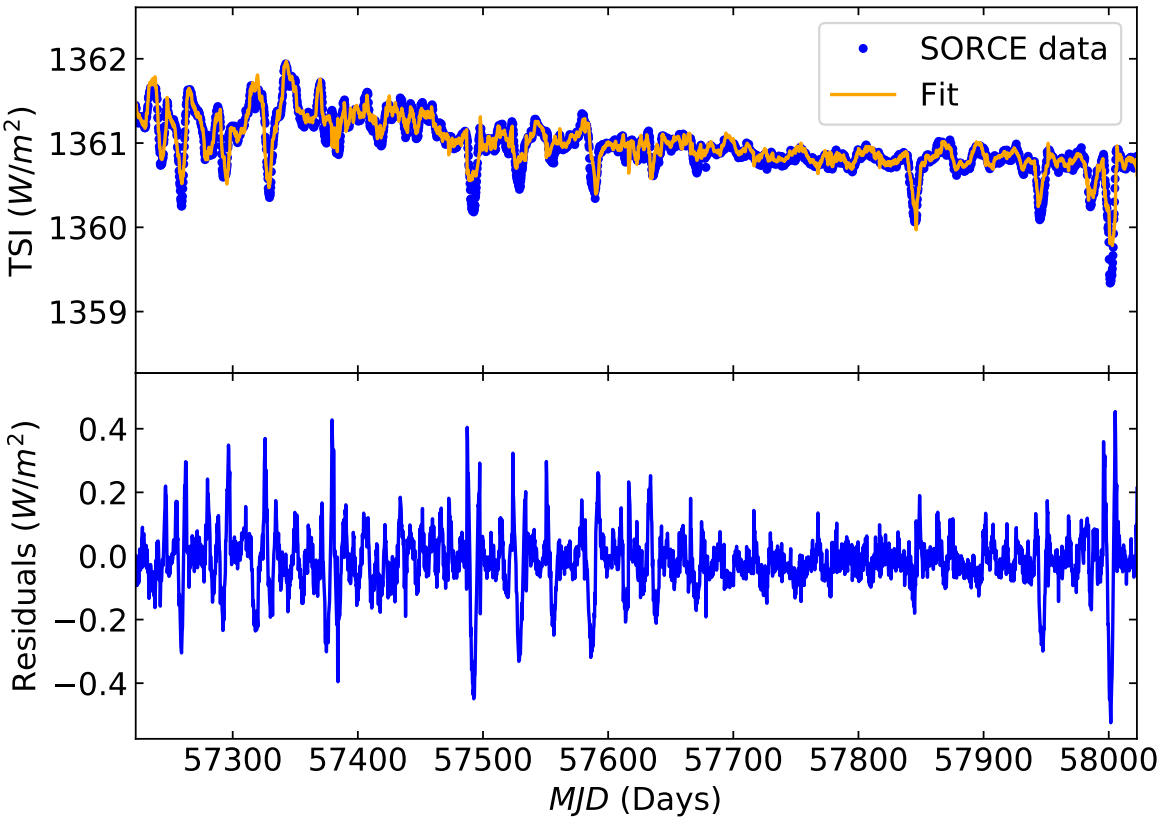

Figure 2.3: *Top*: SORCE/TIM measurements of TSI (blue circles), with reconstructed TSI from SDO/HMI filling factors (orange line). *Bottom*: Fit residuals. The residuals shown are consistent with the typical SORCE/TIM uncertainty per data point, 0.48 W m^{-2}. Note that the correlation between the TSI and the convective magnetic shift (see Sec. 2.4.1) implies an RV scaling with TSI of 3.3 (m s^{-1}) / (W m^{-2}).

2.4 Calculating RV Contributions of Spots/plage

We model the effects of stellar magnetic activity on RV measurements as a combination of two processes: the suppression of convection in magnetically active regions that leads to a net redshift of the spectrum ($\Delta\hat{v}_{\rm conv}$), and the effect of bright and dark active regions on the solar disk that leads to a photometric shift ($\Delta\hat{v}_{\rm phot}$). In the following sections, we discuss the physical origins of each term and their magnitudes as derived from SDO/HMI images. We then reconstruct the solar RVs from a combination of the two processes and fit this model to the RVs measured with the solar telescope/HARPS-N.

2.4.1 Suppression of Convective Blueshift, $\Delta\hat{v}_{\rm conv}$

$\Delta\hat{v}_{\rm conv}$ results from the suppression of solar convective motions by local magnetic fields. Taking an intensity-weighted average of the bright, upflowing plasma in the middle of the convective cells and dark, downflowing plasma at the cell edges results in an overall convective blueshift with an amplitude of approximately 250 m s^{-1} ([73], [7]). The plasma's interaction with solar magnetic fields impedes this convective motion and therefore attenuates this convective blueshift. Note that the convective blueshift of an observed spectral line depends on its formation depth in the photosphere ([74], [75]). The convective shift observed by SDO/HMI using the 6173.3 Å line will therefore differ from the solar telescope/HARPS-N observations, which are averaged over many lines. To account for this systematic difference, we apply a scaling coefficient in our RV reconstruction as discussed in Sec. 2.4.3.

In previous studies, [1], [76], and [2] found $\Delta\hat{v}_{\rm conv}$ to be the dominant source of RV shifts, with a disk integrated amplitude of several m s^{-1}. Using the SDO/HMI dopplergrams in conjunction with the magnetic flux and continuum intensity images, we replicate the analysis of [2] to determine $\Delta\hat{v}_{\rm conv}$ for the full solar telescope/HARPS-N observing period. Several m s^{-1} variations are observed at the synodic rotation period of the Sun along with long-term drifts of a similar amplitude as shown in the upper-left panel of Fig. 2.4.

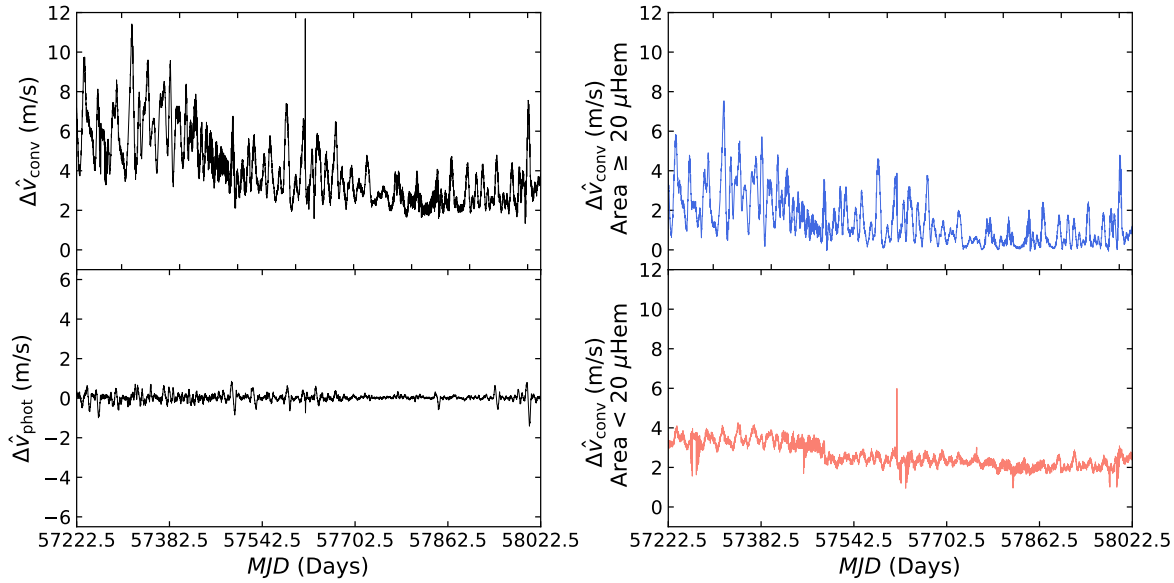

Figure 2.4: *Left:* SDO/HMI-derived estimates of the convective (*top*) and photometric velocities (*bottom*) using all active regions. *Right:* Contributions of plage (area ≥ 20 μHem, *top*) and network (area < 20 μHem, *bottom*) to the suppression of convective blueshift. As in previous works ([1], [2]), we find that $\Delta\hat{v}_{\rm conv}$ dominates the effects of $\Delta\hat{v}_{\rm phot}$.

2.4.2 Photometric Shift, $\Delta\hat{v}_{\rm phot}$

The presence of dark sunspots and bright plage on the solar disk break the Sun's symmetry about its rotation axis. This results in an imbalanced Doppler shift across the solar disk; it is this Doppler imbalance that results in $\Delta\hat{v}_{\rm phot}$, the photometric RV shift due to magnetic activity ([61], [77]). As before, we use the methods of [2] to compute $\Delta\hat{v}_{\rm phot}$ using the SDO/HMI-measured full-disk magnetograms and continuum intensity. This time series is also shown in the lower-left panel of Fig. 2.4 and, as expected, is significantly smaller than the shifts calculated for $\Delta\hat{v}_{\rm conv}$.

2.4.3 Reconstruction of Solar RVs From SDO/HMI Basis Functions

Following [2], we model the total solar telescope RVs, $\Delta{\rm RV}_{\rm model}$, using a linear combination of $\Delta\hat{v}_{\rm conv}$ and $\Delta\hat{v}_{\rm phot}$:

$$\Delta{\rm RV}_{\rm model} = A(t)\Delta\hat{v}_{\rm phot} + B(t)\Delta\hat{v}_{\rm conv} + {\rm RV}_0. \tag{2.2}$$

Here $A(t)$ and $B(t)$ are weighting coefficients for the photometric and convective RV shifts, and RV_0 describes the zero point of HARPS-N. As RV_0 is a purely instrumental parameter, we expect it to remain constant with time. $A(t)$ and $B(t)$ describe the mapping of information from the single $\lambda = 6173.3$ Å spectral line onto the several thousand lines used in the HARPS-N CCF analysis ([44], [45]). The coefficient $A(t)$ accounts for systematic differences between the bright and dark active regions observed with SDO/HMI and the spectrum observed with the solar telescope/HARPS-N, analogous to the scaling factors used in our TSI reconstruction (see Eq. 2.1). The coefficient $B(t)$ accounts for the systematic difference in the convective blueshift due to the different heights of formation of each spectral line. We thus expect $A(t)$ and $B(t)$ to be of order unity, but not necessarily equal to 1. In [2], $A(t)$, $B(t)$, and RV_0 are taken to be constant. However, these coefficients could vary with time perhaps due to additional magnetic processes at work, or some other changes over the activity cycle. We divide each time series into smaller subsections and calculate the fits for each subsection to investigate how $A(t)$ and $B(t)$ evolve in time.

As discussed in Sec. 2.2, we take daily averages of 2.5 years of solar telescope data to mitigate the effects of solar p-modes and granulation. We expect that, on timescales longer than several days, the measured RV variations are dominated by magnetic effects. We then fit Eq. 2.2 to the whole data set, yielding global values of $A(t)$ and $B(t)$, as given in Table 2.2. Following [2], we include an uncorrelated noise parameter s, added in quadrature to the solar telescope/HARPS-N observational errors, to account for instrumental uncertainties and other processes not in our model [78].

We then divide the data into subsections of $N = 112$ days (corresponding to four synodic solar rotations per subsection) and repeat the fit, evaluating $A(t)$, $B(t)$, and RV_0 for each subsection. We chose this value of N to maximize the number of data sections while maintaining sufficiently small statistical uncertainties. The results described below do not depend on the exact value of N.

Since the RV contributions of magnetic active regions are modulated by the Sun's rotation (see Figs. 2.4 and 2.6), we expect our model to fully capture RV variation on timescales of the rotation period (and its harmonics) and above. As shown in Fig. 2.5, the dominant contributions to the observed RV variations occur on these timescales. Below the rotation period, the solar RV is modulated to

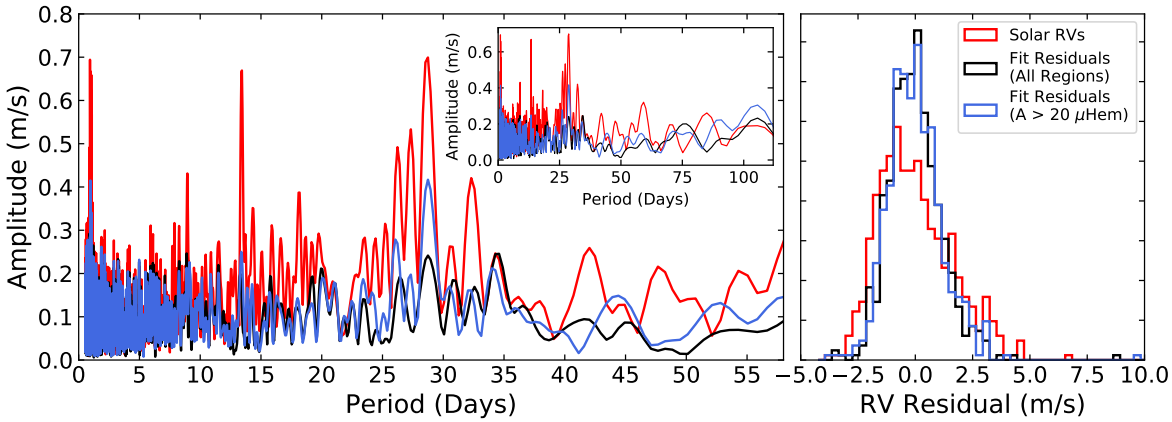

Figure 2.5: *Left:* Periodogram of the solar telescope RVs (red), fit residuals using all active regions (black), and fit residuals using a 20 μHem area threshold (blue). We note that measured solar RVs have an amplitude of 0.72 m s^{-1} at the solar rotation period. Applying our model with no area cut reduces this amplitude to 0.24 m s^{-1}; including an area cut results in an amplitude of 0.42 m s^{-1}. *Inset:* A zoomed-out view of the periodgram. We note that the two fits successfully reduce the RV amplitudes observed on most timescales greater than the rotation period. *Right:* Histogram of the RV residuals. Both fits result in Gaussian-distributed RV residuals: while both fits display decrease the RMS RV residuals, applying an area threshold does not produce a visible change in the fit residuals. The area cut does, however, remove the unphysical trend in RV_0, as discussed in Sec. 2.5.2 and as shown in Fig. 2.7

.

some degree by magnetic region growth and decay, but also by granulation and supergranulation due to surface magneto-convection. Our model is not designed to capture these convective processes, and we therefore do not expect it to capture all RV variations on these short timescales.

2.4.4 Active Region Area Dependence of Convective Shift

In addition to reconstructing the solar telescope/HARPS-N RVs variations we investigate if and how the suppression of convective blueshift associated with a given active region depends on its size. [7] speculated that small intergranular network features and large plage/sunspot regions would have different contributions to the convective blueshift, and [79] ([79]; [80]), observed different center-to-limb velocity variations for solar network and plage. We differentiate the RV contributions of these two classes of active region. The network and plage/spot regions may be distinguished based on their spatial distributions: while small network are uniformly distributed over the solar disk, large plage/spot regions appear only around active latitudes, leading to the well-known butterfly diagram

Basis function	All Active Regions	Area $\geq 20\mu$Hem	[2]	[1]
$\Delta\hat{v}_{phot}$	0.21 m s^{-1}	0.21 m s^{-1}	0.17 m s^{-1}	0.42 m s^{-1}
$\Delta\hat{v}_{conv}$	1.69 m s^{-1}	0.88 m s^{-1}	1.30 m s^{-1}	1.39 m s^{-1}
HARPS-N RV	1.74 m s^{-1}			

Table 2.1: RMS amplitudes of RV time series. We include the time series derived using all regions (left column of Fig. 2.4) and using only plage regions (right column of Fig. 2.4). As a point of comparison, we also include the values of [2] (also derived from SDO/HMI), the values of [1] (derived from the Michelson Doppler Imager onboard the Solar and Heliospheric Observatory), and the solar telescope measurements of the solar RVs (top panel of Fig. 2.1).

(see [81] and references there-in).

To distinguish network from plage/spot regions, we plot the 2D distribution of active region co-latitude Θ and area, as shown in the left panel of Fig. 2.6. There is a sharp cut at approximately 20 micro-hemispheres (that is, 20 parts per million of the visible hemisphere), or 60 Mm^2 in areal units. Active regions smaller than this cutoff are distributed across the Sun, while regions above the cutoff only appear around the equator, at $0.75 < \sin\Theta \leq 1$. We therefore use this area threshold to classify each active region as small network or large spot/plage.

To investigate the differing contributions of the network and spot/plage, we compute $\Delta\hat{v}_{\rm conv}$ as a function of time using only network and only spot/plage regions. From the resulting time-series (see Fig. 2.4) and periodograms (right panel of Fig. 2.6), we observe that the majority of the RV variability at the solar rotation period is the result of large active regions: small regions do not significantly contribute to the suppression of convective blueshift on this timescale. Given these differing contributions, we perform the RV reconstruction of Sec. 2.4.3 first using the convective RV shift calculated using all observed active regions; second using the convective RV shift calculated using only large spots/plage. The results of this analysis are given in Fig. 2.7 and Tables 2.1 and 2.2.

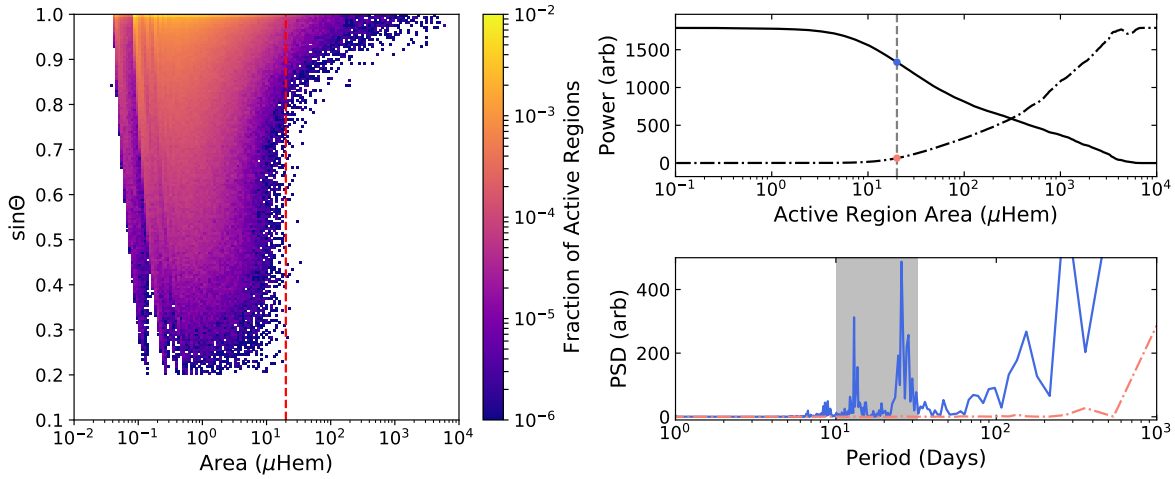

Figure 2.6: *Left*: Fraction of observed solar active regions as a function of region area and co-latitude, Θ (as measured from the north pole). From the observed spatial distribution, we may divide the active regions into small network, which appear all across the solar disk, and large spots/plage, which preferentially appear around activity latitudes. The sharp change in the spatial distribution allows us to infer the presence of a sharp cut-off allows us to infer an area threshold of 20 micro-hemispheres that separates these two regimes. *Upper-right*: Power associated with the solar rotation period and its first harmonic above (solid) and below the area threshold (dashed). Power at these frequencies is evaluated by integrating the power spectral density (PSD) of the RV contributions for each region size over the shaded region, indicated below. *Lower-right*: PSD of RV contributions above (solid blue) and below (red dashed) micro-hemisphere. Below the 20 micro-hemisphere threshold, there is little power associated with the solar rotation period: these small structures therefore do not contribute to the solar RVs on the timescales of interest in this work.

Parameter	Basis function	All Active Regions	Area $\geq 20\mu$Hem	[2]
$A(t)$	$\Delta\hat{v}_{\rm phot}$	2.24 ± 0.60	1.09 ± 0.58	2.45 ± 2.02
$B(t)$	$\Delta\hat{v}_{\rm conv}$	0.93 ± 0.11	1.20 ± 0.15	1.85 ± 0.27
RV_0		102.51 ± 0.06 m s^{-1}	102.36 ± 0.13 m s^{-1}	99.80 ± 0.28 m s^{-1}
s		1.21 m s^{-1}	1.23 m s^{-1}	2.70 m s^{-1}

Table 2.2: Average values of the SDO/HMI-derived $\Delta{\rm RV}_{\rm model}$ to solar telescope/HARPS-N RVs using Eq. 2.2. (*See text.*) We provide values derived using both network and plage regions, replicating the analysis of [2], as well also values derived using only the plage regions. The time variation of these parameters is shown in Fig. 2.7. We also include the results of [2] as a point of comparison. Error bars on each parameter are statistical uncertainties and s is the added white noise beyond the 40 cm s^{-1} noise associated with each solar telescope observation.

