

A Machine Learning approach for correcting radial velocities using physical observables

M. Perger^{1,2}, G. Anglada-Escudé^{1,2}, D. Baroch^{1,2}, M. Lafarga³, I. Ribas^{1,2}, J. C. Morales^{1,2}, E. Herrero^{1,2}, P. J. Amado⁴, J. R. Barnes⁵, J. A. Caballero⁶, S. V. Jeffers⁷, A. Quirrenbach⁸, and A. Reiners⁹

¹Institut de Ciències de l'Espai (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, 08193 Bellaterra, Spain
e-mail: perger@ice.cat

²Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain

³Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom

⁴Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain

⁵School of Physical Sciences, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK

⁶Centro de Astrobiología (CAB), CSIC-INTA, ESAC, Camino bajo del castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain

⁷Max Planck Institut für Sonnensystemforschung (MPS), Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany

⁸Landessternwarte, Zentrum für Astronomie der Universität Heidelberg, Königstuhl 12, 69117 Heidelberg, Germany

⁹Institut für Astrophysik, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Received 29.09.2022 / Accepted 01.03.2023

ABSTRACT

Context. Precision radial velocity (RV) measurements continue to be a key tool to detect and characterise extrasolar planets. While instrumental precision keeps improving, stellar activity remains a barrier to obtain reliable measurements below 1-2 m s⁻¹ accuracy.

Aims. Using simulations and real data, we investigate the capabilities of a Deep Neural Network approach to produce activity free Doppler measurements of stars.

Methods. As case studies we use observations of two known stars (ϵ Eridani and AU Microscopii, both with clear signals of activity induced Doppler variability). Synthetic observations using the `starsim` code are generated for the observables (inputs) and the resulting Doppler signal (labels), and then they are used to train a Deep Neural Network algorithm to predict Doppler corrections. We identify a relatively simple architecture consisting of convolutional layers followed by fully connected layers that is adequate to the task. The indices investigated are mean line-profile parameters (width, bisector, contrast) and multi-band photometry.

Results. We demonstrate that the RV-independent approach can drastically reduce spurious Doppler variability from known physical effects such as spots, rotation and convective blueshift. We identify the combinations of activity indices with most predictive power. When applied to real observations, we observe a good match of the correction with the observed variability, but we also find that the noise reduction is not as good as in the simulations, probably due to the lack of detail in the simulated physics.

Conclusions. We demonstrate that a model-driven machine learning approach is sufficient to clean Doppler signals from activity induced variability for well known physical effects. There are dozens of known activity related observables whose inversion power remains unexplored indicating that the use of additional indicators, more complete models, and more observations with optimised sampling strategies can lead to significant improvements in our detrending capabilities on new and existing data sets.

Key words. Methods: data analysis – Methods: statistical – Techniques: radial velocities – Stars: activity – Stars: starspots – Stars: low-mass

1. Introduction

The majority of exoplanet discoveries in the last two decades have used two methods. One is the transit method monitoring tens of thousands of stars simultaneously, and most effectively done from space with instruments such as Kepler (Borucki et al. 2010) and TESS (Transiting Exoplanet Survey Satellite; Ricker et al. 2015). The other one is the Doppler technique, which is conducted using high-resolution spectrographs such as HIRES (High Resolution Échelle Spectrograph; Vogt et al. 1994), HARPS (High Accuracy Radial velocity Planet Searcher; Mayor et al. 2003), HARPS-N (Cosentino et al. 2012), CARMENES (Calar Alto high-Resolution search for M dwarfs with Exoearths with Near-infrared and optical Échelle Spectrograph; Quirrenbach et al. 2020), or ESPRESSO (Échelle SPectrograph for Rocky Exoplanets and Stable Spectroscopic Observations; Pepe et al. 2021). Whereas the field of exoplanet science

is now progressing from the mere discovery of planetary companions to the analysis of atmospheres of already known, transiting planets, those surveys are still collecting new data providing complementary insight into planet formation and evolution. Significant challenges remain in the detection of small terrestrial planets around Sun-like stars mainly due to spurious effects caused by stellar variability.

There are large datasets, both collected from photometric and spectroscopic observations, which may have the precision to detect objects of high interest, but remain unexploited. In particular, radial velocities (RVs) are extracted from the high-resolution spectra by measuring the wavelength shift of the different lines present in the spectra compared to a laboratory reference standard or to a template created from the stellar spectra themselves. Shifts can be caused by motion of the stellar center of mass caused by a planet, or they can be induced by the stellar activity. There are various effects influencing those measurements.

