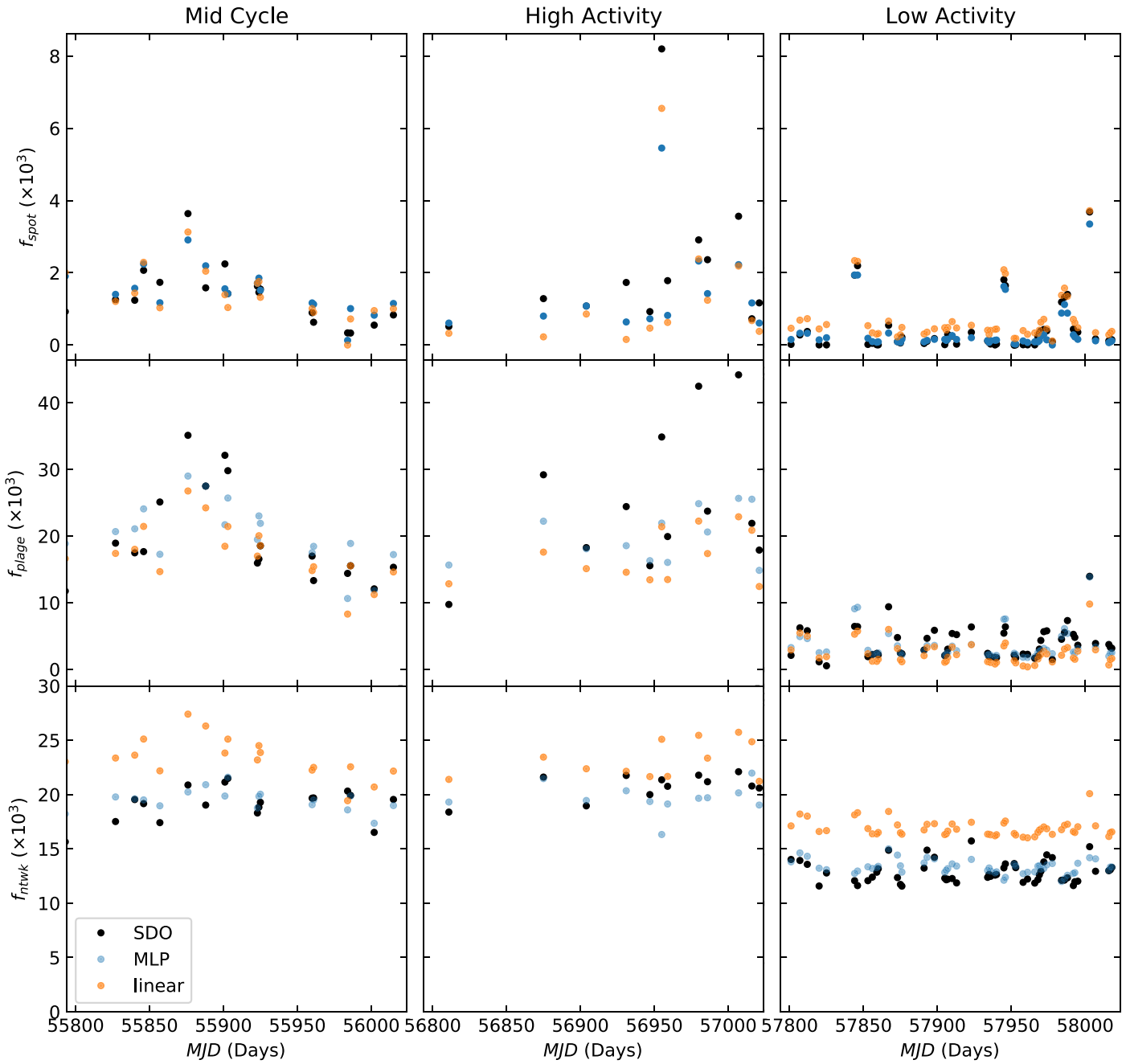


**Figure 5.** 230 day subsets of the time series of the SDO/HMI-observed magnetic filling factors (black), along with the MLP (blue) and linear (orange) estimates derived from the S-index and TSI. Three subsets are shown, taken during the middle of the stellar cycle (left), during solar maximum (middle), and approaching solar minimum (right). Both techniques successfully reproduce $f_{\rm spot}$, $f_{\rm plage}$, and $f_{\rm ntwk}$, with especially good performance at solar minimum.