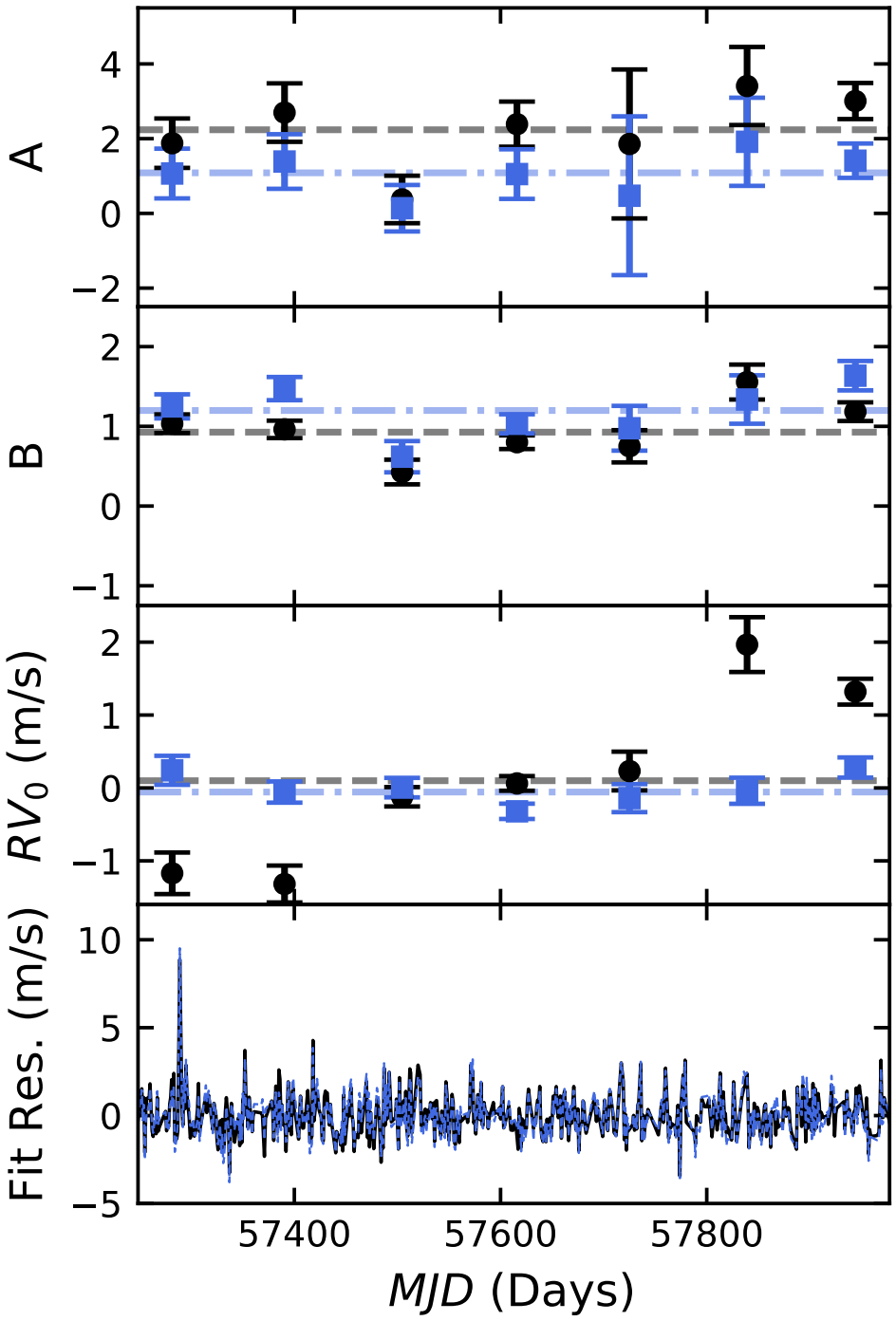

Figure 2.7: *First and second panel*: Fit parameters $A(t)$, $B(t)$, for the unitless scaling parameters for the photometric and convective RV shifts, $\Delta\hat{v}_{\rm phot}$, and $\Delta\hat{v}_{\rm conv}$ derived from SDO/HMI (plotted in Fig. 2.4). *Third panel*: RV offset ($RV_0(t)$) in m s^{-1}. Parameters fitted to HARPS-N solar RVs using Eq. 2.2 (*see text*) in $N = 112$ day sets, using all active regions (black points) or a 20 μHem area threshold (blue squares). Statistical error bars are plotted for each parameter. *Fourth panel*: Residuals of both fits.

2.5 Discussion

2.5.1 Reconstruction of Solar RVs

As shown in Fig. 2.7, both $A(t)$ and $B(t)$ are consistent with constant values, implying that $\Delta\hat{v}_{\rm conv}$ and $\Delta\hat{v}_{\rm phot}$ have a consistent effect on the solar telescope/HARPS-N measurements over the full observation period. The resulting $A(t)$ and $B(t)$ values are consistent with those reported by [2], as shown in Table 2.2. [2] measured the Sun close to solar maximum, observing a total magnetic filling factor $6\% < f_{total} < 10\%$, whereas near solar minimum, we observe $f_{total} < 5\%$. We also note that the RMS amplitudes of $\Delta\hat{v}_{\rm phot}$ and $\Delta\hat{v}_{\rm conv}$ (shown in Table 2.1), calculated using network and plage regions, are consistent with or somewhat smaller than those of [1] and [2], which is consistent with observations performed at different parts of the activity cycle. Given the agreement of our $A(t)$ and $B(t)$ values with those of [2], we may conclude that these parameters do not change significantly as the Sun enters activity minimum, and only weakly depend on the magnetic filling factor, if they do so at all.

Using only large plage/spot regions in our reconstruction of $\Delta\hat{v}_{\rm conv}$ does not significantly change the magnitude of $A(t)$ and $B(t)$ compared to using all active regions. However, when all active regions are considered, the calculated instrumental offset $RV_0(t)$ (see Fig. 2.7) slowly increases over the three year observation period. This slow increase disappears when our model assumes that only large active regions suppress the convective blueshift. We discuss the implications of this result in Sec. 2.5.2.

2.5.2 Long-Timescale Variations: Changes in RV_0

As our model assumes that magnetic activity is fully described by $\Delta\hat{v}_{\rm conv}$ and $\Delta\hat{v}_{\rm phot}$, we expect RV_0 to be an instrument-dependent parameter related to the zero point of HARPS-N, and therefore constant over our observation period. HARPS-N exposures are calibrated using a simultaneous reference with sub-m $\rm s^{-1}$ precision [43], and the SDO/HMI basis functions are calculated rela-

tive to the quiet-Sun velocity[2]; long-term instrumental drifts are therefore calibrated out of each measurement, and should not affect the value of RV_0.

However, in both fits to the solar telescope data, we find a systematic increase in RV_0. When all active regions are considered, we obtain a shift of $\Delta RV_0 = 2.6$ m s^{-1} over the course of the 800 day measurement period. Accounting for the area dependence of the convective velocity eliminates this variation almost entirely. This is consistent with our hypothesis regarding the area dependence of the convective velocity: small active regions do not meaningfully contribute to the suppression of convective blueshift on timescales of the solar rotation period. Instead, as shown in Fig. 2.4 and the right panel of Fig. 2.6, these small regions contribute a systematic drift on timescales of hundreds of days. Differentiating the contributions of small and large magnetic activity is therefore necessary for the successful detection of long-period, low mass planets around solar-type stars.

Physically, this systematic RV shift may be due to the different contributions of plage and network, as demonstrated in [79] ([79]; [80]). Part of the active solar network results from decaying of plage regions, resulting in a correlation between the plage filling factor and the network RV contribution. This, in turn may lead to an overall systematic RV shift. We may also consider a similar scenario for dark spots: Small, dark solar pores lack penumbra, and therefore have a different contribution to the solar RV. As the Sun enters activity minimum, large spots become less common, leading to a larger relative pore contribution, and therefore a systematic RV shift. In both cases, large and small active regions must therefore be treated separately in our RV reconstruction.

2.5.3 RV Residuals and Rotational Modulated Variations

The measured solar telescope/HARPS-N RV variations (Fig. 2.1, top) have an RMS scatter of 1.65 m s^{-1}. Subtracting the reconstructed values of $\Delta\mathrm{RV}_{\mathrm{model}}$ computed using all active regions and using the physically-motivated constant value of RV_0 (assumed to be an instrumental offset) reduces this scatter to 1.31 m s^{-1}. Repeating this analysis with an empirical time-varying value of RV_0 reduces this scatter to 1.18 m s^{-1}. Incorporating the area dependence of the convective shift

into our model results in an RMS scatter of 1.21 m s^{-1}. While including this spatial information does not further improve the RMS scatter of the RV residuals, it almost completely eliminates the observed change in RV_0, and thus constitutes a more physically grounded and complete model, as previously discussed in Sec. 2.5.2.

As discussed at the end of Sec. 2.4.3 and as shown in Fig. 2.6, we expect our model to eliminate the observed power at the solar rotation period. The measured solar telescope RV variations have an amplitude of 0.72 $\pm$ 0.06 m s^{-1} at this timescale: applying our model using all active regions reduces this amplitude to 0.24 $\pm$ 0.08 m s^{-1}. Incorporating the area dependence of the convective shift into our model results in an amplitude of 0.42 $\pm$ 0.08 m s^{-1} (see Fig. 2.5). The residual signal at the rotation period may indicate that our area threshold does not perfectly differentiate plage and network regions: smaller plage regions are not being included in our calculated value of $\Delta\hat{v}_{\rm conv}$, resulting in an imperfect removal of the RV signal at the rotation period. Reducing the area threshold reduces the residual amplitude at the rotation period: however, the inclusion of network regions in the calculated $\Delta\hat{v}_{\rm conv}$ results in a non-zero trend in RV_0. The residual signal may also result from the several-day lag between activity proxies and RV signals observed by [76] and [47]. Lastly, we note that our model assumes the network regions have the same RV contribution as the quiet Sun: in reality, however, we expect network to provide an additional, nontrivial RV contribution ([79]; [80]). A more sophisticated model will be required to fully describe these network-driven variations.

In summary, the reconstructed RVs leave over 1 m s^{-1} of RV variations unaccounted for: these may be the result of supergranulation, which has a physical timescale longer than the 6-8 hour solar observation period at the TNG, ([31], [64]). They may also result from additional surface velocity flows unaccounted for in our model, RV differences of network relative to quiet Sun, or an unaccounted for instrumental systematic.[3]

[3]The HARPS/HARPS-N DRS was recently upgraded to improve the stability of the daily wavelength calibrations. At the time of writing, the HARPS-N solar data had not yet been reprocessed with the new DRS. We understand that the older version of the software has an uncertainty of up to 1 m s^{-1} in the RV zero points between successive days of observation.

2.5.4 Magnetic Activity Indicators and Active Region Area

The agreement between the magnetic filling factors, $R'_{\rm HK}$, and TSI demonstrated in Sec. 2.3 confirms that the traditional metrics for solar activity are all self-consistent. However the correspondence between these activity indicators is not improved by separately considering large and small active regions: the chromospheric emission captured by $\log R'_{\rm HK}$ is strongly correlated with the total magnetic filling factor, not the large region filling factor. Similarly, the TSI may only be accurately reproduced using both large and small active regions. All the magnetic active regions on the Sun have an enhanced chromospheric column density that strengthens the emission reversals in the Ca II H&K line cores, and all bright/dark regions will modulate the Sun's overall brightness. These traditional stellar activity indicators are thus correlated only with the overall coverage of active features, and not the size of each feature. Given the observed dependence of the suppression of convective blueshift on active region size, we may therefore conclude new activity indicators correlated with active region sizes are needed to successfully reproduce RV variations on distant stars.

2.6 Conclusions

In this work, we analyze 3 years of solar observations during the decline of Carrington Cycle 24 to test models of radial-velocity variations of Sun-like stars. We compare solar telescope/HARPS-N measurements of the solar RVs and $\log R'_{\rm HK}$, SDO/HMI disk-resolved activity images, and SORCE/TIM measurements of the total solar irradiance. As expected, the observed values of $\log R'_{\rm HK}$ and TSI are strongly correlated with the overall magnetic filling factor derived from SDO/HMI images.

However, these activity indicators are not straight-forward predictors of the observed solar RV variations. While we see a slow decrease in $\log R'_{\rm HK}$, TSI, and magnetic filling factor as the Sun enters cycle minimum, we do not observe this decrease in the solar telescope/HARPS-N RV variations. To investigate this discrepancy, we model the solar RV as a linear combination of the suppression

of convective blueshift and rotational flux imbalance. Our initial reconstruction of the solar RV variations decreased the RMS scatter from 1.65 m s^{-1} to 1.18 m s^{-1} and reduced the RV amplitude at the rotation period by a factor of 4, but only by introducing an arbitrary systematic drift of 2.6 m s^{-1} over the 800 day observation period. By computing contribution of each active region to the suppression of convective blueshift, we find that active regions smaller than 20 ppm (60 Mm^2) do not significantly suppress the convective blueshift. Including this area dependence in our model does not further decrease the overall RMS scatter, and results in a factor of 2 reduction of the RV amplitude at the rotation period. However, it completely eliminates the need to introduce an arbitrary systematic drift in our reconstructed RVs, resulting in a more physically-grounded model. We propose two possible causes for this drift: small changes in the network coverage which affect our quiet Sun reference velocity due to RV differences between network and the quiet Sun ([79]; [80]), or RV differences between spots (with penumbrae) and pores (without penumbrae), which are modulated by the changing spot filling factor. In either scenario, more detailed studies of the RV contributions of large and small scale features will be required to elucidate the mechanisms involved.

The different contributions of plage and network to the activity-driven RV variations explains why the calcium H/K activity index does not systematically correlate strongly with RV variations in Sun-like stars on timescales comparable to the magnetic cycle. On highly active stars, where large plage regions dramatically outnumber the small network regions, the plage filling factor will be approximately equivalent to the overall filling factor. We therefore expect the traditional activity indicators, such as $\log R'_{\rm HK}$ and optical light curves, to provide a useful proxy for activity-driven RV variations in this regime. On low-activity stars, where the plage and network filling factors are comparable, separating the contributions of plage and network will be necessary to reproduce activity driven RV variations. As the traditional activity indicators are correlated with overall filling factor, they will not provide as useful a proxy of the activity-driven RV variations. For exoplanet RV surveys to be successful for low-activity stars, we must therefore identify correlates for activity region size.

The residuals of our fit still have an RMS spread of over 1.21 m s^{-1}. This additional scatter may be the result of some long-term granulation process [82], additional surface velocity flows, additional

magnetic effects of network ([79]; [80]), or unaccounted for systematic variation in the spectrograph on timescales shorter than the solar rotational period. Determining the physical origin of these residual RV variations, identifying correlates for active region size, and verifying that the observed relationships between RV and active region size hold as the Sun enters the active phase of the magnetic cycle will be the subject of future investigations.

3

Crossing Timescales: The Effects of Active Region Size on Acoustic Oscillations

3.1 Introduction

In the previous chapter, we demonstrated that large and small active regions have different contributions to the suppression of convective blueshift. In order to properly model activity-driven RV variations, it is therefore necessary to have information about the size of active regions in addition to their overall coverage. Most activity indicators do not distinquish between large plage and small network, making it difficult to ascertain the effect of different sized active regions on stellar targets. In order to successfully model activity-driven RV variations for long-period exoplanets, we must therefore identify an activity indicator able to determine the contributions of different active region sizes.

However, it is known that magnetic activity drives drives frequency shifts in the solar p-modes, and that the observed p-mode frequenct therefore varies over timescales of the 11-year solar activity cycle [83–85]. In this work, we investigate the effects of strong, large-scale active regions versus weaker, small-scale ephemeral regions on the observed p-mode shifts, and determine which of these two classes of active regions dominates the oberserved frequency shifts.

3.2 ER vs. AR regions

Here, we consider the magnetic flux from strong-field, active-region (AR) and weak-field, ephemeral-region (ER). These regions are observed by the Wilcox Solar Observatory (WSO, [86]), and are seperated by a 15 G threshold. (Note that this is lower than the $B_{thres} = 3\sigma = 24$ G threshold used in our studies of HMI magnetograms.) These AR and ER may regions may be considered to be qualitatively equivalent to the plage and network regions discussed in Chapters 2 and 4. This anaology will be explored more in Sec. 3.4.

The WSO observations span Solar Cycles 21-24, providing a nearly 40 year baseline for this analysis. These WSO values are compared with comptemporaneous results from higher-resultion HMI magnetographs spanning Solar Cycle 24. These magnetic flux values are calculated using the following equation:

$$F(t) = \frac{\sum_{i,j} I_{ij} B_{ij} / A_{ij}}{\sum_{i,j} I_{ij}} \tag{3.1}$$

Where I_{ij}, B_{ij}, and A_{ij} are the continuum intensity, magnetic field strength, and area associated with the i,j-th pixel. We model the AR flux, $F_{AR}(t)$, using the magnetic flux associated with plage regions: That is, we sum over pixels with magnetic field $B_{thres} > 3\sigma = 24$ G which are located in active regions with area greater than 20 microhemispheres. The flux from ER, $F_{ER}(t)$ is modelled using the network regions: we sum over pixels with magnetic field $B_{thres} \geq 3\sigma = 24$ G, in regions with area greater than 20 microhemispheres. The variations in the resulting AR and ER fluxes show very good agreement with thse determined from WSO observations, as shown in Fig. 3.1.

Note that the finite resolution of WSO means that only a fraction of the "true" ER flux is captured in the magnetograms. We therefore weight the observed ER flux by a factor of β^{-1} — that is, we assume WSO captures a factor of β of the total ER flux. Here, we assume $\beta = 0.4$ (that is, we

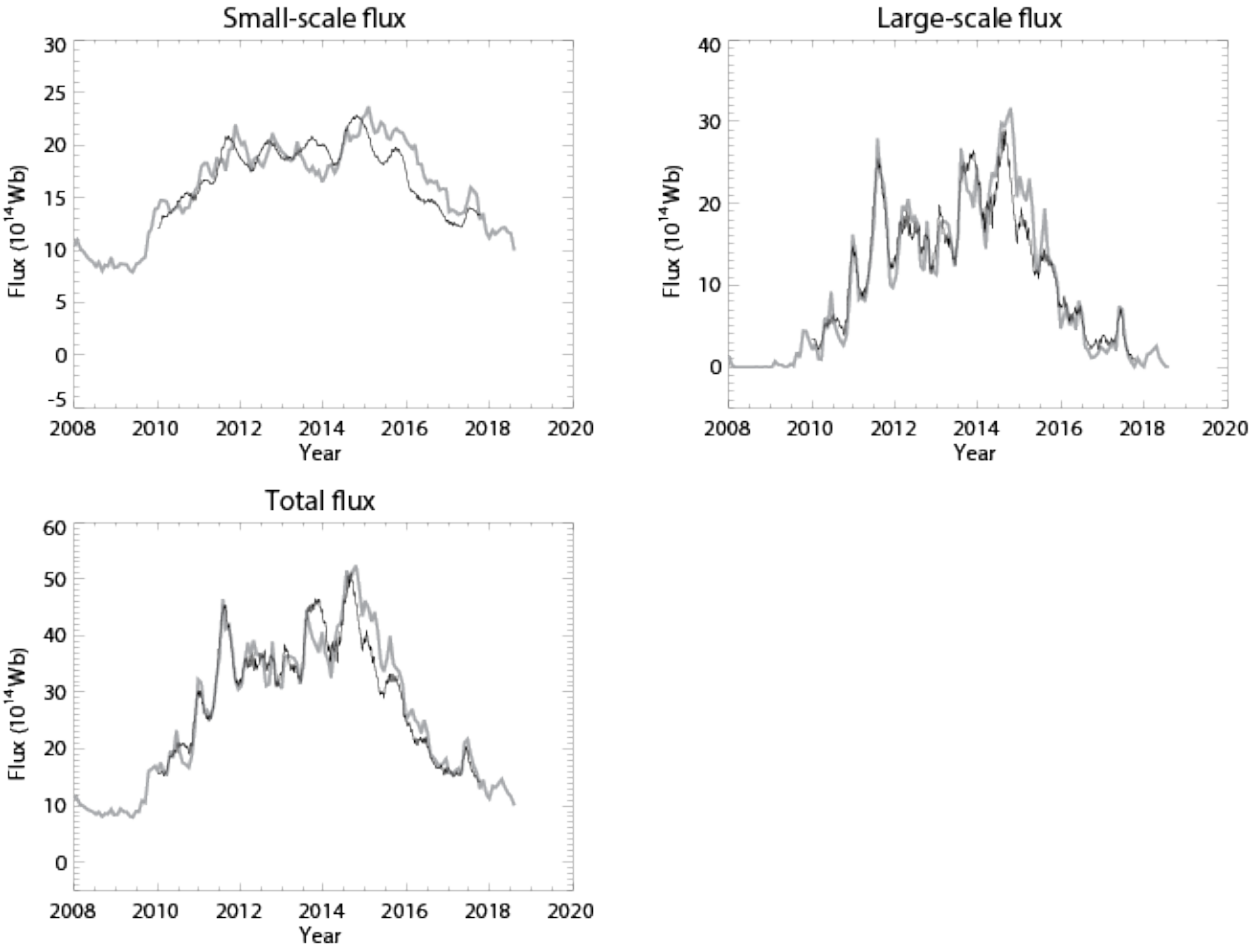

Figure 3.1: Total magnetic flux, along with contributions from small-scale (ER/network) and large-scale(AR, plage) regions, as determined from WSO and HMI observations. The WSO measurements of ER, AR, and total flux are shown in gray; the HMI measurements of network, plage, and total flux are shown in black. Note that the HMI measured small-scale values have been rescaled to match the absolute levels of the WSO values, indicating a mismatch between the WSO and HMI thresholds used to identify and differentiate active regions — however, the variations within each timeseries are consistent, indicating that both analyses are measuring the same contributions.

assume WSO only 40% of the true ER flux, and 60% of the ER flux is not detected due to the instrument's finite resolution.) [87, 88]. The total masured flux, $F_{net,m}(t)$, and the total inferred flux (accounting for the finite resolution of WSO), $F_{net,i}(t)$, are given by the below equations.

$$F_{net,m}(t) = F_{AR}(t) + F_{ER}(t) \tag{3.2}$$

$$F_{net,i}(t) = F_{AR}(t) + \frac{F_{ER}(t)}{\beta} \tag{3.3}$$

3.3 Modelling Solar p-modes

The WSO-derived AR and ER fluxes are then fit to frequency shifts in the p-modes, as observed by Birmingham Solar Oscillations Network (BiSON, [28, 89]). The resulting p-mode shifts are modeled as a linear combination of the AR and ER fluxes:

$$\delta\nu(t) = c_0 + c_1 \left[F_{AR}(t) + \alpha \left(\frac{F_{ER}(t)}{\beta} \right) \right] \tag{3.4}$$

Here α describes the relative contribution of the ER flux versus the AR flux, and the factor of β accounts for the fact that resolutions limits of WSO mean it does not capture the full ER flux, as described above.