Dumusque et al. (2012) give a comprehensive overview on the topic. The effects introduced by stellar surface phenomena, such as dark spots or bright faculae, induce apparent RV wobbles with typical time scales of rotation periods of stars, active region lifetimes, and stellar activity cycles, making it extremely difficult to disentangle such effects from true Doppler signals from planets, which tend to have comparable orbital periods. Most effective *inversion* strategies involve the joint fit of the so-called activity indices with the RVs using a handful of shared parameters in Gaussian processes as discussed in Suárez Mascareño et al. (2015), which still has the drawback of reducing the information content of the time series, and potentially filtering out potential signals of interest.

Currently RV precisions of 1 m s^{-1} are regularly achieved. But this precision rarely translates into accurate measurements of true stellar motion due to the effects of stellar activity. This spurious variability is estimated to be about 2 to 4 m s^{-1} (Perger et al. 2017) for quiet old stars, to hundreds of m s^{-1} for young and active stars. To account for those effects, many empirical techniques have been proposed and applied with limited success. This includes a simple fit of one or two sinusoids accounting for the stellar activity (Suárez Mascareño et al. 2015, often called pre-whitening); in-depth analysis of additional activity indices, photometry and their correlations with observed RV signals; and the application of more sophisticated noise models in Bayesian frameworks, such as Gaussian processes (Perger et al. 2019, 2021). In recent times, as for many fields in astrophysics and science in general, machine learning techniques are being considered to solve similar problems. Shallue & Vanderburg (2018) and Dattilo et al. (2019) developed and applied a method for classifying potential planet signals using machine learning algorithms in the Kepler/K2 data. de Beurs et al. (2020), on the other hand, use the cross-correlation functions (CCFs) created by the binary-mask technique used to extract the RVs from HARPS-N spectra of the Sun (Collier Cameron et al. 2019) and from SOAP 2.0 (Dumusque et al. 2014) synthetic data, and then train a neural network (NN) algorithm to distinguish Doppler shifts due to companions from activity induced shifts based on the information in the cross-correlation profile.

In this article, we investigate whether a machine learning framework trained using synthetic (but realistic and complex) observables is able to produce usable RV corrections without using the information of the Doppler shifts themselves. We present the set up and assumptions of our study in Sect. 2, including the different activity indices studied, two stars used as study-cases, the models and software used to generate synthetic time-series (*starsim*), and the methodologies and machine learning framework that we have developed to address the problem. The results and discussions are given in Sect. 3, and we provide conclusions to our study in Sect. 4.

2. Set up

2.1. Tracers for stellar activity

In the Doppler technique, the stellar spectrum is measured with high spectral resolution ($\mathcal{R} > 50\,000$). The Doppler shift of the spectrum is then obtained by cross-correlating the observed spectrum with a mask (cross-correlation technique as in Cosentino et al. 2012; Lafarga et al. 2020) or a template built from the same observations (Anglada-Escudé & Butler 2012; Zechmeister et al. 2018). The high resolution of the spectra also allows for simultaneous measurement of spectral characteristics other than the RV. In addition to this, contemporaneous photo-

metric observations can be used to infer the coverage of spots on the star, putting constraints on the expected RV variability. When possible, this photometric monitoring is done in various filter bands (or colours) in order to determine not only the spot coverage but also properties of the spots such as their effective temperatures (Rosich et al. 2020).

The process of measuring precision Doppler shifts with the cross-correlation method consists of consolidating the information of hundreds (thousands) of spectral lines in a cross-correlation profile obtained by performing the correlation of the observed spectra with a predefined weighted mask (list of wavelengths where spectral lines should appear at zero Doppler shift). The resulting CCF profile is equivalent to the *mean* shape of the spectral lines for each observation. The computation of this CCF profile is typically produced by instrument pipelines such as the Data Reduction Systems (DRS) of the instruments HARPS, HARPS-N, and ESPRESSO, and the *raccoon* code (Radial velocities and Activity indicators from Cross-CORrelatiON with masks; Lafarga et al. 2020) developed in the context of the CARMENES project. In this study we use *raccoon* to obtain Doppler shifts and precision measurements of the shape parameters of the CCF.