**Table 3**
Pearson Correlation Coefficients between HMI-Derived Estimate of the Suppression of Convective Blueshift, $\Delta v_{\rm conv}$ and the Activity-driven RVs Derived from Equation (13)

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

**Note.** 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 MH19.

**Table 4**
Pearson Correlation Coefficients between HMI-derived Estimate of the Photometric Velocity Shift, $\Delta v_{\rm phot}$ and the Activity-driven RVs Derived from Equation (14)

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

**Note.** The relatively high correlation coefficients indicate that the spot filling factors successfully estimate the photometric RV shifts, as expected from MH19.

**Table 5**
RMS RV Residuals from Several Models and Methods

| | rms ($m\,s^{-1}$) |
|---|---|
| Full solar data set | 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 |

**Note.** Using our estimates of $f_{spot}$, $f_{plage}$, and $f_{ntwk}$ in Equation (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.

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 Figure 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 MH19, 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 $\Delta v_{conv}$ and $\Delta v_{phot}$, respectively.

### 5.2. Direct MLP Modelling of Solar 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 de Beurs et al. (2020), 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 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 5) 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 (Miklos et al. 2020). 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., Leroy 1962; Stift & Leone 2003), has an RV signal that depends on the overall filling factor, as well as a given line's wavelength, effective Landé $g$ value, and magnetic field strength (Reiners et al. 2013). 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.

**Table 6**
Pearson Correlation Coefficients between HMI Ground Truth Filling Factors and the Linear and MLP Estimates for Each Class of Filling Factor

| | Spots | Plage | Network |
|---|---|---|---|
| Linear | 0.81 | 0.87 | 0.81 |
| MLP | 0.85 | 0.87 | 0.89 |

The direct MLP fit to the solar RVs and its residuals are plotted in Figure 6. The effects of HARPS-N cryostat cold plate warm-ups, discussed in Collier Cameron et al. (2019), Dumusque et al. (2021), 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.

### 5.3. 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 (Buchhave et al. 2012), 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 temperature as a prior. The stellar distance may be straightforwardly determined

**Table 7**
Best-Fit Coefficients for Fitting Equation (15) to the HARPS-N Solar RVs Using SDO Filling Factors, Linear Estimates of the Filling Factors, and MLP Estimates of the Filling Factors

| 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 | $7.7 \pm 1.2$ | $165 \pm 76$ | $244 \pm 15$ | $52 \pm 27$ | $-2.0 \pm 0.3$ |
| Linear | $7.1 \pm 1.3$ | $470 \pm 80$ | $281 \pm 11$ | $0^*$ | $-1.32 \pm 0.08$ |
| MLP | $7.5 \pm 1.4$ | $479 \pm 218$ | $242 \pm 52$ | $2 \pm 90$ | $-1.3 \pm 1.2$ |

**Note.** Note that the estimates of $f_{\rm plage}$ and $f_{\rm ntwk}$ 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.

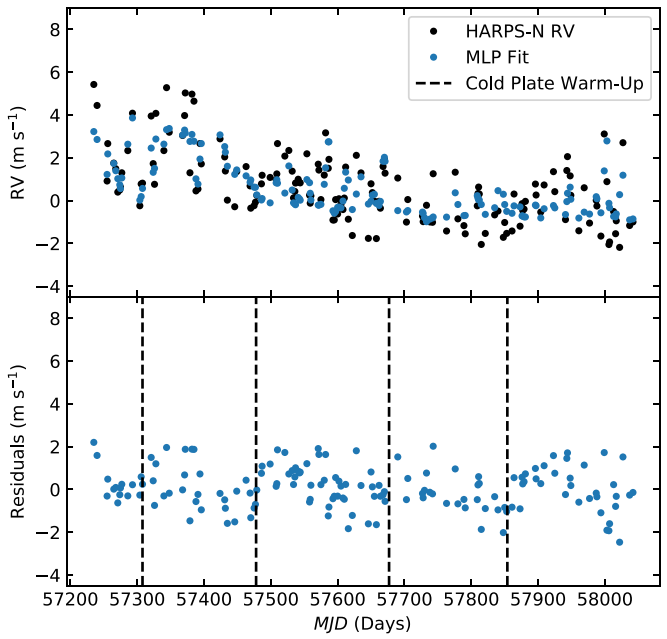


**Figure 6.** MLP fit to the HARPS-N solar telescope data. HARPS-N RVs are shown in black and MLP estimates of the RVs are in blue. Fit residuals are shown in the bottom panels: HARPS-N cryostat warm-up dates (*see text*) are indicated with black dashed lines.