The resulting measured (and inferred, in the case of ER) fluxes, as well as the observed and modelled p-mode shifts are shown in Fig. 3.2.

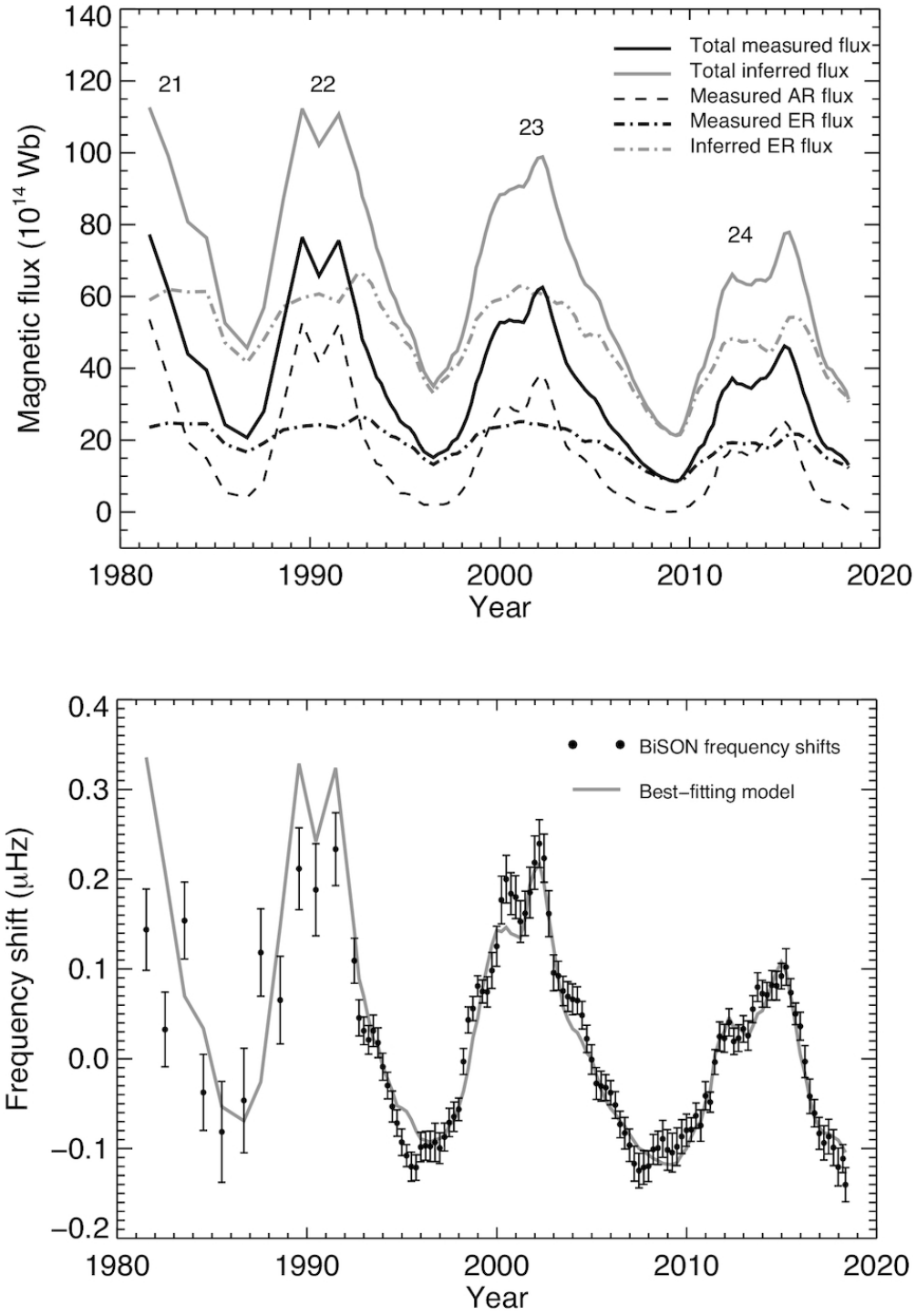

Figure 3.2: *Top panel*: Measured AR (black dashed line) and ER (black dot–dashed line) fluxes, and total measured flux (black solid line). The grey dot–dashed line shows the inferred total ER flux, i.e. having taken into account that not all the ER flux is captured by the observations. The inferred total flux (the sum of the measured AR and inferred ER) is shown in solid grey. Also shown are the numbers of each activity cycle. *Bottom panel*: BiSON mean frequency shifts (points with error bars) and modelled frequency shifts (solid line) given by using the best-fitting coefficients. *(Figure credit: [3])*

3.4 Results and Extensions to Studies of Network and Plage

Fitting this model gives $\alpha = 0.11 \pm 0.09(\text{stat}) \pm 0.02(\text{sys})$, indicating that low-order p-modes are indeed significantly less sensitive to magnetic flux from the ER component of magnetic flux (that is, the magnetic flux produced by network regions) than from the AR component (the flux produced by the plage regions). Using this model, the authors may therefore estimate the p-mode frequency shifts that would be expected for a Sun-like star at the Maunder minimum, a "grand minimum" of solar activity between 1645 and 1715, during which sunspots were extremely rare [90]. The seismic behavoir of the Sun at this grand minimum approximates the behavoir of magnetically quiet stars, making this limit of interest for future astroseismology studies.

This result also shows that it is, in principal, possible to use disk-integrated activity indicies to distinquish the effects of differet classes of active regions. Here, we demonstrate that strong-field, large scale AR and smaller, weak-field, ER have different effects on the observed p-mode frequency. As shown in Fig. 3.1, the ER and AR fluxes measured by WSO are strongly correlated with the plage and network fluxes measured by HMI, indicating that they capture the same physical contributions. Difference in the absolute scaling of these fluxes is likely the result of thresholds used to distinquish active regions on each instrument: On HMI, a 24 G magnetic field threshold used to identify active regions, and the 20 μHem area thresholded used to distinquish plage and network regions. WSO uses a 15 G threshold to distinquish AR from ER. To see if a more quantative correspondence between the AR and ER regions from this work with the SDO/HMI-measured plage and network features may be established, we plot the average magnetic field strength and size of active regions (with $|B| > 8$ G) from 100 randomly-sampled HMI images in Fig. 3.3.

While we do note that, while magnetic flux does indeed increase with active region size, there is a great deal of scatter around this relationship: HMI observes small scale active regions with a wide range of field strengths. Also note the 15 G magnetic field threshold used to differentiate AR and ER in this work is within 2σ of the HMI magnetogram shot noise limit of 8 G — it is therefore difficult to ascribe quantitative signficance to a mapping between the WSO observations of AR, ER,

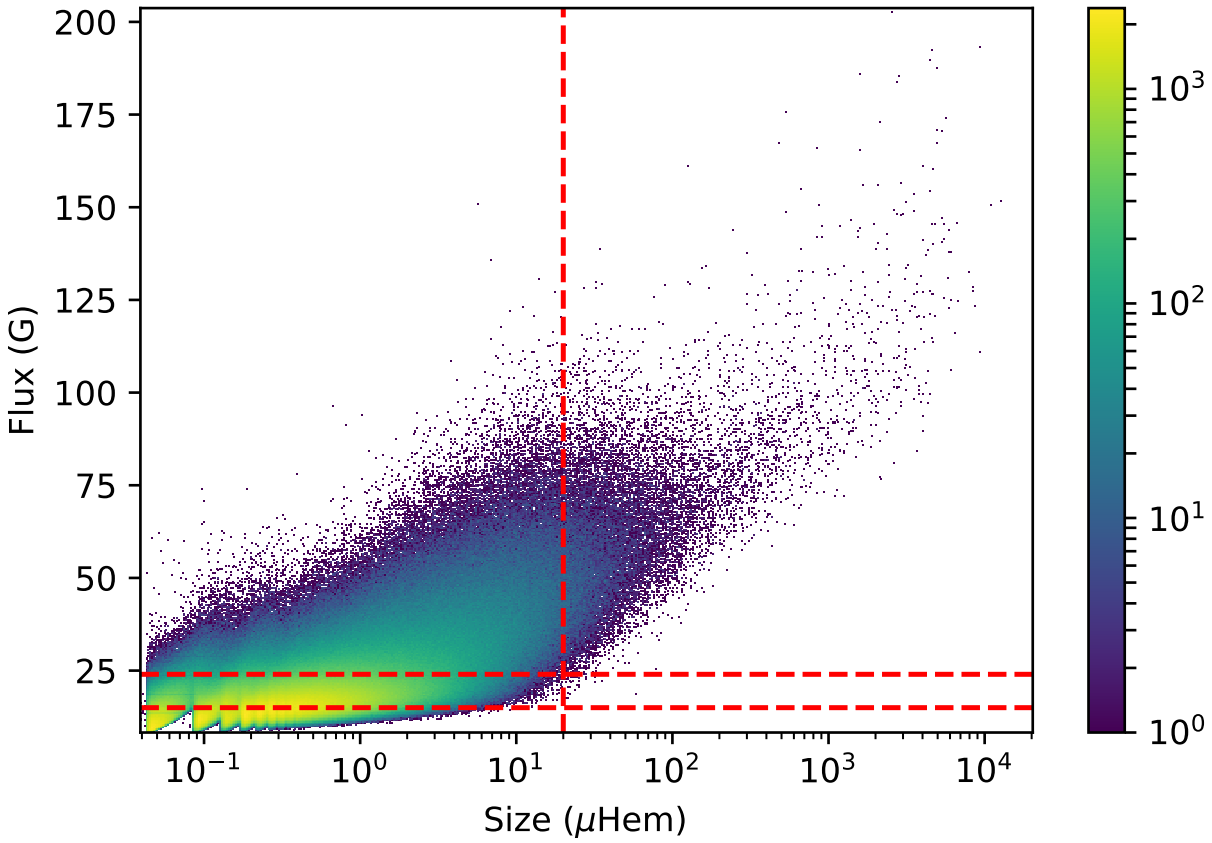

Figure 3.3: 2D histogram of magnetic field and size of solar active regions (with$|B| > 8$ G) from 100 randomly selected HMI images. The colorbar represents the number of active regions with a given size/flux bin. The dotted red lines show the area and magnetic field cuts used to identify and classify HMI active regions.

plage, and network.

In spite of these difficultities, the close qualitative agreement between the WSO and HMI measurements shown in Fig. 3.1 does appear to indicate that network and plage regions have different contributions to variations in the p-mode frequency. However, successfully measuring these $\sim 0.1\mu$Hz variations on stellar targets will require short-cadence ($<$5 minute) measurements taken over a star's activity cycle: while this is certainly possible with dedicated observations, most exoplanet host stars are observed less frequency, over shroter baselines. In the next chapter, we use a combination of photometric and spectroscopic measurements to distinquish the effects of large plage and small network active regions.

4

Estimating Feature Specific Magnetic Filling Factors From Simultaneous Spectroscopy and Photometry

4.1 Introduction

In 2021, T. Milboure et al., submitted "Estimating magnetic filling factors from simultaneous spectroscopy and photometry: disentangling spots, plage, and network," to The Astrophysics Journal (*preprint arXiv:2105.09113*, 2021). In this work, the authors develop two implementation of a technique to model RV the effects of specific active regions using only photometric and spectroscopic observations. Their results are presented in the following chapter.

State of the art radial velocity (RV) searches for low-mass, long-period exoplanets are limited by the effects of stellar magnetic activity. An Earth-mass planet in the habitable zone of a Sun-like star has an RV amplitude on the order of 10 cm s^{-1}. However, stellar activity processes on host stars, such as acoustic oscillations, magnetoconvection, suppression of convective blueshift, and long-term activity cycles, can produce signals with amplitudes exceeding 1 m s^{-1}. A variety of techniques exist to mitigate the effect of these processes on the measured RVs: [29] discuss optimal exposure times to average out acoustic oscillations; [35] and [91] present strategies for mitigaing the effects of

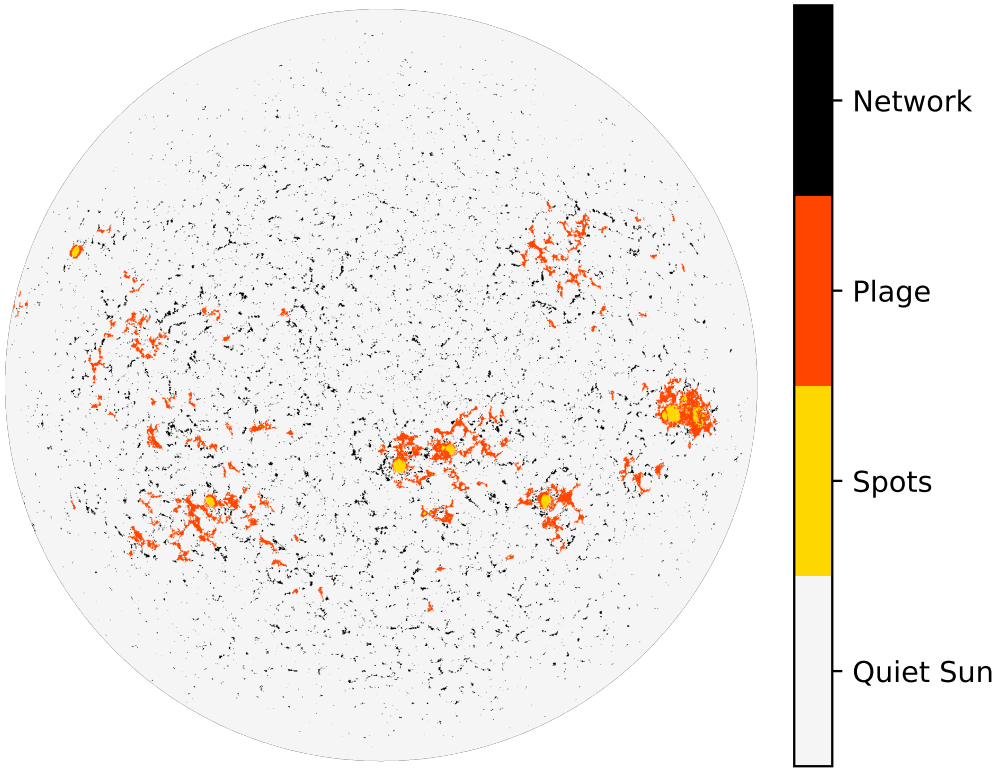

Figure 4.1: A representative HMI map of the three classes of active regions considered in this work. Spots, plage, and network are identified using the thresholding algorithm described by [2] and Chap. 2, with the threshold values given by [4]. This algorithm is briefly recapped in Sec. 4.2.3. Image taken January 1st, 2015 at 0:0:0.00 UT.

granulation; and [40], [39], [37], [92], and numerous others discuss statistically and physically-driven techniques for removing the effects of large-scale magnetic regions from RV measurements.

On timescales of the stellar rotation period, the apparent radial velocity is modulated by three main types of active regions: dark sunspots; large, bright plage; and small, bright network regions. These different regions may be identified using full-disk solar images, as shown in Fig. 4.1. [93] (hereafter referred to as Chap. 2) found that the large-scale photospheric plage contribute differently to the solar suppression of convective blueshift than the smaller network. Failure to account for this different contribution leads to a significant RV shift over the 800 day span their of observations. Some of the long-term variation reported in Chap. 2 may also be attributed to instrumental systematics: re-reducing the HARPS-N solar data with the ESPRESSO DRS [94, 95] reduces this shift from 2.6 ± 0.3 m s^{-1} to 1.6 ± 0.5 m s^{-1}. However, the remaining RV shift can only be fully removed by properly accounting for network regions in the calculated activity-driven RVs. While this analysis is possible on the Sun using high-resolution full disk images, traditional spectroscopic activity indicators, such as the Mt. Wilson S-Index [70, 71] and the derivative index $\log(R'_{HK})$ [49, 50] do not differentiate between large and small active regions. A new activity index or combination of activity indices is therefore necessary to successfully model the suppression of convective blueshift on stellar targets.

In this work, we demonstrate a new technique using simultaneous spectroscopy and photometry to estimate spot, plage, and network filling factors, and demonstrate that these filling factors may be used to model RV variations. In Section 2, we discuss the solar data used by our technique. An analytical implementation of the technique is described in Section 3, and a neural network implementation is presented in Section 4. The resulting solar filling factors, a model of the solar RVs, and possible applications to stellar targets are analyzed in Section 5.

4.2 Measurements

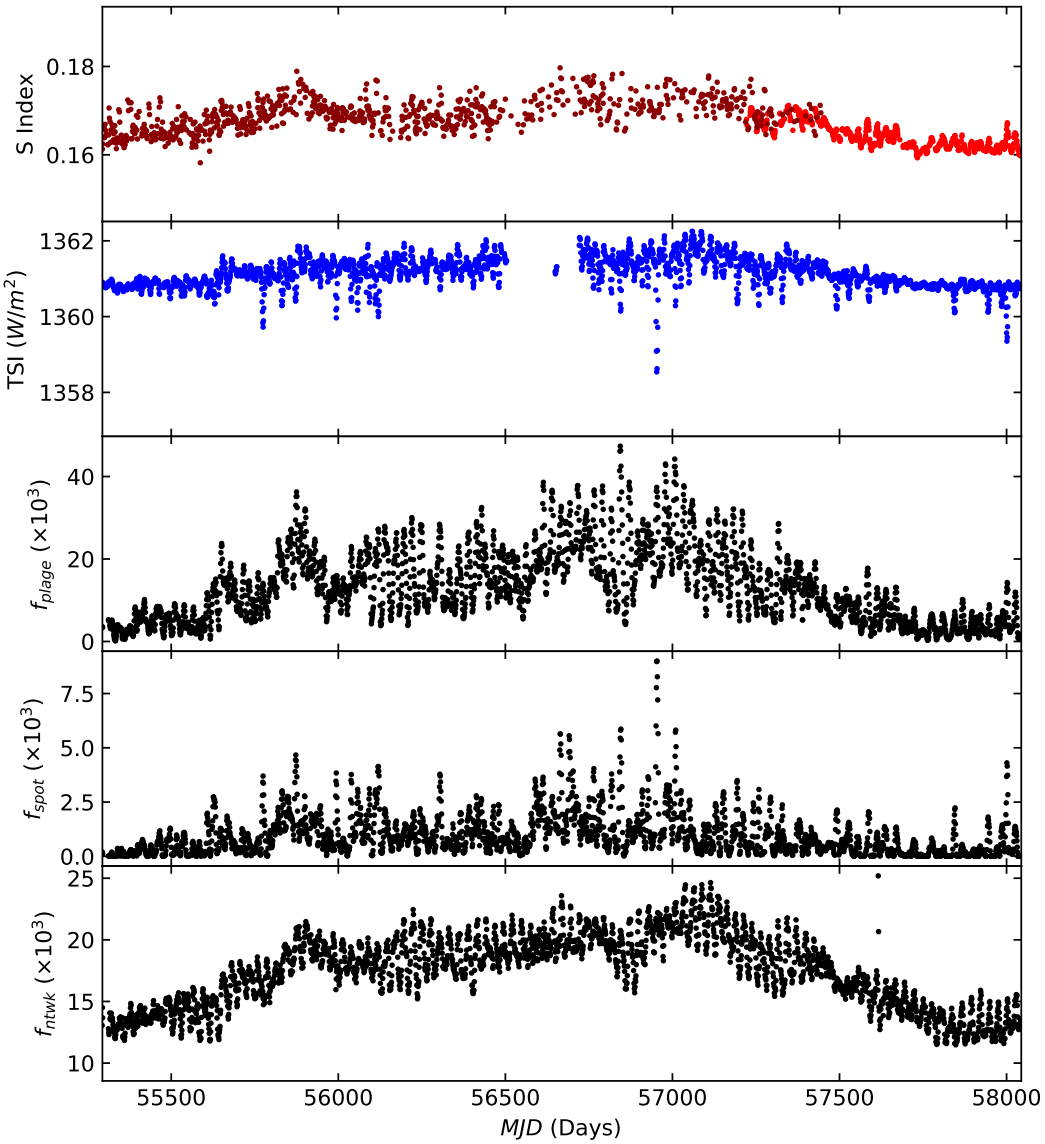

Figure 4.2: Time series of solar observations used in this work. From top to bottom: Mt. Wilson and HARPS-N solar telescope observations of the calcium S-index (red); Total Solar Irradiance (TSI) from SORCE/TIM (blue); and SDO/HMI plage, spot, and network filling factors (black). Note that the consistent overall shapes of the S-index, TSI, and bright filling factors, and that dips in the TSI are coincident with peaks in the spot filling factor. Observations are taken between April 2010 through October 2017. Note the two different reds used in the S-index plot: the darker red points correspond to measurements by Mt. Wilson, and bright red points are from the HARPS-N solar telescope. Note that due to an instrumental anomaly, no TSI data is availble from SORCE/TIM from mid July 2013 until March 2014 - we therefore do not use any times in this period in our analysis [5, 6]

4.2.1 HARPS-N/Mt. Wilson Survey

We use the HARPS-N solar telescope [25, 96] measurements of the S-index, as described in Chap. 2 and [47]. The S-index quantitatively represents activity-driven chromospheric re-emission in the Calcium II H and K lines. The presence of spots, plage, and network all increase the S-index. The solar telescope takes exposures every five minutes while the Sun is visible. Each measurement of the S-index has an average precision of 2.5×10^{-4}, or a fractional uncertainty of 0.0016.

Note that HARPS-N solar telescope observations began in mid 2015. To cover the rest of the solar cycle, we use data from the Mt. Wilson S-index survey, as presented by [97]. Observations from the two instruments overlap between July 2015 and February 2016 (JD 2457222 and JD 2457444), allowing us to combine these time series. The solar telescope dataset is rescaled so that the points in the overlapping time interval have the same mean and variance as the Mt. Wilson data from the same time interval, as described in [92]. The resulting combined dataset is shown in the top panel of Fig. 4.2.

4.2.1.1 HARPS-N Solar RVs

We use the HARPS-N solar telescope's measurements of the solar RVs to assess our ability to model realistic RV variations using our estimated filling factors. [1] [25, 47, 95], as well as our estimated values derived from the linear and MLP techniques. The HARPS-N RVs used in this work span the period from July 2015 to October 2017, with exposures taken every five minutes while the Sun is visible. Each RV measurement has an average precision of 23 cm s^{-1}.

4.2.2 SORCE

We use the Total Irradiance Monitor (TIM) onboard the Solar Radiation and Climate Experiment (SORCE) [5, 6, 51] to measure photometry for the whole solar cycle. The total solar irradiance

[1] We use publicly available HARPS-N solar telescope observations reduced using the most recent ESPRESSO pipeline, available at https://dace.unige.ch/sun/

(TSI) is the solar analogue of the light curves obtained by *Kepler*, K2, TESS, and CHEOPS [12**?** –14], though the Sun's proximity means it can be observed continuously over much longer periods. The TIM level 3 data products are averaged over 6 hours, with a precision of 0.005 W/m^2.

We expect the overall brightness of the Sun to vary with the stellar cycle. Its relative brightness increases with the presence of plage and network, and decreases with the presence of spots. This modulation makes the TSI, and stellar light curves in general, useful tools for isolating the effects of stellar magnetic activity [40]. The time series of the TSI is shown in the second panel of Fig. 4.2.

4.2.3 SDO

We use images from the Helioseismic and Magnetic Imager (HMI) instrument onboard the Solar Dynamics Observatory (SDO, [53–55]) to independently calculate solar filling factors. HMI measures the 6173.3 Å iron line at six points in wavelength space using two polarizations. From these measurements, they reconstruct the Doppler shift and magnetic field strength along with the continuum intensity, line width, and line depth at each point on the solar disk.

Spots, plage, and network are identified on HMI images using a simple threshold algorithm:

- An HMI pixel is considered magnetically active if the radial component of the magnetic field is over three times greater than the expected noise floor: $|B_r| > 3\sigma_{B_r}$.
- Active pixels below the intensity threshold of [4], such that $I < 0.89 I_{quiet}$, are labelled as spots. Here, I_{quiet} is the average intensity of inactive pixels on a given image.
- Active regions exceeding the above intensity threshold that span an area > 20 micro-hemispheres (that is, 20 parts per million of the visible hemisphere), or 60 Mm^2, are labelled as plage.
- Active regions exceeding the intensity threshold that span an area < 20 micro-hemispheres are labelled as network.

These calculations are explained in further detail in Chap. 2 and Appendix A.

The resulting filling factors for each feature are plotted in the bottom three panels of Fig. 4.2. We use one HMI image taken every four hours in our analysis. The photon noise at disk center for the magnetograms and continuum intensity for these HMI images are $\sigma_{B_r} = 8$ G and $\sigma_{I_c} = 0.01\%$ respectively [55]. This corresponds to uncertainties $< 0.1\%$ in the resulting magnetic filling factors.

Since SDO/HMI allows us to perform precise direct, independent measurements of the three filling factors of interest, we use these results as the "ground truth" in our analysis.

An SDO analogue does not exist for non-solar stars, so if we wish to determine feature-specific filling factors of stellar targets we must make indirect estimates of the filling factors using spectroscopic and photometric data. In the next section, we discuss two processes to do so. To mitigate the effects of acoustic oscillations, granulation, and other short-timescale activity process, we take daily averages of each set of observations used in our analysis. We also interpolate the HARPS-N/Mt. Wilson observations, SORCE/TIM observations, and SDO/HMI filling factors onto a common time grid of one observation each day, when all three instruments have measurements.

4.3 Linear Technique

4.3.1 Modelling Irradiance Variations Using Filling Factors

In Chap. 2, we reproduce the observed TSI using a linear combination of the spot and plage filling factors. Following [7], they use the SDO/HMI derived plage and spot filling factors, and assume that the solar irradiance follows the Stefan-Boltzmann law for blackbodies:

$$TSI = \mathcal{A}\sigma \left[(1 - f_{spot} - f_{bright})T_{quiet}^4 + f_{spot}(T_{quiet} + \Delta T_{spot})^4 + f_{bright}(T_{quiet} + \Delta T_{bright})^4\right] \quad (4.1)$$

where σ is the Stefan-Boltzmann constant, $\mathcal{A} = (R_\odot/1\text{ AU})^2$ is a geometrical constant relating the energy emitted at the solar surface to the energy received at Earth, T_{quiet} is the quiet Sun temperature, ΔT_{spot} and ΔT_{bright} are the effective temperature contrasts of spots and plage/network regions, and f_{spot} and f_{bright} are the HMI spot and plage/network filling factors. Expanding as a power series yields the following approximation:

$$TSI \approx \mathcal{A}\sigma T_{quiet}^4 \left(1 + 4\frac{\Delta T_{spot}}{T_{quiet}} f_{spot} + 4\frac{\Delta T_{bright}}{T_{quiet}} f_{bright}\right) \tag{4.2}$$

In Chap. 2, we show that HMI observations of filling factors may be used to reproduce SORCE TSI given temperature contrasts for plage/network features and spots and the effective temperature of the quiet Sun. In this work, we invert the process and use the resulting effective temperatures of each type of active region, along with the correlations between the TSI, S index, and filling factors demonstrated in Chap. 2, to reproduce the observed magnetic filling factors for each type of active region. The potential stellar applications of this technique are discussed in more detail in Sec. 4.5.2.

We begin by fitting the SORCE TSI to Eq. 4.2 using the SDO/HMI measurements of f_{spot} and f_{bright}. This fit yields $T_{quiet} = 5769.85 \pm 0.01$ K, $\Delta T_{spot} = -525 \pm 8$ K, and $\Delta T_{spot} = 46.1 \pm 0.5$ K. Also note that the solar radius varies as a function of wavelength. To be consistent with HMI, we use $R_\odot = 695982 \pm 13$ km, the solar radius measured at 6173.3 Å [98].

4.3.2 Differentiating Bright and Dark Regions

The brightness of a Sun-like star may be modulated by the long-term stellar activity cycle (*e.g.*, the 11-year solar cycle). Since $\Delta T_{bright} > 0$ and $\Delta T_{spot} < 0$, we see from Eq. 4.2 that TSI variations (on timescales of the rotation period) below the current value from the activity cycle must be the result of spots. Since the Sun is plage dominated [99], plage and network are the primary source of variation of the TSI and S-index, with spots making negligible contributions to the variability of the irradiance on timescales of the solar cycle. This is also visible in comparing the plots of the

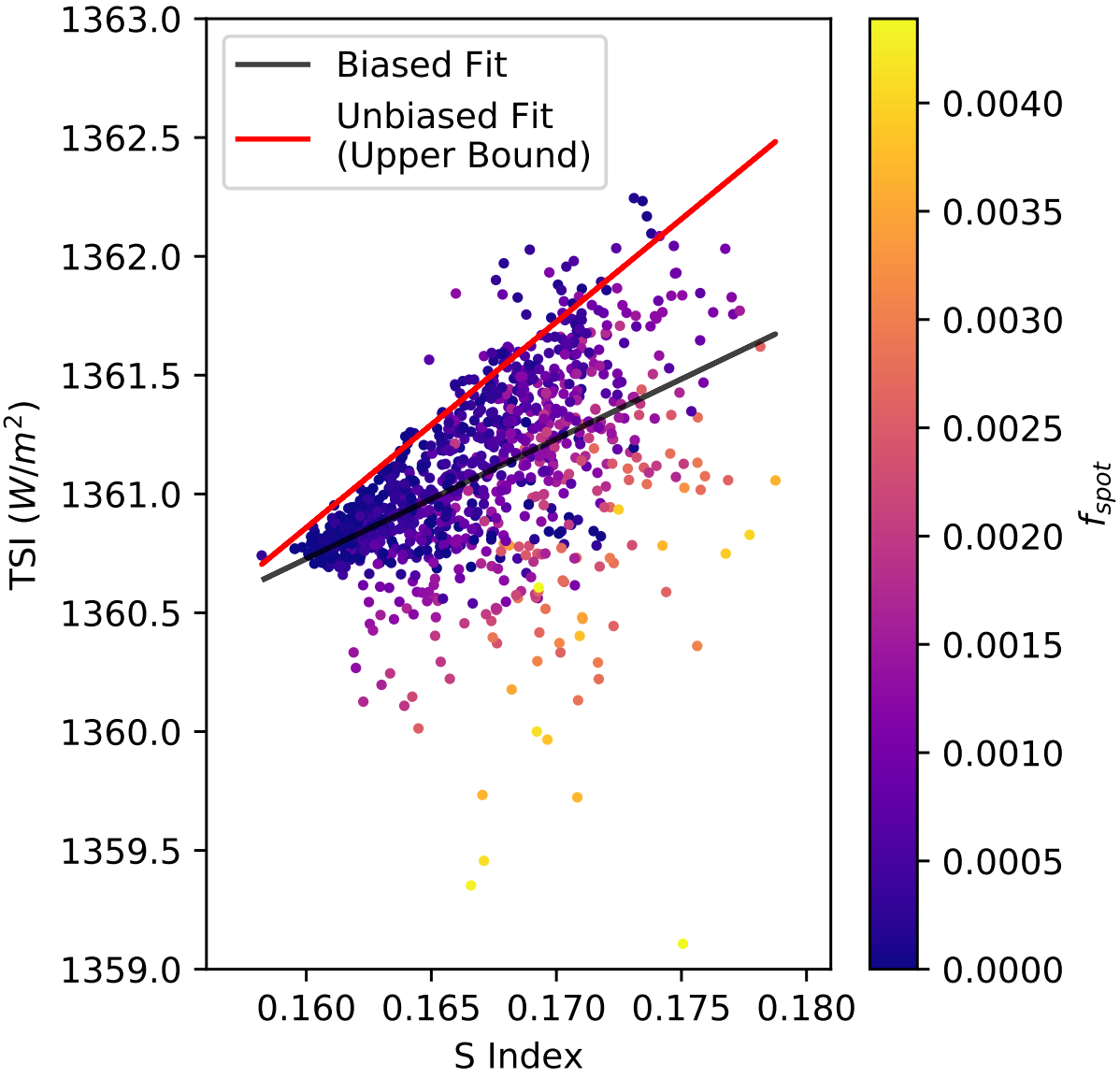

Figure 4.3: A plot of the SORCE TSI versus the HARPS-N/Mt. Wilson S index. The color of each point corresponds to the value of f_{spot}. We see that the S index is highly correlated with the TSI, as expected. We may use this correlation to estimate the plage/network filling factors on the Sun. However, increased spot coverage results in a lower TSI value for a given S index, which will bias our estimate of bright region filling factor. This, in turn, will result in a less accurate estimate of the spot filling factor. The black line shows the result of the straightforward linear fit of TSI and S index, which is biased as described above. To isolate the plage and network driven TSI variations, we find the 50% most densely clustered points in the above scatter plot, and fit a line to the upper boundary of this region. (This choice in point density is arbitrary, but the resulting best-fit line is robust to variations in this parameter.) The resulting fit line, shown in red, is unbiased by the presence of spots.

TSI, S index, f_{plage}, and f_{ntwk} shown in Fig. 4.2, and is also discussed in detail by Chap. 2. We may therefore use a linear transformation of the S-index to provide an initial estimate of f_{bright}, and then use the TSI to estimate f_{spot}. The full calculation is as follows:

(1) We begin by assuming that the S-index is directly proportional to the total plage and network filling factor, as shown in Fig. 4.3 (that is, $f_{bright,1} = m_1 S_{HK} + b_1$), and that the plage and network are the dominant drivers of TSI variation. Our first estimate of the spot filling factor is therefore $f_{spot,1} = 0$. We then estimate the values of m_1 and b_1 by fitting the TSI as a linear transformation of the S-index:

$$TSI_{fit,1}(m_1, b_1) = \mathcal{A}\sigma T_{quiet}^4 \left(1 + 4\frac{\Delta T_{bright}}{T_{quiet}} (m_1 S_{HK} + b_1)\right). \tag{4.3}$$

Note that we have included the physical constants for normalization though they are degenerate with the fit parameters.

It is not sufficient to perform a simple linear fit to the TSI and S index. In the above step, we model the activity-driven variations of the TSI due to the presence of bright regions. However, as first noted by A. Mortier[2] and as shown in Fig. 4.3, the presence of spots produces scatter in this relationship. This scatter is in one direction: as f_{spot} increases, the observed TSI for a given value of the S-index decreases. To isolate the activity-driven TSI variations due to f_{bright}, we determine the 50% most densely clustered points in Fig. 4.3, and fit the upper boundary of this region. The best fit line of this upper bound gives us the values of m_1 and b_1 used above.

(2) Next, we assume any deviation from the fit above are driven by spots, which are not included in this model. We then make a second estimate of the spot filling factor, $f_{spot,2}$, from the residuals

[2]The authors would like to thank A. Mortier in particular for her contribution to this analysis. Failure to account for this effect results in a dramatic underestimation of the f_{spot}, and hinders the use of this technique for estimating RV variations.

to the above fit:

$$f_{spot,2} = \begin{cases} \frac{TSI - TSI_{fit,1}(m_1,b_1)}{4\mathcal{A}\sigma\Delta T_{spot}T_{quiet}^3} & TSI - TSI_{fit,1}(m_1,b_1) \leq 0 \\ 0 & TSI - TSI_{fit,1}(m_1,b_1) > 0 \end{cases} \tag{4.4}$$

Essentially, any point below the line of best fit in Fig. 4.3 is assumed to be due to spot-driven brightness variations. This increases the importance of avoiding spot-driven biases in Step 1. If a simple linear fit is used in Step 1 instead of the fit to the upper boundary described above, the presence of spots will reduce the slope of the best-fit line, which will result in an artificially reduced f_{bright} value, and will also exclude real spot-driven variations from our calculation of $f_{spot,2}$.

(3) We determine our final estimate of f_{bright} and f_{spot} by fitting the following expression to the TSI:

$$TSI_{fit,2} = \mathcal{A}\sigma T_{quiet}^4 \left(1 + 4\frac{\Delta T_{spot}}{T_{quiet}}\left(a_2 f_{spot,2}\right) + 4\frac{\Delta T_{bright}}{T_{quiet}}\left(m_2 S_{HK} + b_2\right)\right). \tag{4.5}$$

where our estimated values of f_{bright} and f_{spot} are given by

$$f_{bright} = m_2 S_{HK} + b_2 \tag{4.6}$$

$$f_{spot} = a_2 f_{spot,2} \tag{4.7}$$

and the parameters a_2, m_2, and b_2 are determined by the above fit. The resulting best-fit parameters derived from the solar case are given in Table 4.1. Note that we do not expect m_2 and b_2 to be very different from the parameters m_1 and b_1 found previously, nor do we expect a_2 to be very different from 1. However, since we exclude any negative residuals from our estimate of f_{spot} in Step 2 above,

m_1	2.02 ± 0.07
b_1	-0.31 ± 0.01
a_2	0.9912 ± 0.0008
m_2	2.072 ± 0.002
b_2	-0.3169 ± 0.0003

Table 4.1: Best-fit parameters for the linear filling factor estimation technique. As expected, m_1 and m_2 are consistent within error bars, as are b_1 and b_2. Similarly, a_2 is very close to 1, as expected..

we perform this final fit in case this excluded information changes the best-fit parameters in any way.

Note that this technique only requires knowledge of the star's distance and radius, along with estimates of the spot and plage/network temperature contrasts and the quiet star effective temperature - this means that it can be used to estimate filling factors without prior knowledge of the filling factors from full-disk images. If the plage and network features are only being used to decorrelate activity-driven RV variations, only time series correlated with the spot, network, and plage filling factors are needed, and the above terms may be absorbed into the fit coefficients in Eqs. 4.3, 4.5—this is discussed further in Sec. 4.5.2.

4.3.3 Differentiating the Network and Plage Filling Factor

In the discussion above, we extract f_{bright}, the combined plage and network filling factor. However, we may consider these two separately by adding a network term to Eq. 4.1:

$$TSI = \mathcal{A}\sigma \left[(1 - f_{spot} - f_{plage} - f_{ntwk})T_{quiet}^4 + f_{spot}(T_{quiet} + \Delta T_{spot})^4 + f_{plage}(T_{quiet} + \Delta T_{plage})^4 + f_{ntwk}(T_{quiet} + \Delta T_{ntwk})^4\right] \quad (4.8)$$

Fitting this equation to the TSI using the SDO observed filling factors reveals that the plage and network have distinct effective temperatures, $\Delta T_{plage} = 32 \pm 1K$ and $\Delta T_{ntwk} = 79 \pm 2K$. This is consistent with the intensity maps produced by HMI, which show that network regions are, on

average, indeed brighter than plage These temperature contrasts are necessary for seperating the plage and network contributions to the filling factor.

Setting Eq. 4.1 equal to Eq. 4.8 and expanding as a power series, we find

$$f_{bright}\Delta T_{bright} \approx f_{plage}\Delta T_{plage} + f_{ntwk}\Delta T_{ntwk}.$$

Since areas are additive, we also expect

$$f_{bright} = f_{plage} + f_{ntwk}.$$

Combining these equations and solving for f_{ntwk} and f_{plage} in terms of f_{bright} yields the following expressions:

$$f_{ntwk} = \frac{\Delta T_{plage} - \Delta T_{bright}}{\Delta T_{plage} - \Delta T_{ntwk}} f_{bright} + \mathcal{B}, \tag{4.9}$$

and

$$f_{plage} = \frac{\Delta T_{bright} - \Delta T_{ntwk}}{\Delta T_{plage} - \Delta T_{ntwk}} f_{bright} - \mathcal{B}. \tag{4.10}$$

Note the prefactors for each estimate, which simply rescale the brightness contributions of each class of active region to account for the different effective temperatures. Also note the offset $\mathcal{B}$ in our estimate of f_{plage}: this accounts for the fact that f_{plage} goes to 0 at solar minimum, while f_{ntwk} has a basal value at solar minimum. In this analysis, the value of $\mathcal{B}$ may be found from the expected value of f_{bright} at solar minimum:

$$\mathcal{B} = \frac{\Delta T_{bright} - \Delta T_{ntwk}}{\Delta T_{plage} - \Delta T_{ntwk}} \min\left(f_{bright}\right). \tag{4.11}$$

Determining this offset therefore requires TSI and S index observations taken at solar minimum, which may increase the observational load associated with this technique. However, modeling the effects of network on the activity-driven solar RV variations only requires a quantity *correlated with* f_{plage}. The value of this offset is therefore unimportant for our purposes.

Hidden Layer Sizes	α	β
(64,64)	0.0001	0.001

Table 4.2: Hyperparameter values for MLP filling factor calculation as optimized from cross-validation. Here α gives the L_2 regularization parameter and β is the learning rate.

4.4 Machine Learning Technique

While machine learning techniques are predicatively powerful, their black-box nature makes them not physically explanatory, and therefore not necessarily useful for some scientific applications. However, the existence of a clear causal connection between the S-index, TSI, and filling factors makes machine learning a strong candidate for the problem of estimating feature-specific magnetic filling factors from spectroscopic and photometric information. We already know the physics connecting these variables, and can therefore have machine learning "discover" and refine the relationships found above. A neural network used as a universal function approximator [100] may be able to determine subtle details of these relationships that are not incorporated into our linear model, such as the different effects of network vs. plage, how underlying spatial distributions of active regions affect the resulting filling factors and activity indicators, and the correlations between spots and plage.

We therefore compare the linear technique discussed in the previous section with a type of neural network known as a Multilayer Perceptron (MLP, [101]). The MLP consists of an input layer, several fully-connected hidden layers, and an output layer. It is one of the simplest neural networks that may be used as a universal function approximator, making it ideal for this application. We implement the MLP using the `MLPRegressor` class in the `scikit-learn` package in Python [102, 103].

We train the MLP using the TSI and S-index inputs, and using the SDO plage, spot, and network filling factors as outputs. 75% of the total available data (taken over the whole solar cycle) is used for training, with 25% set aside to test the performance of the trained network. The MLP uses two

hidden network layers, each with 64 neurons. We optimized the size and number of these layers as well as the L_2 regularization parameter, α, which combats overfitting by constraining the size of the fit parameters as measured with an L_2 norm; and the learning rate, β, which controls the step-size in the parameter space search using five-fold cross validation. That is, we randomly shuffled the training data and divided it into five groups. We trained the network on four of these groups, and then tested the network on the remaining group. We repeated this process using each of the five groups as a test set to mitigate the effects of overfitting on our network, and then repeated the entire five-fold process using each combination of network parameters to determine which combination of hyperparameters resulted in the best performance. The resulting values are summarized in Table 4.2. The network was optimized to minimize square error using a stochastic gradient descent algorithm (SGD), and was trained for a maximum of 10^4 steps (though the algorithm may stop training earlier once the network converges.)