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 Equations (3) and (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 $f_{\rm plage}$ requires observations near stellar minimum, modeling RV variations only requires time series which are proportional to $f_{\rm spot}$, $f_{\rm plage}$ and $f_{\rm ntwk}$—in this case, this offset is unimportant, and no additional constraint is placed on the stellar observations. Furthermore, the values of $\Delta T_{\rm spot}$, $\Delta T_{\rm plage}$ and $\Delta T_{\rm ntwk}$ may be absorbed into the fit coefficients in Equations (3), (5), and (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 (Dumusque et al. 2014), StarSim (Herrero et al. 2016), 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.

## 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 an 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 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 $\sqrt{(1.10\ \mathrm{m\ s^{-1}})^2 - (0.91\ \mathrm{m\ s^{-1}})^2} = 0.62\ \mathrm{m\ s^{-1}}$ 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 an 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 $f_{\rm spot}$, $f_{\rm plage}$, and $f_{\rm ntwk}$ 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.

This work was primarily supported by NASA award number NNX16AD42G and the Smithsonian Institution. The solar telescope used in these observations was built and maintained with support from the Smithsonian Astrophysical Observatory, the Harvard Origins of Life Initiative, and the TNG.

This material is also based upon work supported by NASA under grants No. NNX15AC90G and NNX17AB59G issued through the Exoplanets Research Program. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant Agreement No. 313014 (ETAEARTH).

The HARPS-N project has been funded by the Prodex Program of the Swiss Space Office (SSO), the Harvard University Origins of Life Initiative (HUOLI), the Scottish Universities Physics Alliance (SUPA), the University of Geneva, the Smithsonian Astrophysical Observatory (SAO), and the Italian National Astrophysical Institute (INAF), the University of St. Andrews, Queen's University Belfast, and the University of Edinburgh.

We would like to acknowledge the excellent discussions and scientific input by members of the International Team, "Toward Earth-like Alien Worlds: Know Thy Star, Know Thy Planet", supported by the International Space Science Institute (ISSI, Bern).

This work was partially performed under contract with the California Institute of Technology (Caltech)/Jet Propulsion Laboratory (JPL) funded by NASA through the Sagan Fellowship Program executed by the NASA Exoplanet Science Institute (R.D.H.).

S.H.S. is grateful for support from NASA Heliophysics LWS grant NNX16AB79G and NASA XRP grant 80NSSC21K0607.

A.M. acknowledges support from the senior Kavli Institute Fellowships.

A.C.C. acknowledges support from the Science and Technology Facilities Council (STFC) consolidated grant number ST/R000824/1.

X.D. is grateful to the Branco-Weiss Fellowship for continuous support. This project has received funding from the European Research Council (ERC) under the European Unions' Horizon 2020 research and innovation program (grant agreement No. 851555).

C.A.W. acknowledges support from Science and Technology Facilities Council grant ST/P000312/1.

H.M.C. acknowledges financial support from the National Centre for Competence in Research (NCCR) Planets, supported by the Swiss National Science Foundation (SNSF), as well as UK Research and Innovation Future Leaders Fellowship grant No. MR/S035214/1.

We thank the entire TNG staff for their continued support of the solar telescope project at HARPS-N.

*Facility*: TNG: HARPS-N (Cosentino et al. 2014), Solar Telescope (Dumusque et al. 2015; Phillips et al. 2016).

*Software*: HARPS-N DRS, Python (van Rossum 1995): `NumPy` (Harris et al. 2020), `SciPy` (Virtanen et al. 2020), `scikit-learn` (Pedregosa et al. 2011).

## ORCID iDs

T. W. Milbourne https://orcid.org/0000-0001-5446-7712
D. F. Phillips https://orcid.org/0000-0001-5132-1339
N. Langellier https://orcid.org/0000-0003-2107-3308
A. Mortier https://orcid.org/0000-0001-7254-4363
R. D. Haywood https://orcid.org/0000-0001-9140-3574
S. H. Saar https://orcid.org/0000-0001-7032-8480
H. M. Cegla https://orcid.org/0000-0001-8934-7315
A. Collier Cameron https://orcid.org/0000-0002-8863-7828
X. Dumusque https://orcid.org/0000-0002-9332-2011
D. W. Latham https://orcid.org/0000-0001-9911-7388
L. Malavolta https://orcid.org/0000-0002-6492-2085
J. Maldonado https://orcid.org/0000-0002-2218-5689
S. Thompson https://orcid.org/0000-0002-8039-194X
L. A. Buchhave https://orcid.org/0000-0003-1605-5666
M. Cecconi https://orcid.org/0000-0001-5701-2529
R. Cosentino https://orcid.org/0000-0003-1784-1431
A. Ghedina https://orcid.org/0000-0003-4702-5152
M. Lodi https://orcid.org/0000-0002-5432-9659
M. López-Morales https://orcid.org/0000-0003-3204-8183
A. Sozzetti https://orcid.org/0000-0002-7504-365X
R. L. Walsworth https://orcid.org/0000-0003-0311-4751