Note that we may also use this MLP approach to fit the solar RVs directly using the TSI and S-index, without first computing magnetic filling factors. In Sec. 4.5.1, we compare a direct MLP fit of this form to RV models dervived from our estimated filling factors to determine if there is any additional RV information in the TSI and S index which is not incorporated into our filling factor estimates.

4.5 Results and Discussion

Fig. 4.4 shows that both the linear and MLP-based techniques successfully reproduce the directly-observed values of f_{spot}, f_{plage}, and f_{ntwk}. Fig. 4.5 shows the same information as Fig. 4.4, but for three 230 day regions taken in the middle of the solar cycle, at solar maximum, and at solar minimum. We see that, again, both the linear and MLP techniques are able to reproduce the SDO-measured values of f_{spot}, f_{plage}, and f_{ntwk} at all points in the stellar activity cycle on these timescales.

Note that there is a systematic ~ 0.004 offset between the linear estimates of f_{ntwk} and the SDO

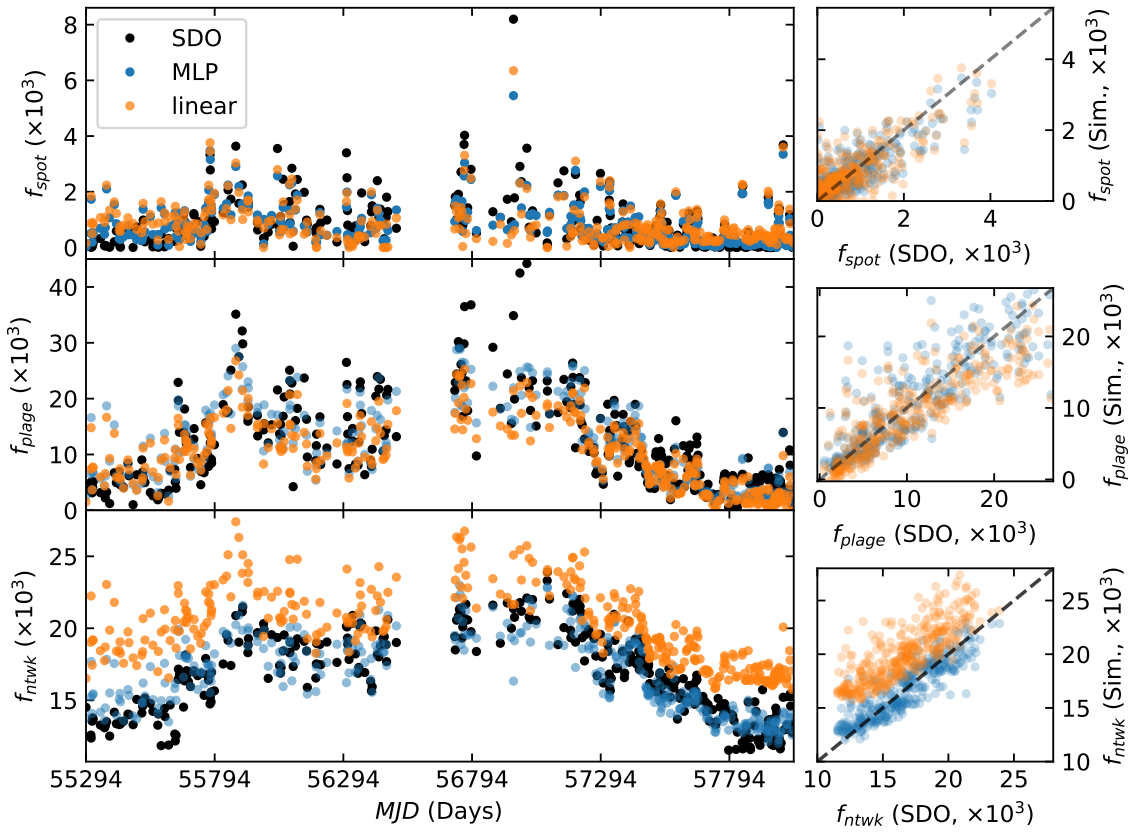

Figure 4.4: Comparison of the SDO/HMI-measured magnetic filling factors (black) to the machine learning (blue) and linear (orange) estimates derived from the S-index and TSI. The time series for the three filling factors are plotted in the left column. The estimated filling factors are plotted as a function of the HMI filling factors in the right column—the grey dashed lines indicate a slope of 1, and are meant to guide the eye. Both the linear and machine learning techniques reproduce the directly-observed values of f_{spot}, f_{plage}, and f_{ntwk}. Note that there is a slight offset between the linear estimate of f_{ntwk} and the SDO measurements. However, this offset is well within the expected 20% - 50% definitional variations reported by [7]

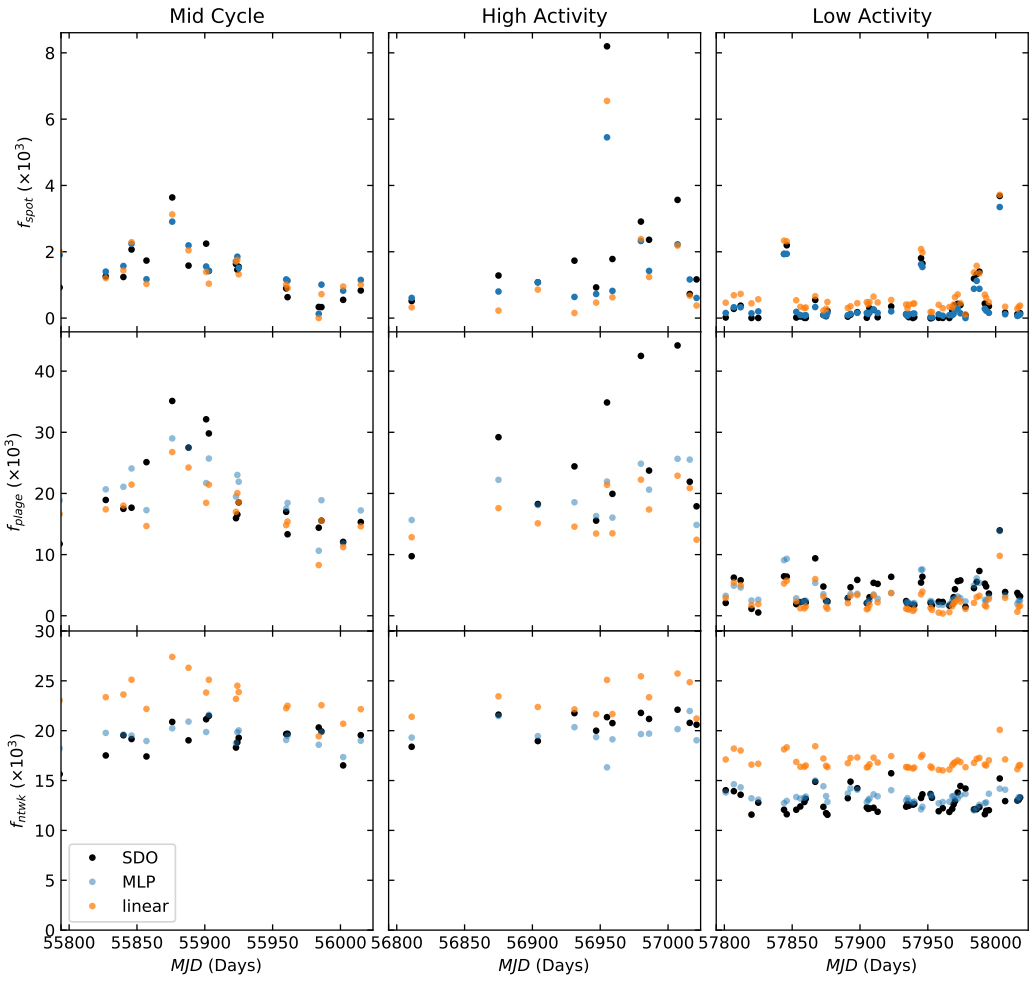

Figure 4.5: 230 day subsets of the time series of the SDO/HMI-observed magnetic filling factors (black), along with the MLP (blue) and linear (orange) estimates derived from the S-index and TSI. Three subsets are shown, taken during the middle of the stellar cycle (left), during solar maximum (middle), and approaching solar minimum (right). Both techniques successfully reproduce f_{spot}, f_{plage}, and f_{ntwk}, with especially good performance at solar minimum.

measured values. This is likely the result of the significant covariance between b_1 and $\mathcal{A}\sigma T_{quiet}^4$ in Eq. 4.3. Any systematic errors in the measured values of $R_\odot$ and T_{quiet}^4 will change the resulting value of b_1, resulting in an offset in the estimated values of f_{ntwk}. Small changes to these parameters can dramatically change the observed offset in f_{ntwk}: artificially increasing T_{quiet}^4 by 0.15 K eliminates the offset entirely. This is well below the precision achieved for measurement of stellar temperatures. While we attain good precision in the solar case, in general linear estimates of f_{ntwk} should assumed to be true up to a constant offset. As stated previously, using these filling factors to remove activity-driven signals from RV measurements only requires values correlated with the filling factor value, making this offset unimportant.

We also note that, while $R_\odot$ and T_{quiet}^4 are assumed to be constants in our model, they do change in time as the result of physical processes not included in our model. These quantities also vary with wavelength: Since here we are using the Ca II H&K lines and integrated visible intensity to reproduce filling factors measured at 6173.3 Å, uncertainties in these parameters associated with their wavelength dependence are inevitable. Indeed, [7] note that measured filling factors will vary by 20% to 50% as a result of these dependencies and other definitional differences: our estimated f_{ntwk} values are certainly consistent with the SDO measured values within these margins.

In Table 4.3, we list the Pearson correlation coefficients between the HMI derived filling factors and our estimates from the linear and MLP techniques. (For the sake of consistency, note that for both the linear and MLP estimates, we compute correlation coefficients only for results generated using the fraction of data reserved for testing the MLP.) We see that both techniques reproduce the information contained in the HMI filling factors, with the MLP performing slightly better than the linear model on all three filling factors. This may indicate that there is some additional information about the filling factors present in the TSI and S index observations that is not being used by the linear technique. However, given the high degree of correlation produced by both techniques, we can use both estimates of the magnetic filling factors to reduce the effects of activity on observed RVs.

	Spots	Plage	Network
Linear	0.81	0.87	0.81
MLP	0.85	0.87	0.89

Table 4.3: Pearson correlation coefficients between HMI ground-truth filling factors and the linear and MLP estimates for each class of filling factor.

Filling Factor Source	$r\left(\Delta v_{conv}, RV(f_{spot}, f_{ntwk}, f_{plage})\right)$
HMI	0.92
Linear Estimate	0.84
MLP Estimate	0.83

Table 4.4: Pearson correlation coefficients between HMI derived estimate of the suppression of convective blueshift, Δv_{conv} and the activity driven RVs derived from Eq. 4.13. The very high correlation coefficients indicate that the plage and network filling factors successfully estimate the RV contribution of the suppression of convective blueshift, as expected from Chap. 2.

4.5.1 Application to Solar RVs

In Chap. 2, we found that the HARPS-N solar radial velocities were well-represented by a linear combination of Δv_{conv}, the suppression of convective blueshift, and Δv_{phot}, the photometric velocity shift due to bright and dark active regions breaking the symmetry of the solar rotational profile:

$$RV = A_1 \Delta v_{phot} + B_1 \Delta v_{conv} + RV_0. \tag{4.12}$$

Here, we see if we can perform a similar reconstruction using our estimates of the magnetic filling factors. Since the presence of active regions drives the suppression of convective blueshift, we expect the Δv_{conv} to be proportional to the spot, plage, and network filling factors. Based on the results of Chap. 2, we also expect network and plage regions to have different contributions to Δv_{conv}. We therefore model the suppression of convective blueshift as:

Filling Factor Source	$r\left(\Delta v_{phot}, RV(f_{spot})\right)$
HMI	0.62
Linear Estimate	0.70
MLP Estimate	0.67

Table 4.5: Pearson correlation coefficients between HMI derived estimate of the photometric velocity shift, Δv_{phot} and the activity driven RVs derived from Eq. 4.14. The relatively high correlation coefficients indicate that the spot filling factors successfully estimate the photometric RV shifts, as expected from Chap. 2.

	RMS (m s^{-1})
Full solar dataset	1.64
Decorrelated with S index	1.10
Decorrelated with HMI filling factors	0.91
Decorrelated with linear filling factor estimates	1.04
Decorrelated with MLP filling factor estimates	1.02
Decorrelated with MLP RV estimate	0.96

Table 4.6: RMS RV residuals from several models and methods. Using our estimates of f_{spot}, f_{plage}, and f_{ntwk} in Eq. 4.15 reduces the RMS RVs by 60 cm s^{-1}. However, using the HMI-observed filling factors reduces the RMS residuals by a further 13 cm s^{-1}, indicating there is additional information in these filling factors not captured by our estimates. A direct MLP fit to the solar RVs, using the S index and TSI as inputs, performs better than our estimated filling factors, but does not perform as well as the fit to HMI filling factors. This indicates that, while our estimated filling factors are highly correlated with the observed values, the S index and TSI alone are insufficient to completely characterize the filling factors of each feature.

$$\Delta v_{conv} = B f_{spot} + C f_{plage} + D f_{network} + E. \tag{4.13}$$

While plage and network occupy a greater area than spots on the Sun, and therefore dominate the suppression of convective blueshift, the higher brightness contrast of spots means that they drive the photometric RV shift, Δv_{phot}. We expect Δv_{phot} to scale with number and size of the spots rotating across the solar surface. However, we also expect a phase lag between f_{spot} and Δv_{phot}. For a single spot moving across the solar disk, f_{spot} is at its maximum value when the spot is on the center of the solar disk. However, the absolute value of Δv_{phot} is maximized when the spot is at the solar limb, rotating toward or away from the observer, and is zero when the spot is at disk center. We therefore expect Δv_{phot} to also depend on the derivative of the filling factor with respect to time:

$$\Delta v_{phot} \propto f_{spot} \times \left(\frac{df_{spot}}{dt} \right). \tag{4.14}$$

Note that this formulation mirrors the FF′ method developed by [40].

By combining Eqs. 4.12, 4.13, and 4.14, we therefore produce a model of the solar RVs based on

Filling Factor Source	A (10^5 m)	B (m s^{-1})	C (m s^{-1})	D (m s^{-1})	RV_0 (m s^{-1})
SDO	7.7 ± 1.2	165 ± 76	244 ± 15	52 ± 27	-2.0 ± 0.3
Linear	7.1 ± 1.3	470 ± 80	281 ± 11	0^*	-1.32 ± 0.08
MLP	7.5 ± 1.4	479 ± 218	242 ± 52	2 ± 90	-1.3 ± 1.2

Table 4.7: Best fit coefficients for fitting Eq. 4.15 to the HARPS-N solar RVs using SDO filling factors, linear estimates of the filling factors, and MLP estimates of the filling factors. Note that the estimates of f_{plage} and f_{ntwk} derived from the linear technique are both linear transformations of the S index. We therefore require $D = 0$ to avoid degeneracies when using the filling factor estimates derived from the linear technique.

our estimated feature-specific magnetic filling factors:

$$RV = Af_{spot}\left(\frac{df_{spot}}{dt}\right) + Bf_{spot} + Cf_{plage} + Df_{network} + RV_0. \tag{4.15}$$

Note that the offset E in Eq. 4.13 has been absorbed into RV_0.

We then fit the HARPS-N solar telescope RVs to Eq. 4.15 using the directly-measured SDO filling factors. The results of each fit is given in Table 4.7. Note that since the linear estimates of f_{bright}, f_{ntwk}, and f_{plage} are linear transformations of the S index, as shown in Eqs. 4.6, 4.9, and 4.10, we require $D = 0$ to avoid degeneracies when using the filling factor estimates derived from the linear technique. Fitting the linear estimated filling factors to Eq. 4.15 without this constraint is equivalent to fitting to $RV = Af_{spot}\left(\frac{df_{spot}}{dt}\right) + Bf_{spot} + C'S_{HK} + D'S_{HK} + E'$. We therefore set $D = 0$ to avoid having degeneracies between C and D in our fit. No such constraint on D is necessary when considering the SDO or MLP filling factors.

To ensure our fit is indeed reproducing the suppression of convective blueshift and the photometric RV shift, as expected, we compare the relevant terms of Eq. 4.15 to the SDO/HMI estimates of these RV perturbations, as calculated in Chap. 2 In Tables 4.4 and 4.5, we compare the estimates of Δv_{conv} and Δv_{phot} computed from Eqs. 4.13 and 4.14 using the filling factors measured by SDO, and estimated using the linear and MLP techniques to the values of Δv_{conv} and Δv_{phot} derived from HMI observations in [2] and Chap. 2. We see that all of the estimated values of Δv_{conv} are highly correlated with the HMI-derived velocities. Our estimates of Δv_{phot} are less correlated

with the actual photometric shift, but still show good agreement. Interestingly, including the contributions of plage and network regions in Eq. 4.14 — that is, adding terms $\propto f_{plage} \times \left(\frac{df_{plage}}{dt}\right)$ and $\propto f_{ntwk} \times \left(\frac{df_{ntwk}}{dt}\right)$ — does not appear to increase the correlation coefficient. However, we may still conclude that the RVs calculated using Eq. 4.15 indeed do correspond to the combination of suppression of convective blueshift and photometric RV shift described by Eq. 4.12.

For both the SDO measured and MLP-derived filling factors, we see $C > D$. This is consistent with the idea that the denser magnetic interconnections available in photospheric plages are more successful in inhibiting convection, and thus convective blueshifts, than the sparser network magnetizations, as suggested in Chap. 2. Indeed, we see that, using MLP estimates, the network contribution is consistent with zero, and using SDO observations, the network contribution is only $\sim 2\sigma$ above zero.

The B coefficients, which describe the spot contributions to the suppression of convective blueshift, vary depending on the filling factors used. The linear and MLP estimates of f_{spot} receive a heavy weighting, while the SDO weighting is about a factor of 3 smaller. The MLP estimates also have a contribution only $\sim 2\sigma$ above zero. However, the Sun is a plage-dominated star, and f_{spot} is about a factor of 100 times smaller than f_{plage}, as shown in Fig. 4.2. So, while the precise weighting of f_{spot} varies based on the values used, in all cases their contribution to the suppression of convective blueshift will be negligible compared to that of f_{plage}. We may therefore conclude that, as suggested by Chap. 2, plage regions are the dominate contribution to the solar suppression of convective blueshift, while spots are the dominant contribution to the photometric RV shift: knowledge of the plage and spot filling factors are therefore sufficient to reproduce Δv_{conv} and Δv_{phot} respectively.

Fitting to the HMI-observed filling factors reduces the HARPS-N solar RV residuals from 1.64 m s^{-1} to 0.91 m s^{-1}, as shown in Table 4.6. In comparison, the usual technique of simply decorrelating the S-index from the RV measurements (i.e., fitting $RV = AS_{HK} + B$) results in an RMS of only 1.10 m s^{-1}, indicating that spots, plage, and network regions have different contributions to the S index, and have different effects on the suppression of convective blueshift [1].

Repeating this fit with both our linear and MLP estimates of f_{spot}, f_{ntwk}, and f_{plage} reduces the RMS RV to 1.04 m s^{-1} and 1.02 m s^{-1} respectively. This implies that, while our estimates are highly correlated with the true values of the filling factors, there is additional information in the true filling factors that is not captured by either technique, resulting in less precise estimates of the convective blueshift and photometric RV shifts. Interestingly, while the linear filling factor estimates cannot distinguish the RV contributions of the spots and network, as discussed above, the linear and MLP estimates result in similar RMS RVs. The fact that the linear estimates reduce the RMS RVs below the level obtained from the S index despite this limitation highlights the importance of spots in our models of activity-driven RVs.

To see if it is possible for a more refined technique to extract further RV information from our inputs, we fit an MLP directly to the solar RVs using the S-index and TSI as inputs. This is similar to the technique proposed by [104], but replacing the residual cross-correlation function with the S-index and TSI. The hyperparameters of this MLP are the same as those given in Table 4.2. As before, we divide our data into training and test sets, and the quoted residuals are derived from the test set. This fit results in an RMS residual of 0.96 m s^{-1}, (as shown in Table 4.6) indicating that there is indeed more RV information to be gained from this set of observations. Interestingly, however, while this RMS value is below the residuals obtained from both sets of estimated filling factors, it is greater than the 0.91 m s^{-1} residuals obtained by using the direct SDO measurements of the filling factors. This appears to indicate that, while the S index and TSI contain more information than our linear and MLP estimates could obtain, they do not contain *all* the information about the solar plage, spot, and network coverage.

This is unsurprising: we note that network regions can form from decaying plage regions. Due to the geometry of the magnetic flux tubes associated with these regions, a network region may rotate onto the limb, become a plage region as it rotates onto disk center, and then become a network region again as it rotates back onto the limb. The linear technique directly uses the different temperature contrasts of network and plage to provide a useful first pass at differentiating these regions, but does not capture these links between them. That is, there are additional physical effects that further complicate the relationship between photometry, spectroscopy, filling factors, and RVs [105].

While the underlying behavior of the MLP is unknown, it likely employs a similar, slightly more complex technique to differentiate the two classes of regions. The magnetic intensification effect, which strengthens lines in the presence of a magnetic field (e.g., [106, 107]), has an RV signal which depends on the overall filling factor, as well as a given line's wavelength, effective Landé g value, and the magnetic field strength [108]. HMI monitors the photospheric 6173.3 Å iron line: these wavelength-dependent effects mean that the filling factors derived from HMI may not be consistent with those derived from the chromospheric calcium H and K lines. The center-to-limb dependence of the calcium H and K lines are different than the 6173.3 Å line as well, which could lead to mismatches in the derived filling factors as a function of rotational phase. More complicated linear and MLP-based filling factor estimates could use spectroscopic measurements of additional absorption lines, and photometric measurements integrated over different wavelength bands to compensate for these effects, and to exploit different wavelength-dependent contrasts of each feature to better separate these three classes of magnetic active regions.

The direct MLP fit to the solar RVs and its residuals are plotted in Fig. 4.6. The effects of HARPS-N cryostat cold plate warm-ups, discussed in [47, 95], are clearly visible in the fit residuals, indicating that the MLP is not "learning" instrumental systematics, and that the residual RV variations below this level are likely dominated by a combination of instrumental systematics and activity processes not reflected by variations in the S-index and TSI. Further work is necessary to identify these remaining activity processes, and to disentangle them from instrumental effects.

4.5.2 Application to the Stellar Case

The techniques developed in this work should be applicable to Sun-like stars with the proper observational cadence. To reproduce properly scaled filling factors, the linear technique requires precise knowledge of radius, effective temperature, and distance to the target, as well as the temperature contrasts of the plage, spots, and network. The effective temperature may be calculated spectroscopically [109], while the temperature contrasts may be assumed to be Sun-like in the case of G-class stars. The stellar radius may then be calculated photometrically, using the spectroscopic

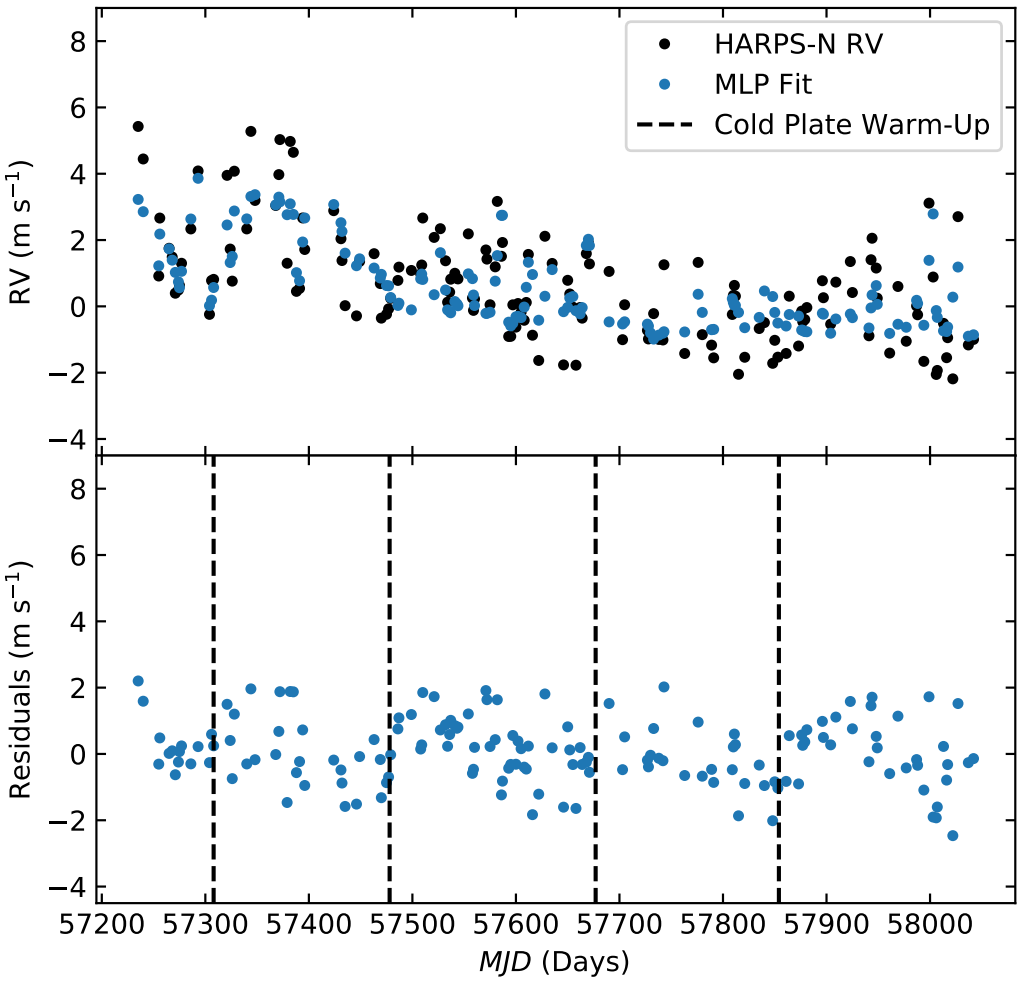

Figure 4.6: MLP fit to the HARPS-N solar telescope data. HARPS-N RVs are shown in black, and MLP estimates of the RVs are in blue. Fit residuals are shown in the bottom panels: HARPS-N cryostat warm-up dates (*see text*) are indicated with black dashed lines.

temperature as a prior. The stellar distance may be straightforwardly determined through parallax measurements, while the rotation period may be obtained via photometry or through RV measurements.

Although the techniques presented here assume a plage-dominated star, it is straightforward to rework Eqs. 4.3 and 4.5 for a spot-dominated target: in this case, the S index is assumed to be correlated with spot-driven variations of the TSI, with positive deviations indicating the presence of bright, plage regions.

While properly scaling the linear estimate of f_{plage} requires observations near stellar minimum, modelling RV variations only requires time series which are proportional to f_{spot}, f_{plage} and f_{ntwk}—in this case, this offset is unimportant, and no additional constraint is placed on the stellar observations. Furthermore, the values of ΔT_{spot}, ΔT_{plage} and ΔT_{ntwk} may be absorbed into the fit coefficients in Eqs. 4.3, 4.5, and 4.15, further simplifying matters.

The MLP machine learning technique, in contrast, requires less knowledge of the target star: while the mathematical and physical transformations learned by MLP are unknown to the user, the MLP is presumably learning a more sophisticated version of the linear technique, and implicitly "learns" the solar values for feature temperature contrast and quiet temperature as it identifies higher-order correlations between the TSI, S index, and filling factors. This makes the MLP straightforward to implement when precise contrast values are unknown. Furthermore, since the MLP uses no timing information, it places no constraints on the observational cadence or baseline: it only requires simultaneous photometric and spectroscopic measurements.

However, since ground-truth filling factors can only be directly measured in the solar case, the MLP must be trained using solar data. Its stellar application is therefore limited to Sun-like stars (that is, stars with very similar surface filling factors as the Sun, or possibly even only solar twins), making it less generalizable to other stellar targets. Since stars other than the Sun cannot be resolved spatially at high resolution, assessing just how "Sun-like" a target needs to be for the machine learning technique to yield meaningful results is challenging. One possibility is to generate

synthetic stellar images for targets with a variety of spectral types, activity levels, feature contrasts, and viewing angles using SOAP 2.0 [76], StarSim [110], or a similar platform, computing the light curve and S-index for these images, and seeing if an MLP trained on the solar case reproduces the filling factors expected for each image. Such an analysis is beyond the scope of this work.

4.6 Conclusions

We assess two techniques to extract spot, plage, and network filling factors using simultaneous spectroscopy and photometry. The first technique involves a straightforward analytical manipulation of the S-index and TSI time series, while the second uses a neural network machine learning technique known as a Multilayer Perceptron (MLP) trained on ground-truth filling factors derived from full-disk solar images. Both techniques yield filling factor estimates which are highly correlated with values derived from full-disk solar images, with Spearman correlation coefficients ranging from 0.81 and 0.89 from each technique.

We show that decorrelating a nearly-three-year time series of solar RVs using HMI-observed spot, plage, and network filling factors effectively reproduces the expected RV variations due to the convective blueshift and rotational imbalance due to flux inhomogeneities, reducing the residual activity-driven RVs more than the typical technique of decorrelating using spectroscopic activity indices alone. Fitting to HMI filling factors reduces the RV RMS from 1.64 m s^{-1} to 0.91 m s^{-1}, while fitting to the S-index alone results in an RMS variation of 1.10 m s^{-1}. Including this additional information about spots, plage, and network thus accounts for an additional $\sqrt{(1.10\text{m s}^{-1})^2 - (0.91\text{m s}^{-1})^2} = 0.62\text{m s}^{-1}$ of RMS variation. The filling factor estimates from both the linear and MLP techniques offer some improvement to the RMS residuals beyond what is obtained from only the S-index. Decorrelating with the linear estimates reduces the RMS variation to 1.04 m s^{-1}, and the MLP estimated filling factors reduces the RMS to 1.02 m s^{-1}.

Using a MLP trained directly on the solar RVs, we reduced the RMS to 0.96 m s^{-1}. While this indicates that the S-index and TSI contain more RV information than obtained by either estimate

of our filling factors, it does not lower the RMS RVs below the 0.91 m s^{-1} limit obtained using direct measurements of the magnetic filling factors. This suggests that, while our initial estimates of f_{spot}, f_{plage}, and f_{ntwk} are highly correlated with the expected value, more information is needed to fully characterize these feature-specific filling factors. To match the performance of the HMI filling factors, a more sophisticated version of this technique, using additional spectral lines and photometric bands will likely be necessary.

Both the analytical and machine learning techniques may be used to extract filling factors on other stars: the analytical technique is more widely generalizable, but requires detailed knowledge of the star and good temporal sampling, ideally with observations of the target at activity minimum. The machine learning technique, in contrast, requires no additional knowledge of the target star, and applies no constraints on the observing schedule—however, it is only applicable to stars with very similar filling factor properties as the Sun.

5

Outlook

In this work, we investigate the effects of specific active regions on activity-driven RVs using the solar case. We also develop techniques for measuring these effects on Sun-like stars. Here, we briefly discuss the possible applications of these techniques to non-solar targets, and the remaining challenges surrounding RV exoplanet searches.

The HARPS-N solar data set consists of 35,000 exposure taken approximately once every five minutes over a period of three years. The Solar Dynamics Observatory (SDO) has taken full disk images once every 45 seconds over all of Solar Cycle 24. Similarly, SORCE and the Mt. Wilson S-index project have watched the Sun for all of Solar Cycle 24, and BiSON has observed the solar p-modes since 1978.

The majority of stellar targets do not have this long observational baseline. The best observed RV targets belong to the HARPS-N Rocky Planet Search (RPS) [111]. The RPS survey consists of 51 nearby, low-activity stars with spectral types G8 - M0, $\log R'_{HK} < -4.69$, and an average RV dispersion of about 2 m s^{-1}. The best-observed targets have over 1000 observations, though most only have a few hundred, as shown in Fig. 5.1.

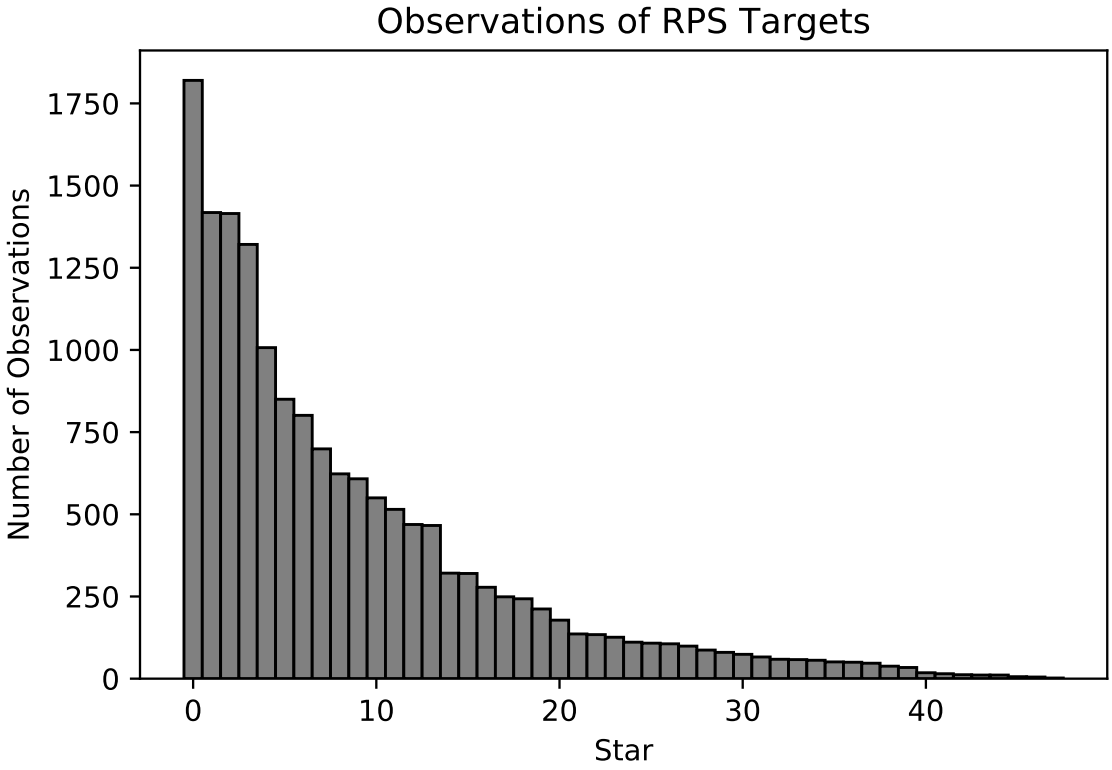

Figure 5.1: The number of observations for the RPS survey.

Using the p-mode shifts as a proxy to distiniquish contributions from small and large active regions, as discussed in Chap. 3, will require short-cadence observations of stellar targets over the duration of a star's activity cycle. (BiSON observes the Sun with a 40 second cadence [28, 89]). Similarly, the techniques presented in Chap. 4 require high signal-to-noise spectroscopic observations, with near simulatenous photometry. These constraints may be probitive for anything outside of dedicated observations of the brightest targets.

While this work represents great progress in modelling activity-driven RV variations, other challenges still remain. Beyond the suppression of convective blueshift and flux inhomogenities modelled in this work, the RVs are effected by other processes, as shown in Figs. 2.7, 4.6, and Table 4.6. While the techniques presented in this work accounts of 1.3 m s^{-1} of RMS variation, 0.91 m s^{-1} still remains to be understood. For instance, this work does not account for the effects of granulation [30, 33, 73, 74], supergranulation [31, 64]. Additionally, while tremendous work has already been done to understand and calibrate the instrumental effects of the HARPS-N spectrograph [24, 25, 27], the effects of HARPS-N are still visible in the RV residuals, as evidenced by the coldplate warm-ups visible in Fig. 4.6. Further work is necessary to completely characterize exoplanet-hunting spectrographs, including the implementation of the turn-key astro-comb developed by A. Ravi, et al. [17, 112] at HARPS-N. However, the continued progress understanding and modelling these physical and instrumental effects brings us closer and closer to the detection of true exo-Earths, other habitable worlds, and answers to the questions "what is our place in the universe" and "are we alone in the universe?"

Appendix

A

Description of HMI Image Analysis

In Chapters 2 and 4 of this work, we compute the filling factors of sunspots, plage, and network, and the radial velocities (RVs) associated with the suppression of convective blueshift and rotational imbalance. These calculations are based on the methods of [2], with some differences. Here, we briefly review the methods of that paper, highlighting the differences in our implementation.

A.1 Identifying Active Regions

We identify solar active regions using the same thresholding methods as [2]. We identify active regions using the line of sight HMI magnetograms. Active pixels have a magnetic field greater than

$$|B| > 3\sigma/\mu$$

where $\sigma = 8$G, the shot noise per SDO/HMI pixel, and $\mu = \cos\theta$, where θ gives the angular position from the center of the Sun.

To differentiate between dark spots and bright plage, we apply an intensity threshold. We compute the average quiet-sun intensity $I_{\rm quiet}$ by averaging all the inactive pixels identified using the above

threshold. Pixels are identified as spots using the intensity threshold of [4] - that is, if

$$I_{ij} < 0.89 * I_{\rm quiet}$$

The overall, spot, and plage filling factors are calculated simply by computing the fraction of SDO/HMI pixels corresponding to a given active region type relative to the number of pixels on the solar disk, N_{sun}:

$$f_{total} = \frac{1}{N_{sun}} \sum_{ij} W_{ij}$$

Here $W_{ij} = 1$ if pixel-ij corresponds to an active region, and is 0 otherwise. The same equation may be used to calculate the bright (plage/network) and spot filling factors, f_{bright} and f_{spot}: in those cases, $W_{ij} = 1$ if pixel-ij corresponds to falls above or below the intensity threshold described above. We therefore find that $f_{total} = f_{bright} + f_{spot}$. Similarly, we may differentiate $f_{plage}$and f_{ntwk} by taking $W = 1$ if pixel-ij corresponds to falls above or below the area threshold.

A.2 Calculation of Active Region Velocities

A.2.1 The convective velocity, $\Delta\hat{v}_{\rm conv}$

Our calculation of the activity-driven RV shifts differs slightly from that of [2]. Our calculation of $\Delta\hat{v}_{\rm conv}$ is given by computing the disk-averaged Doppler velocity, $\hat{v}$, and subtracted the disk-averaged quiet-sun velocity, $\hat{v}_{\rm quiet}$:

$$\Delta\hat{v}_{\rm conv} = \hat{v} - \hat{v}_{\rm quiet}$$

$\hat{v}$ is the intensity-weighted average of the dopplergram, with the spacecraft velocity and rotation profile ($\hat{v}_{sc}$ and $\hat{v}_{rot}$) removed:

$$\hat{v} = \frac{\sum_{ij} (v_{ij} - v_{sc,ij} - v_{rot,ij}) I_{ij}}{\sum_{ij} I_{ij}}$$

and $\hat{v}_{\text{quiet}}$ is the intensity-weighted average over the quiet pixels only:

$$\hat{v}_{\text{quiet}} = \frac{\sum_{ij} (v_{ij} - v_{sc,ij} - v_{rot,ij}) I_{ij} \bar{W}_{ij}}{\sum_{ij} I_{ij} \bar{W}_{ij}}$$

where $\bar{W}_{ij} = 1$ for inactive pixels, and is 0 otherwise.

A.2.2 The photometric velocity, $\Delta\hat{v}_{\text{phot}}$

The photometric velocity is calculated

$$\hat{v}_{\text{phot}} = \frac{\sum_{ij} v_{rot,ij} (I_{ij} - \hat{K} L_{ij}) W_{ij}}{\sum_{ij} I_{ij}}$$

here $\hat{K}$ is the average quiet-sun intensity at disk center, and L_{ij} gives the limb darkening at the ij-th pixel.

Bibliography

[1] N. Meunier, A.-M. Lagrange, and M. Desort, "Reconstructing the solar integrated radial velocity using MDI/SOHO," A&A **519**, A66 (2010), arXiv:1005.4764 [astro-ph.SR] .

[2] R. D. Haywood, A. Collier Cameron, Y. C. Unruh, C. Lovis, A.F. Lanza, J. Llama, M. Deleuil, R. Fares, M. Gillon, C. Moutou, F. Pepe, D. Pollacco, D. Queloz, and D. SÃ©gransan, "The sun as a planet-host star: proxies from sdo images for harps radial-velocity variations," Monthly Notices of the Royal Astronomical Society **457**, 3637–3651 (2016).

[3] William J Chaplin, Rachel Howe, Sarbani Basu, Yvonne Elsworth, Timothy W Milbourne, RaphaÃ«lle D Haywood, Guy R Davies, Steven J Hale, Andrea Miglio, and Eddie Ross, "Sensitivity of low-degree solar p modes to active and ephemeral regions: frequency shifts back to the maunder minimum," Monthly Notices of the Royal Astronomical Society: Letters **489**, L86–L90 (2019).

[4] K. L. Yeo, S. K. Solanki, and N. A. Krivova, "Intensity contrast of solar network and faculae," A&A **550**, A95 (2013).

[5] Greg Kopp and George Lawrence, "The total irradiance monitor (tim): Instrument design," Solar Physics **230**, 91–109 (2005).

[6] Greg Kopp, Karl Heuerman, and George Lawrence, "The total irradiance monitor (tim): Instrument calibration," Solar Physics **230**, 111–127 (2005).

[7] N. Meunier, M. Desort, and A.-M. Lagrange, "Using the sun to estimate earth-like planets detection capabilities ii. impact of plages," Astronomy and Astrophysics **512**, 3637–3651 (2010), www.aanda.org/articles/aa/pdf/2010/04/aa13551-09.pdf .

[8] M. Kuzuhara, M. Tamura, T. Kudo, M. Janson, R. Kandori, T. D. Brandt, C. Thalmann, D. Spiegel, B. Biller, J. Carson, Y. Hori, R. Suzuki, A. Burrows, T. Henning, E. L. Turner, M. W. McElwain, A. Moro-Martín, T. Suenaga, Y. H. Takahashi, J. Kwon, P. Lucas, L. Abe, W. Brandner, S. Egner, M. Feldt, H. Fujiwara, M. Goto, C. A. Grady, O. Guyon, J. Hashimoto, Y. Hayano, M. Hayashi, S. S. Hayashi, K. W. Hodapp, M. Ishii, M. Iye, G. R. Knapp, T. Matsuo, S. Mayama, S. Miyama, J.-I. Morino, J. Nishikawa, T. Nishimura, T. Kotani, N. Kusakabe, T.-S. Pyo, E. Serabyn, H. Suto, M. Takami, N. Takato, H. Terada, D. Tomono,

M. Watanabe, J. P. Wisniewski, T. Yamada, H. Takami, and T. Usuda, "DIRECT IMAGING OF a COLD JOVIAN EXOPLANET IN ORBIT AROUND THE SUN-LIKE STAR GJ 504," The Astrophysical Journal **774**, 11 (2013).