## References

Aigrain, S., Pont, F., & Zucker, S. 2012, MNRAS, 419, 3147
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Sci, 327, 977
Buchhave, L. A., Latham, D. W., Johansen, A., et al. 2012, Natur, 486, 375
Cegla, H. M. 2019, Geosc, 9, 114
Cessa, V., Beck, T., Benz, W., et al. 2017, Proc. SPIE, 10563, 468
Chaplin, W. J., Cegla, H. M., Watson, C. A., Davies, G. R., & Ball, W. H. 2019, AJ, 157, 163
Collier Cameron, A., Mortier, A., Phillips, D., et al. 2019, MNRAS, 487, 1082
Cosentino, R., Lovis, C., Pepe, F., et al. 2014, Proc. SPIE, 9147, 2658
Couvidat, S., Schou, J., Hoeksema, J. T., et al. 2016, SoPh, 291, 1887
Cybenko, G. 1989, MCSS, 2, 303
de Beurs, Z. L., Vanderburg, A., Shallue, C. J., et al. 2020, arXiv:2011.00003
Dumusque, X., Boisse, I., & Santos, N. C. 2014, ApJ, 796, 132
Dumusque, X., Cretignier, M., Sosnowska, D., et al. 2021, A&A, 648, A103
Dumusque, X., Glenday, A., Phillips, D. F., et al. 2015, ApJL, 814, L21
Egeland, R., Soon, W., Baliunas, S., et al. 2017, ApJ, 835, 25
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Haywood, R. D., Collier Cameron, A., Unruh, Y. C., et al. 2016, MNRAS, 457, 3637
Haywood, R. D., Milbourne, T. W., Saar, S. H., et al. 2020, arXiv:2005.13386
Herrero, E., Ribas, I., Jordi, C., et al. 2016, A&A, 586, A131
Hinton, G. E. 1989, ArtInt, 40, 185

Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398
Kopp, G., Heuerman, K., & Lawrence, G. 2005, SoPh, 230, 111
Kopp, G., & Lawrence, G. 2005, SoPh, 230, 91
Langellier, N., Milbourne, T. W., Phillips, D. F., et al. 2021, AJ, 161, 287
Leroy, J. L. 1962, AnAp, 25, 127
Linksy, J. L., & Avrett, E. H. 1970, PASP, 82, 169
Meunier, N., Desort, M., & Lagrange, A.-M. 2010a, A&A, 512, A39
Meunier, N., Lagrange, A. M., & Desort, M. 2010b, A&A, 519, A66
Meunier, N., Lagrange, A.-M., & Borgniet, S. 2017, A&A, 607, A6
Miklos, M., Milbourne, T. W., Haywood, R. D., et al. 2020, ApJ, 888, 117
Milbourne, T. W., Haywood, R. D., Phillips, D. F., et al. 2019, ApJ, 874, 107
Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. 1984, ApJ, 279, 763
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825, https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf
Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, A&A, 645, A96
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3
Phillips, D. F., Glenday, A. G., Dumusque, X., et al. 2016, Proc. SPIE, 9912, 99126Z
Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269
Reiners, A., Shulyak, D., Anglada-Escudé, G., et al. 2013, A&A, 552, A103
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, JATIS, 1, 1
Rottman, G. 2005, SoPh, 230, 7
Rozelot, J. P., Kosovichev, A., & Kilcik, A. 2015, ApJ, 812, 91
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229
Shapiro, A. I., Solanki, S. K., Krivova, N. A., Yeo, K. L., & Schmutz, W. K. 2016, A&A, 589, A46
Stift, M. J., & Leone, F. 2003, A&A, 398, 411
van Rossum, G. 1995, Python Tutorial Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), https://ir.cwi.nl/pub/5007/05007D.pdf
Vaughan, A., Preston, G., & Wilson, O. 1978, PASP, 90, 267
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wilson, O. C. 1968, ApJ, 153, 221
Yeo, K. L., Solanki, S. K., & Krivova, N. A. 2013, A&A, 550, A95