[9] Markus Kasper, Jean-Luc Beuzit, Christophe Verinaud, Raffaele G. Gratton, Florian Kerber, Natalia Yaitskova, Anthony Boccaletti, Niranjan Thatte, Hans Martin Schmid, Christoph Keller, Pierre Baudoz, Lyu Abe, Emmanuel Aller-Carpentier, Jacopo Antichi, Mariangela Bonavita, Kjetil Dohlen, Enrico Fedrigo, Hiddo Hanenburg, Norbert Hubin, Rieks Jager, Visa Korkiakoski, Patrice Martinez, Dino Mesa, Olivier Preis, Patrick Rabou, Ronald Roelfsema, Graeme Salter, Mathias Tecza, and Lars Venema, "EPICS: direct imaging of exoplanets with the e-ELT," in *Ground-based and Airborne Instrumentation for Astronomy III*, edited by Ian S. McLean, Suzanne K. Ramsay, and Hideki Takami (SPIE, 2010).

[10] Gregory W. Henry, Geoffrey W. Marcy, R. Paul Butler, and Steven S. Vogt, "A transiting "51 peg–like" planet," The Astrophysical Journal **529**, L41–L44 (2000).

[11] David Charbonneau, Timothy M. Brown, David W. Latham, and Michel Mayor, "Detection of planetary transits across a sun-like star," The Astrophysical Journal **529**, L45–L48 (2000).

[12] William J. Borucki, David Koch, Gibor Basri, Natalie Batalha, Timothy Brown, Douglas Caldwell, John Caldwell, Jørgen Christensen-Dalsgaard, William D. Cochran, Edna DeVore, Edward W. Dunham, Andrea K. Dupree, Thomas N. Gautier, John C. Geary, Ronald Gilliland, Alan Gould, Steve B. Howell, Jon M. Jenkins, Yoji Kondo, David W. Latham, Geoffrey W. Marcy, Søren Meibom, Hans Kjeldsen, Jack J. Lissauer, David G. Monet, David Morrison, Dimitar Sasselov, Jill Tarter, Alan Boss, Don Brownlee, Toby Owen, Derek Buzasi, David Charbonneau, Laurance Doyle, Jonathan Fortney, Eric B. Ford, Matthew J. Holman, Sara Seager, Jason H. Steffen, William F. Welsh, Jason Rowe, Howard Anderson, Lars Buchhave, David Ciardi, Lucianne Walkowicz, William Sherry, Elliott Horch, Howard Isaacson, Mark E. Everett, Debra Fischer, Guillermo Torres, John Asher Johnson, Michael Endl, Phillip MacQueen, Stephen T. Bryson, Jessie Dotson, Michael Haas, Jeffrey Kolodziejczak, Jeffrey Van Cleve, Hema Chandrasekaran, Joseph D. Twicken, Elisa V. Quintana, Bruce D. Clarke, Christopher Allen, Jie Li, Haley Wu, Peter Tenenbaum, Ekaterina Verner, Frederick Bruhweiler, Jason Barnes, and Andrej Prsa, "Kepler planet-detection mission: Introduction and first results," Science **327**, 977–980 (2010), http://science.sciencemag.org/content/327/5968/977.full.pdf .

[13] Steve B. Howell, Charlie Sobeck, Michael Haas, Martin Still, Thomas Barclay, Fergal Mullally, John Troeltzsch, Suzanne Aigrain, Stephen T. Bryson, Doug Caldwell, William J. Chaplin, William D. Cochran, Daniel Huber, Geoffrey W. Marcy, Andrea Miglio, Joan R. Najita, Marcie Smith, J. D. Twicken, and Jonathan J. Fortney, "The k2 mission: Characterization and early results," Publications of the Astronomical Society of the Pacific **126**, 398 (2014).

[14] George R. Ricker, Joshua N. Winn, Roland Vanderspek, David W. Latham, GÃ¡spÃ¡r Ã. Bakos, Jacob L. Bean, Zachory K. Berta-Thompson, Timothy M. Brown, Lars Buchhave, Nathaniel R. Butler, R. Paul Butler, William J. Chaplin, David B. Charbonneau, JÃžrgen Christensen-Dalsgaard, Mark Clampin, Drake Deming, John P. Doty, Nathan De Lee, Courtney Dressing, Edward W. Dunham, Michael Endl, FranÃ§ois Fressin, Jian Ge, Thomas Henning, Matthew J. Holman, Andrew W. Howard, Shigeru Ida, Jon M. Jenkins, Garrett Jernigan, John Asher Johnson, Lisa Kaltenegger, Nobuyuki Kawai, Hans Kjeldsen, Gregory Laughlin, Alan M. Levine, Douglas Lin, Jack J. Lissauer, Phillip MacQueen, Geoffrey Marcy, Peter R.

McCullough, Timothy D. Morton, Norio Narita, Martin Paegert, Enric Palle, Francesco Pepe, Joshua Pepper, Andreas Quirrenbach, Stephen A. Rinehart, Dimitar Sasselov, Bunâei Sato, Sara Seager, Alessandro Sozzetti, Keivan G. Stassun, Peter Sullivan, Andrew Szentgyorgyi, Guillermo Torres, Stephane Udry, and Joel Villasenor, "Transiting exoplanet survey satellite," Journal of Astronomical Telescopes, Instruments, and Systems **1**, 1 – 1 – 10 (2014).

[15] V. Cessa, T. Beck, W. Benz, C. Broeg, D. Ehrenreich, A. Fortier, G. Peter, D. Magrin, I. Pagano, J.-Y. Plesseria, M. Steller, J. Szoke, N. Thomas, R. Ragazzoni, and F. Wildi, "CHEOPS: a space telescope for ultra-high precision photometry of exoplanet transits," in *International Conference on Space Optics â ICSO 2014*, Vol. 10563, edited by Zoran Sodnik, Bruno Cugny, and Nikos Karafolas, International Society for Optics and Photonics (SPIE, 2017) pp. 468 – 476.

[16] Perryman, M., *The Exoplanet Handbook* (Cambridge University Press, 2011).

[17] Aakash Ravi, *Topics in precision astrophysical spectroscopy*, Ph.D. thesis, Harvard University (2020).

[18] A. Mortier, S. G. Sousa, V. Zh. Adibekyan, I. M. Brandão, and N. C. Santos, "Correcting the spectroscopic surface gravity using transits and asteroseismology," Astronomy & Astrophysics **572**, A95 (2014).

[19] L. Malavolta, C. Lovis, F. Pepe, C. Sneden, and S. Udry, "Atmospheric stellar parameters from cross-correlation functions," Monthly Notices of the Royal Astronomical Society **469**, 3965–3975 (2017).

[20] S. G. Sousa, N. C. Santos, G. Israelian, C. Lovis, M. Mayor, P. B. Silva, and S. Udry, "Spectroscopic characterization of a sample of metal-poor solar-type stars from the HARPS planet search program," Astronomy & Astrophysics **526**, A99 (2011).

[21] Rosario Cosentino, Christophe Lovis, Francesco Pepe, Andrew Collier Cameron, David W. Latham, Emilio Molinari, Stephane Udry, Naidu Bezawada, Martin Black, Andy Born, Nicolas Buchschacher, Dave Charbonneau, Pedro Figueira, Michel Fleury, Alberto Galli, Angus Gallie, Xiaofeng Gao, Adriano Ghedina, Carlos Gonzalez, Manuel Gonzalez, Jose Guerra, David Henry, Keith Horne, Ian Hughes, Dennis Kelly, Marcello Lodi, David Lunney, Charles Maire, Michel Mayor, Giusi Micela, Mark P. Ordway, John Peacock, David Phillips, Giampaolo Piotto, Don Pollacco, Didier Queloz, Ken Rice, Carlos Riverol, Luis Riverol, Jose San Juan, Dimitar Sasselov, Damien Segransan, Alessandro Sozzetti, Danuta Sosnowska, Brian Stobie, Andrew Szentgyorgyi, Andy Vick, and Luc Weber, "Harps-n: the new planet hunter at tng," **8446**, 8446 – 8446 – 20.

[22] Liang Yu, Ian J. M. Crossfield, Joshua E. Schlieder, Molly R. Kosiarek, Adina D. Feinstein, John H. Livingston, Andrew W. Howard, BjÃ¶rn Benneke, Erik A. Petigura, Makennah Bristow, Jessie L. Christiansen, David R. Ciardi, Justin R. Crepp, Courtney D. Dressing, Benjamin J. Fulton, Erica J. Gonzales, Kevin K. Hardegree-Ullman, Thomas Henning, Howard Isaacson, Sébastien Lépine, Arturo O. Martinez, Farisa Y. Morales, and Evan Sinukoff, "Planetary candidates from k2 campaign 16," The Astronomical Journal **156**, 22 (2018).

[23] G. Kovács, S. Zucker, and T. Mazeh, "A box-fitting algorithm in the search for periodic

transits," Astronomy & Astrophysics **391**, 369–377 (2002).

[24] Alexander G. Glenday, Chih-Hao Li, Nicholas Langellier, Guoqing Chang, Li-Jin Chen, Gabor Furesz, Alexander A. Zibrov, Franz Kärtner, David F. Phillips, Dimitar Sasselov, Andrew Szentgyorgyi, and Ronald L. Walsworth, "Operation of a broadband visible-wavelength astro-comb with a high-resolution astrophysical spectrograph," **2**, 250–254.

[25] D. F. Phillips, Alex G. Glenday, Xavier Dumusque, Nicolas Buchschacher, Andrew Collier Cameron, Massimo Cecconi, David Charbonneau, Rosario Cosentino, Adriano Ghedina, RaphÃ€elle Haywood, David W. Latham, Chih-Hao Li, Marcello Lodi, Christophe Lovis, Emilio Molinari, Francesco Pepe, Dimitar Sasselov, Andrew Szentgyorgyi, Stephane Udry, and Ronald L. Walsworth, "An astro-comb calibrated solar telescope to search for the radial velocity signature of venus," Proc.SPIE **9912**, 9912 – 9912 – 8 (2016).

[26] Xavier Dumusque, Alex Glenday, David F. Phillips, Nicolas Buchschacher, Andrew Collier Cameron, Massimo Cecconi, David Charbonneau, Rosario Cosentino, Adriano Ghedina, David W. Latham, Chih-Hao Li, Marcello Lodi, Christophe Lovis, Emilio Molinari, Francesco Pepe, StÃ©phane Udry, Dimitar Sasselov, Andrew Szentgyorgyi, and Ronald Walsworth, "Harps-n observes the sun as a star," **814**, L21 ().

[27] Tilo Steinmetz, Tobias Wilken, Constanza Araujo-Hauck, Ronald Holzwarth, Theodor W. Hänsch, Luca Pasquini, Antonio Manescau, Sandro D'Odorico, Michael T. Murphy, Thomas Kentischer, Wolfgang Schmidt, and Thomas Udem, "Laser frequency combs for astronomical observations," **321**, 1335–1337, http://science.sciencemag.org/content/321/5894/1335.full.pdf .

[28] A.-M. Broomhall, W. J. Chaplin, G. R. Davies, Y. Elsworth, S. T. Fletcher, S. J. Hale, B. Miller, and R. New, "Definitive sun-as-a-star p-mode frequencies: 23 years of bison observations," **396**, L100–L104, /oup/backfile/content$_public/journal/mnrasl/396/1/10.1111/j.1745-3933.2009.00672.x/2/396-1-l100.pdf$.

[29] W. J. Chaplin, H. M. Cegla, C. A. Watson, G. R. Davies, and W. H. Ball, "Filtering solar-like oscillations for exoplanet detection in radial velocity observations," The Astronomical Journal **157**, 163 (2019).

[30] Del Moro, D., "Solar granulation properties derived from three different time series," **428**, 1007–1015.

[31] D. Del Moro, F. Berrilli, T. L. Duvall, Jr., and A. G. Kosovichev, "Dynamics and Structure of Supergranulation," **221**, 23–32.

[32] Hans Kjeldsen, Timothy R. Bedding, R. Paul Butler, JÃžrgen Christensen-Dalsgaard, Laszlo L. Kiss, Chris McCarthy, Geoffrey W. Marcy, Christopher G. Tinney, and Jason T. Wright, "Solar-like oscillations in alpha centauri b," **635**, 1281.

[33] H. M. Cegla, S. Shelyag, C. A. Watson, and M. Mathioudakis, "Stellar surface magneto-convection as a source of astrophysical noise. i. multi-component parameterization of absorption line profiles," **763**, 95.

[34] P. L. Palle, A. Jimenez, F. Perez Hernandez, C. Regulo, T. Roca Cortes, and L. Sanchez, "A measurement of the background solar velocity spectrum," **441**, 952–959.

[35] H. M. Cegla, "The Impact of Stellar Surface Magnetoconvection and Oscillations on the Detection of Temperate, Earth-Mass Planets Around Sun-Like Stars," Geosciences **9**, 114 (2019).

[36] Stephen R. Walton, "Flux tube models of solar plages," The Astrophysical Journal **312**, 909 (1987).

[37] N. Langellier, T. W. Milbourne, D. F. Phillips, R. D. Haywood, S. H. Saar, A. Mortier, L. Malavolta, S. Thompson, A. Collier Cameron, X. Dumusque, H. M. Cegla, D. W. Latham, J. Maldonado, C. A. Watson, N. Buchschacher, M. Cecconi, D. Charbonneau, R. Cosentino, A. Ghedina, M. Gonzalez, C-H. Li, M. Lodi, M. López-Morales, G. Micela, E. Molinari, F. Pepe, E. Poretti, K. Rice, D. Sasselov, A. Sozzetti, S. Udry, and R. L. Walsworth, "Detection limits of low-mass, long-period exoplanets using gaussian processes applied to HARPS-n solar radial velocities," The Astronomical Journal **161**, 287 (2021).

[38] V. Rajpaul, S. Aigrain, and S. Roberts, "Ghost in the time series: no planet for Alpha Cen B," MNRAS **456**, L6–L10 (2016), arXiv:1510.05598 [astro-ph.EP] .

[39] V. Rajpaul, S. Aigrain, M. A. Osborne, S. Reece, and S. Roberts, "A Gaussian process framework for modelling stellar activity signals in radial velocity data," Monthly Notices of the Royal Astronomical Society **452**, 2269–2291 (2015), https://academic.oup.com/mnras/article-pdf/452/3/2269/4912584/stv1428.pdf .

[40] S. Aigrain, F. Pont, and S. Zucker, "A simple method to estimate radial velocity variations due to stellar activity using photometry*," Monthly Notices of the Royal Astronomical Society **419**, 3147–3158 (2012), https://academic.oup.com/mnras/article-pdf/419/4/3147/9506243/mnras0419-3147.pdf .

[41] Xavier Dumusque, Francesco Pepe, Christophe Lovis, Damien Ségransan, Johannes Sahlmann, Willy Benz, François Bouchy, Michel Mayor, Didier Queloz, Nuno Santos, and Stéphane Udry, "An earth-mass planet orbiting centauri b," Nature **491**, 207–211 (2012).

[42] Julián D. Alvarado-Gómez, Gaitee A. J. Hussain, Jeremy J. Drake, Jean-François Donati, Jorge Sanz-Forcada, Beate Stelzer, Ofer Cohen, Eliana M. Amazo-Gómez, Jason H. Grunhut, Cecilia Garraffo, Sofia P. Moschou, James Silvester, and Mary E. Oksala, "Far beyond the sun – i. the beating magnetic heart in horologium," Monthly Notices of the Royal Astronomical Society **473**, 4326–4338 (2017).

[43] R. Cosentino, Christophe Lovis, Francesco Pepe, Andrew Collier Cameron, David W. Latham, Emilio Molinari, Stephane Udry, Naidu Bezawada, Nicolas Buchschacher, Pedro Figueira, Michel Fleury, Adriano Ghedina, Alexander G. Glenday, Manuel Gonzalez, Jose Guerra, David Henry, Ian Hughes, Charles Maire, Fatemeh Motalebi, and David Forrest Phillips, "HARPS-N @ TNG, two year harvesting data: performances and results," in *Ground-based and Airborne Instrumentation for Astronomy V*, Vol. 9147, edited by Suzanne K. Ramsay, Ian S. McLean, and Hideki Takami, International Society for Optics and Photonics (SPIE, 2014) pp. 2658 – 2669.

[44] A. Baranne, D. Queloz, M. Mayor, G. Adrianzyk, G. Knispel, D. Kohler, D. Lacroix, J.-P. Meunier, G. Rimbaud, and A. Vin, "ELODIE: A spectrograph for accurate radial velocity measurements." **119**, 373–390 (1996).

[45] D. Sosnowska, M. Lodi, X. Gao, N. Buchschacher, A. Vick, J. Guerra, M. Gonzalez, D. Kelly, C. Lovis, F. Pepe, E. Molinari, A. C. Cameron, D. Latham, and S. Udry, "Harps-n: software path from the observation block to the image," Proc.SPIE **8451**, 8451 – 8451 – 12 (2012).

[46] J. D. Giorgini, D. K. Yeomans, A. B. Chamberlin, P. W. Chodas, R. A. Jacobson, M. S. Keesey, J. H. Lieske, S. J. Ostro, E. M. Standish, and R. N. Wimberly, "JPL's On-Line Solar System Data Service," in *AAS/Division for Planetary Sciences Meeting Abstracts #28*, Bulletin of the American Astronomical Society, Vol. 28, p. 1158.

[47] A Collier Cameron, A Mortier, D Phillips, X Dumusque, R D Haywood, N Langellier, C A Watson, H M Cegla, J Costes, D Charbonneau, A Coffinet, D W Latham, M Lopez-Morales, L Malavolta, J Maldonado, G Micela, T Milbourne, E Molinari, S H Saar, S Thompson, N Buchschacher, M Cecconi, R Cosentino, A Ghedina, A Glenday, M Gonzalez, C-H Li, M Lodi, C Lovis, F Pepe, E Poretti, K Rice, D Sasselov, A Sozzetti, A Szentgyorgyi, S Udry, and R Walsworth, "Three years of Sun-as-a-star radial-velocity observations on the approach to solar minimum," Monthly Notices of the Royal Astronomical Society **487**, 1082–1100 (2019), http://oup.prod.sis.lan/mnras/article-pdf/487/1/1082/28753321/stz1215.pdf .

[48] Isaacson, H. and Fischer, D., "CHROMOSPHERIC ACTIVITY AND JITTER MEASUREMENTS FOR 2630 STARS ON THE CALIFORNIA PLANET SEARCH," The Astrophysical Journal **725**, 875–885 (2010).

[49] A.H. Vaughan, G.W. Preston, and O.C. Wilson, "Flux measurements of ca ii h and k emission," PASP **90**, 267–274 (1978).

[50] R. W. Noyes, L. W. Hartmann, S. L. Baliunas, D. K. Duncan, and A. H. Vaughan, "Rotation, convection, and magnetic activity in lower main-sequence stars," **279**, 763–777 (1984).

[51] Gary Rottman, "The sorce mission," Solar Physics **230**, 7–25 (2005).

[52] R. D. Haywood, A. Collier Cameron, D. Queloz, S. C. C. Barros, M. Deleuil, R. Fares, M. Gillon, A. F. Lanza, C. Lovis, C. Moutou, F. Pepe, D. Pollacco, A. Santerne, D. Ségransan, and Y. C. Unruh, "Planets and stellar activity: hide and seek in the CoRoT-7 system," MNRAS **443**, 2517–2531 (2014), arXiv:1407.1044 [astro-ph.EP] .

[53] J. Schou, P. H. Scherrer, R. I. Bush, R. Wachter, S. Couvidat, M. C. Rabello-Soares, R. S. Bogart, J. T. Hoeksema, Y. Liu, T. L. Duvall, D. J. Akin, B. A. Allard, J. W. Miles, R. Rairden, R. A. Shine, T. D. Tarbell, A. M. Title, C. J. Wolfson, D. F. Elmore, A. A. Norton, and S. Tomczyk, "Design and ground calibration of the helioseismic and magnetic imager (hmi) instrument on the solar dynamics observatory (sdo)," Solar Physics **275**, 229–259 (2012).

[54] W. Dean Pesnell, B. J. Thompson, and P. C. Chamberlin, "The solar dynamics observatory (sdo)," Solar Physics **275**, 3–15 (2012).

[55] S. Couvidat, J. Schou, J. T. Hoeksema, R. S. Bogart, R. I. Bush, T. L. Duvall, Y. Liu, A. A. Norton, and P. H. Scherrer, "Observables processing for the helioseismic and magnetic imager

instrument on the solar dynamics observatory," Solar Physics **291**, 1887–1938 (2016).

[56] Michel Mayor and Didier Queloz, "A jupiter-mass companion to a solar-type star," **378**, 335–359.

[57] D. A. Fischer, G. Anglada-Escude, P. Arriagada, R. V. Baluev, J. L. Bean, F. Bouchy, L. A. Buchhave, T. Carroll, A. Chakraborty, J. R. Crepp, R. I. Dawson, S. A. Diddams, X. Dumusque, J. D. Eastman, M. Endl, P. Figueira, E. B. Ford, D. Foreman-Mackey, P. Fournier, G. Fűrész, B. S. Gaudi, P. C. Gregory, F. Grundahl, A. P. Hatzes, G. Hébrard, E. Herrero, D. W. Hogg, A. W. Howard, J. A. Johnson, P. Jorden, C. A. Jurgenson, D. W. Latham, G. Laughlin, T. J. Loredo, C. Lovis, S. Mahadevan, T. M. McCracken, F. Pepe, M. Perez, D. F. Phillips, P. P. Plavchan, L. Prato, A. Quirrenbach, A. Reiners, P. Robertson, N. C. Santos, D. Sawyer, D. Segransan, A. Sozzetti, T. Steinmetz, A. Szentgyorgyi, S. Udry, J. A. Valenti, S. X. Wang, R. A. Wittenmyer, D. A. Wright, J. T.Fischer, G. Anglada-Escude, P. Arriagada, R. V. Baluev, J. L. Bean, F. Bouchy, L. A. Buchhave, T. Carroll, A. Chakraborty, J. R. Crepp, R. I. Dawson, S. A. Diddams, X. Dumusque, J. D. Eastman, M. Endl, P. Figueira, E. B. Ford, D. Foreman-Mackey, P. Fournier, G. Fűrész, B. S. Gaudi, P. C. Gregory, F. Grundahl, A. P. Hatzes, G. Hébrard, E. Herrero, D. W. Hogg, A. W. Howard, J. A. Johnson, P. Jorden, C. A. Jurgenson, D. W. Latham, G. Laughlin, T. J. Loredo, C. Lovis, S. Mahadevan, T. M. McCracken, F. Pepe, M. Perez, D. F. Phillips, P. P. Plavchan, L. Prato, A. Quirrenbach, A. Reiners, P. Robertson, N. C. Santos, D. Sawyer, D. Segransan, A. Sozzetti, T. Steinmetz, A. Szentgyorgyi, S. Udry, J. A. Valenti, S. X. Wang, R. A. Wittenmyer, and J. T. Wright, "State of the field: Extreme precision radial velocities," PASP **128**, 066001 (2016).

[58] M. Auvergne, P. Bodin, L. Boisnard, J.-T. Buey, S. Chaintreuil, G. Epstein, M. Jouret, T. Lam-Trong, P. Levacher, A. Magnan, R. Perez, P. Plasson, J. Plesseria, G. Peter, M. Steller, D. Tiphène, A. Baglin, P. Agogué, T. Appourchaux, D. Barbet, T. Beaufort, R. Bellenger, R. Berlin, P. Bernardi, D. Blouin, P. Boumier, F. Bonneau, R. Briet, B. Butler, R. Cautain, F. Chiavassa, V. Costes, J. Cuvilho, V. Cunha-Parro, F. de Oliveira Fialho, M. Decaudin, J.-M. Defise, S. Djalal, A. Docclo, R. Drummond, O. Dupuis, G. Exil, C. Fauré, A. Gaboriaud, P. Gamet, P. Gavalda, E. Grolleau, L. Gueguen, V. Guivarc'h, P. Guterman, J. Hasiba, G. Huntzinger, H. Hustaix, C. Imbert, G. Jeanville, B. Johlander, L. Jorda, P. Journoud, F. Karioty, L. Kerjean, L. Lafond, V. Lapeyrere, P. Landiech, T. Larqué, P. Laudet, J. Le Merrer, L. Leporati, B. Leruyet, B. Levieuge, A. Llebaria, L. Martin, E. Mazy, J.-M. Mesnager, J.-P. Michel, J.-P. Moalic, W. Monjoin, D. Naudet, S. Neukirchner, K. Nguyen-Kim, M. Ollivier, J.-L. Orcesi, H. Ottacher, A. Oulali, J. Parisot, S. Perruchot, A. Piacentino, L. Pinheiro da Silva, J. Platzer, B. Pontet, A. Pradines, C. Quentin, U. Rohbeck, G. Rolland, F. Rollenhagen, R. Romagnan, N. Russ, R. Samadi, R. Schmidt, N. Schwartz, I. Sebbag, H. Smit, W. Sunter, M. Tello, P. Toulouse, B. Ulmer, O. Vandermarcq, E. Vergnault, R. Wallner, G. Waultier, and P. Zanatta, "The CoRoT satellite in flight: description and performance," **506**, 411–424, 0901.2206 .

[59] Li Zeng and Dimitar Sasselov, "A detailed model grid for solid planets from 0.1 through 100 earth masses," **125**, 227.

[60] F. A. Pepe, Stefano Cristiani, Rafael Rebolo Lopez, Nuno C. Santos, Antonio Amorim, Gerardo Avila, Willy Benz, Piercarlo Bonifacio, Alexandre Cabral, Pedro Carvas, Roberto Cirami, João Coelho, Maurizio Comari, Igor Coretti, Vincenzo De Caprio, Hans Dekker,

Bernard Delabre, Paolo Di Marcantonio, Valentina D'Odorico, Michel Fleury, Ramón García, José Miguel Herreros Linares, Ian Hughes, Olaf Iwert, Jorge Lima, Jean-Louis Lizon, Gaspare Lo Curto, Christophe Lovis, Antonio Manescau, Carlos Martins, Denis Mégevand, André Moitinho, Paolo Molaro, Mario Monteiro, Manuel Monteiro, Luca Pasquini, Christoph Mordasini, Didier Queloz, José L. Rasilla, José M. Rebordão, Samuel Santana Tschudi, Paolo Santin, Danuta Sosnowska, Paolo Spanò, Fabio Tenegi, Stéphane Udry, Eros Vanzella, Matteo Viel, Maria Rosa Zapatero Osorio, and Filippo Zerbi, "ESPRESSO: the Echelle spectrograph for rocky exoplanets and stable spectroscopic observations," in *Proc.SPIE*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7735 (2010) p. 77350F.

[61] Steven H. Saar and Robert A. Donahue, "Activity-related radial velocity variation in cool stars," **485**, 319.

[62] C. J. Schrijver and C. Zwaan, *Solar and Stellar Magnetic Activity*, Cambridge Astrophysics (Cambridge University Press).

[63] X. Dumusque, S. Udry, C. Lovis, N. C. Santos, and M. J. P. F. G. Monteiro, "Planetary detection limits taking into account stellar noise - i. observational strategies to reduce stellar oscillation and granulation effects," **525**, A140 ().

[64] N. Meunier, A.-M. Lagrange, S. Borgniet, and M. Rieutord, "Using the sun to estimate earth-like planet detection capabilities - vi. simulation of granulation and supergranulation radial velocity and photometric time series," **583**, A118.

[65] J.W. Harvey, F. Hill, J.R. Kennedy, J.W. Leibacher, and W.C. Livingston, "The global oscillation network group (gong)," **8**, 117 – 120.

[66] M. Fligge, S. K. Solanki, and Y. C. Unruh, "Modelling irradiance variations from the surface distribution of the solar magnetic field," **353**, 380–388 (2000).

[67] J. Löhner-Böttcher and R. Schlichenmaier, "Correlations between sunspots and their moat flows," **551**, A105.

[68] A. Reiners, U. Lemke, F. Bauer, B. Beeck, and P. Huke, "Radial velocity observations of the 2015 mar. 20 eclipse - a benchmark rossiter-mclaughlin curve with zero free parameters," **595**, A26.

[69] J. T. Hoeksema, C. S. Baldner, R. I. Bush, J. Schou, and P. H. Scherrer, "On-orbit performance of the helioseismic and magnetic imager instrument onboard the solar dynamics observatory," **293**, 45.

[70] J. L. Linksy and E. H. Avrett, "The solar h and k lines," Publications of the Astronomical Society of the Pacific **82**, 169 (1970).

[71] O. C. Wilson, "Flux measurements at the centers of stellar h- and k-lines," Astrophysical Journal **153**, 221 (1968).

[72] A. I. Shapiro, Solanki, S. K., Krivova, N. A., Schmutz, W. K., Ball, W. T., Knaack, R., Rozanov, E. V., and Unruh, Y. C., "Variability of sun-like stars: reproducing observed photometric trends," **569**, A38.

[73] D. Dravins, L. Lindegren, and A. Nordlund, "Solar granulation - Influence of convection on spectral line asymmetries and wavelength shifts," **96**, 345–364.

[74] David F. Gray, "The third signature of stellar granulation," **697**, 1032.

[75] David F. Gray and Benjamin Oostra, "The solar-flux third granulation signature," **852**, 42.

[76] X. Dumusque, I. Boisse, and N. C. Santos, "SOAP 2.0: A Tool to Estimate the Photometric and Radial Velocity Variations Induced by Stellar Spots and Plages," ApJ **796**, 132 (2014), arXiv:1409.3594 [astro-ph.SR] .

[77] A.-M. Lagrange, M. Desort, and N. Meunier, "Using the Sun to estimate Earth-like planets detection capabilities . I. Impact of cold spots," **512**, A38, 1001.1449 .

[78] A. Collier Cameron, D. Pollacco, R. A. Street, T. A. Lister, R. G. West, D. M. Wilson, F. Pont, D. J. Christian, W. I. Clarkson, B. Enoch, A. Evans, A. Fitzsimmons, C. A. Haswell, C. Hellier, S. T. Hodgkin, K. Horne, J. Irwin, S. R. Kane, F. P. Keenan, A. J. Norton, N. R. Parley, J. Osborne, R. Ryans, I. Skillen, and P. J. Wheatley, "A fast hybrid algorithm for exoplanetary transit searches," **373**, 799–810, astro-ph/0609418 .

[79] M. L. Palumbo, R. D. Haywood, S. H. Saar, *et al.*, "Radial-Velocity Signatures of Magnetic Features on the Sun Observed as a Star," ().

[80] M. L. Palumbo, R. D. Haywood, S. H. Saar, *et al.*, "," ().

[81] David H. Hathaway, "The solar cycle," **12**, 4.

[82] N. Meunier, "Solar chromospheric emission and magnetic structures from plages to intranetwork: Contribution of the very quiet sun," **615**, A87.

[83] Martin F. Woodard and Robert W. Noyes, "Change of solar oscillation eigenfrequencies with the solar cycle," Nature **318**, 449–450 (1985).

[84] Y. Elsworth, R. Howe, G. R. Isaak, C. P. McLeod, and R. New, "Variation of low-order acoustic solar oscillations over the solar cycle," Nature **345**, 322–324 (1990).

[85] K. G. Libbrecht and M. F. Woodard, "Solar-cycle effects on solar oscillation frequencies," Nature **345**, 779–782 (1990).

[86] Thomas L. Duvall, John M. Wilcox, Leif Svalgaard, Philip H. Scherrer, and Patrick S. McIntosh, "Comparison of h synoptic charts with the large-scale solar magnetic field as observed at stanford," Solar Physics **55**, 63–68 (1977).

[87] N. A. Krivova, L. Balmaceda, and S. K. Solanki, "Reconstruction of solar total irradiance since 1700 from the surface magnetic flux," Astronomy & Astrophysics **467**, 335–346 (2007).

[88] L. E. A. Vieira and S. K. Solanki, "Evolution of the solar magnetic flux on time scales of years to millenia," Astronomy and Astrophysics **509**, A100 (2010).

[89] S. J. Hale, R. Howe, W. J. Chaplin, G. R. Davies, and Y. P. Elsworth, "Performance of the birmingham solar-oscillations network (BiSON)," Solar Physics **291**, 1–28 (2015).

[90] Ilya G. Usoskin, Rainer Arlt, Eleanna Asvestari, Ed Hawkins, Maarit KÃ€pylÃ€, Gennady A. Kovaltsov, Natalie Krivova, Michael Lockwood, Kalevi Mursula, Jezebel O'Reilly, Matthew Owens, Chris J. Scott, Dmitry D. Sokoloff, Sami K. Solanki, Willie Soon, and José M. Vaquero, "The maunder minimum (1645–1715) was indeed a grand minimum: A reassessment of multiple datasets," Astronomy & Astrophysics **581**, A95 (2015).

[91] N. Meunier, Lagrange, A.-M., and Borgniet, S., "A new method of correcting radial velocity time series for inhomogeneous convection," A&A **607**, A6 (2017).

[92] R. D. Haywood, T. W. Milbourne, S. H. Saar, A. Mortier, D. Phillips, D. Charbonneau, A. Collier Cameron, H. M. Cegla, N. Meunier, and M. L. Palumbo III, "Unsigned magnetic flux as a proxy for radial-velocity variations in sun-like stars," (2020), arXiv:2005.13386 [astro-ph.SR] .

[93] T. W. Milbourne, R. D. Haywood, D. F. Phillips, S. H. Saar, H. M. Cegla, A. C. Cameron, J. Costes, X. Dumusque, N. Langellier, D. W. Latham, J. Maldonado, L. Malavolta, A. Mortier, M. L. Palumbo III, S. Thompson, C. A. Watson, F. Bouchy, N. Buchschacher, M. Cecconi, D. Charbonneau, R. Cosentino, A. Ghedina, A. G. Glenday, M. Gonzalez, C-H. Li, M. Lodi, M. López-Morales, C. Lovis, M. Mayor, G. Micela, E. Molinari, F. Pepe, G. Piotto, K. Rice, D. Sasselov, D. Ségransan, A. Sozzetti, A. Szentgyorgyi, S. Udry, and R. L. Walsworth, "HARPS-n solar RVs are dominated by large, bright magnetic regions," The Astrophysical Journal **874**, 107 (2019).

[94] Pepe, F., Cristiani, S., Rebolo, R., Santos, N. C., Dekker, H., Cabral, A., Di Marcantonio, P., Figueira, P., Lo Curto, G., Lovis, C., Mayor, M., Mégevand, D., Molaro, P., Riva, M., Zapatero Osorio, M. R., Amate, M., Manescau, A., Pasquini, L., Zerbi, F. M., Adibekyan, V., Abreu, M., Affolter, M., Alibert, Y., Aliverti, M., Allart, R., Allende Prieto, C., Álvarez, D., Alves, D., Avila, G., Baldini, V., Bandy, T., Barros, S. C. C., Benz, W., Bianco, A., Borsa, F., Bourrier, V., Bouchy, F., Broeg, C., Calderone, G., Cirami, R., Coelho, J., Conconi, P., Coretti, I., Cumani, C., Cupani, G., D´Odorico, V., Damasso, M., Deiries, S., Delabre, B., Demangeon, O. D. S., Dumusque, X., Ehrenreich, D., Faria, J. P., Fragoso, A., Genolet, L., Genoni, M., Génova Santos, R., González Hernández, J. I., Hughes, I., Iwert, O., Kerber, F., Knudstrup, J., Landoni, M., Lavie, B., Lillo-Box, J., Lizon, J.-L., Maire, C., Martins, C. J. A. P., Mehner, A., Micela, G., Modigliani, A., Monteiro, M. A., Monteiro, M. J. P. F. G., Moschetti, M., Murphy, M. T., Nunes, N., Oggioni, L., Oliveira, A., Oshagh, M., Pallé, E., Pariani, G., Poretti, E., Rasilla, J. L., Rebordão, J., Redaelli, E. M., Santana Tschudi, S., Santin, P., Santos, P., Ségransan, D., Schmidt, T. M., Segovia, A., Sosnowska, D., Sozzetti, A., Sousa, S. G., Spanò, P., Suárez Mascareño, A., Tabernero, H., Tenegi, F., Udry, S., and Zanutta, A., "Espresso at vlt - on-sky performance and first results," A&A **645**, A96 (2021).

[95] X. Dumusque, M. Cretignier, D. Sosnowska, N. Buchschacher, C. Lovis, D. F. Phillips, F. Pepe, F. Alesina, L. A. Buchhave, J. Burnier, M. Cecconi, H. M. Cegla, R. Cloutier, A. Collier Cameron, R. Cosentino, A. Ghedina, M. Gonzalez, R. D. Haywood, D. W. Latham, M. Lodi, M. Lopez-Morales, J. Maldonado, L. Malavolta, G. Micela, E. Molinari, A. Mortier, H. Perez Ventura, M. Pinamonti, E. Poretti, K. Rice, L. Riverol, C. Riverol, J. San Juan, D. Segransan, A. Sozzetti, S. J. Thompson, S. Udry, and T. G Wilson, "Three years of harps-n high-resolution spectroscopy and precise radial velocity data for the sun," (2020), arXiv:2009.01945 [astro-ph.SR] .

[96] X. Dumusque, A. Glenday, D. F. Phillips, N. Buchschacher, A. Collier Cameron, M. Cecconi, D. Charbonneau, R. Cosentino, A. Ghedina, D. W. Latham, C.-H. Li, M. Lodi, C. Lovis, E. Molinari, F. Pepe, S. Udry, D. Sasselov, A. Szentgyorgyi, and R. Walsworth, "HARPS-N Observes the Sun as a Star," ApJl **814**, L21 (2015), arXiv:1511.02267 [astro-ph.EP] .

[97] Ricky Egeland, Willie Soon, Sallie Baliunas, Jeffrey C. Hall, Alexei A. Pevtsov, and Luca Bertello, "THE MOUNT WILSON OBSERVATORYS-INDEX OF THE SUN," The Astrophysical Journal **835**, 25 (2017).

[98] Jean Pierre Rozelot, Alexander Kosovichev, and Ali Kilcik, "SOLAR RADIUS VARIATIONS: AN INQUISITIVE WAVELENGTH DEPENDENCE," The Astrophysical Journal **812**, 91 (2015).

[99] A. I. Shapiro, Solanki, S. K., Krivova, N. A., Yeo, K. L., and Schmutz, W. K., "Are solar brightness variations faculae- or spot-dominated?" A&A **589**, A46 (2016).

[100] G. Cybenko, "Approximation by superpositions of a sigmoidal function," Mathematics of Control, Signals and Systems **2**, 303–314 (1989).

[101] Geoffrey E. Hinton, "Connectionist learning procedures," Artificial Intelligence **40**, 185–234 (1989).

[102] G van Rossum, *Python tutorial*, Tech. Rep. CS-R9526 (Centrum voor Wiskunde en Informatica (CWI), Amsterdam, 1995).

[103] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research **12**, 2825–2830 (2011).

[104] Zoe L. de Beurs, Andrew Vanderburg, Christopher J. Shallue, Xavier Dumusque, Andrew Collier Cameron, Lars A. Buchhave, Rosario Cosentino, Adriano Ghedina, RaphaÃ«lle D. Haywood, Nicholas Langellier, David W. Latham, Mercedes López-Morales, Michel Mayor, Giusi Micela, Timothy W. Milbourne, Annelies Mortier, Emilio Molinari, Francesco Pepe, David F. Phillips, Matteo Pinamonti, Giampaolo Piotto, Ken Rice, Dimitar Sasselov, Alessandro Sozzetti, StÃ©phane Udry, and Christopher A. Watson, "Identifying exoplanets with deep learning. iv. removing stellar activity signals from radial velocity measurements using neural networks," (2020), arXiv:2011.00003 [astro-ph.EP] .

[105] M. Miklos, T. W. Milbourne, R. D. Haywood, D. F. Phillips, S. H. Saar, N. Meunier, H. M. Cegla, X. Dumusque, N. Langellier, J. Maldonado, L. Malavolta, A. Mortier, S. Thompson, C. A. Watson, M. Cecconi, R. Cosentino, A. Ghedina, C-H. Li, M. López-Morales, E. Molinari, Ennio Poretti, D. Sasselov, A. Sozzetti, and R. L. Walsworth, "Testing the spectroscopic extraction of suppression of convective blueshift," The Astrophysical Journal **888**, 117 (2020).

[106] J. L. Leroy, "Contributions à l'étude de la polarisation de la lumière solaire," Annales d'Astrophysique **25**, 127 (1962).

[107] M. J. Stift and F. Leone, "Magnetic intensification of spectral lines," **398**, 411–421 (2003).

[108] A. Reiners, D. Shulyak, G. Anglada-Escudé, S. V. Jeffers, J. Morin, M. Zechmeister, O. Kochukhov, and N. Piskunov, "Radial velocity signatures of zeeman broadening," Astronomy & Astrophysics **552**, A103 (2013).

[109] Lars A. Buchhave, David W. Latham, Anders Johansen, Martin Bizzarro, Guillermo Torres, Jason F. Rowe, Natalie M. Batalha, William J. Borucki, Erik Brugamyer, Caroline Caldwell, Stephen T. Bryson, David R. Ciardi, William D. Cochran, Michael Endl, Gilbert A. Esquerdo, Eric B. Ford, John C. Geary, Ronald L. Gilliland, Terese Hansen, Howard Isaacson, John B. Laird, Philip W. Lucas, Geoffrey W. Marcy, Jon A. Morse, Paul Robertson, Avi Shporer, Robert P. Stefanik, Martin Still, and Samuel N. Quinn, "An abundance of small exoplanets around stars with a wide range of metallicities," **486**, 375–377 (2012).

[110] Herrero, Enrique, Ribas, Ignasi, Jordi, Carme, Morales, Juan Carlos, Perger, Manuel, and Rosich, Albert, "Modelling the photosphere of active stars for planet detection and characterization," A&A **586**, A131 (2016).

[111] F. Motalebi, Udry, S., Gillon, M., Lovis, C., Ségransan, D., Buchhave, L. A., Demory, B. O., Malavolta, L., Dressing, C. D., Sasselov, D., Rice, K., Charbonneau, D., Collier Cameron, A., Latham, D., Molinari, E., Pepe, F., Affer, L., Bonomo, A. S., Cosentino, R., Dumusque, X., Figueira, P., Fiorenzano, A. F. M., Gettel, S., Harutyunyan, A., Haywood, R. D., Johnson, J., Lopez, E., Lopez-Morales, M., Mayor, M., Micela, G., Mortier, A., Nascimbeni, V., Philips, D., Piotto, G., Pollacco, D., Queloz, D., Sozzetti, A., Vanderburg, A., and Watson, C. A., "The harps-n rocky planet search - i. hd134a transiting rocky planet in a multi-planet system at 6.5 pc from the sun," A&A **584**, A72 (2015).

[112] Aakash Ravi, David F. Phillips, Matthias Beck, Leopoldo L. Martin, Massimo Cecconi, Adriano Ghedina, Emilio Molinari, Albrecht Bartels, Dimitar Sasselov, Andrew Szentgyorgyi, and Ronald L. Walsworth, "Astro-comb calibrator and spectrograph characterization using a turnkey laser frequency comb," Journal of Astronomical Telescopes, Instruments, and Systems **3**, 1 (2017).