The RV is measured by fitting a Gaussian function and measuring its mean. Other useful measurements can be obtained from the fitted Gaussian profile, such as the full-width-at-half maximum (or FWHM, in m s^{-1}) and the relative depth -or contrast- relative to the continuum (CON, in %). As an indicator of possible changes in the symmetry of the CCF profile, the Bisector Inverse Slope (BIS, in m s^{-1}) is usually measured. As abundantly discussed in the literature, all these *indices* (FWHM, CON, and BIS, among others) are known to be affected by the presence of spots and other physical effects that are also responsible for the observed spurious Doppler shifts. Therefore the information needed to produce a *clean* RV measurement could be, in principle, available in the same observation through the indices. This production of clean RVs from other spectroscopic indices (or *inversion* hereafter) has been the topic of many research efforts during the last decade achieving only partial success.

2.2. Observational data

In order to base our experiments with synthetic data on realistic observations, we use measurements of two well studied and representative stars: the very nearby K2-dwarf ε Eridani (Gliese 144; ε Eri hereafter), and the very young M1-dwarf AU Microscopii (Gliese 803; AU Mic hereafter). They are both quite active and show strong stellar contributions to the RVs as well as to all CCF indices. In Table 1, we provide the parameters of these test stars as found in the literature. Both datasets were chosen for having high signal to noise spectra (so observations would not be limited by photon noise), and their observing cadence was quite regular over a relatively short time interval (2–3 months).

ε Eri is a young, nearby star with a rotation period of about 11.2d (Fröhlich 2007). It has a debris disk inclined by 30° (Quillen & Thorndike 2002) with respect to the sky plane, and at least one planetary companion of at least $0.7 M_{\text{Jup}}$ in a 7-year period orbit (Campbell et al. 1988) causing a Doppler signal with a RV semi-amplitude of about 11 m s^{-1} . For this study, we use 204 publicly available spectra obtained with HARPS ($3\,750 \text{ \AA} < \lambda < 6\,900 \text{ \AA}$, $\mathcal{R} = 115\,000$) obtained on 66 quasi-consecutive nights between 5 Oct 2019 (BJD = 2458762 d) and

Table 1. Literature and `starsim` parameters for ε Eridani, and AU Microscopii, and their planetary companions.

parameter	unit	ε Eridani			AU Microscopii		
		literature value	Ref.	<code>starsim</code>	literature value	Ref.	<code>starsim</code>
Right ascension RA	h:m:s	03:32:55.84±0.12	(1)	...	20:45:09.532±0.017	(1)	...
Declination Dec	°:':"	-09:27:29.739±0.093	(1)	...	-31:20:27.238±0.011	(1)	...
age	Myr	400 - 800	(2)	...	22±3	(4)	...
distance	pc	3.216±0.002	(1)	...	39.714±0.002	(1)	...
μ_α	mas a ⁻¹	-974.76±0.16	(1)	...	281.319±0.022	(1)	...
μ_δ	mas a ⁻¹	20.88±0.12	(1)	...	-360.148±0.019	(1)	...
$v \sin i$	km s ⁻¹	2.4±0.5	(3)	...	9.3±1.2	(5)	...
magnitude G	mag	3.4658±0.0031	(1)	...	7.8434±0.0029	(1)	...
spectral type SpT	...	K2.0V	(6)	HARPS K5 mask	M1.0V	(17)	CARMENES M1.5 mask
stellar mass $M_{*,v}$	M_\odot	0.82±0.05	(7)	0.8 - 0.9	0.50±0.03	(4)	0.45 - 0.55
		0.847±0.042	(8)				
stellar radius R_*	R_\odot	0.74±0.01	(7)	0.65 - 0.75	0.75±0.03	(4)	0.7 - 0.8
		0.702±0.035	(8)				
effective temperature T_{eff}	K	5076±30	(10)	4900 - 5300	3700±100	(4)	3550 - 3850
surface gravity log g	dex	4.30±0.08	(11)	4.5 (fixed)	4.39±0.03	(1)	4.5 (fixed)
metallicity $[Fe/H]$...	-0.13±0.04	(12)	0.0 (fixed)	-0.12±0.11	(18)	0.0 (fixed)
inclination i	°	60	(13)	30 - 90	90	(19)	60 - 90
rotation period P_{rot}	day	11.2	(3)	11 - 12.5	4.863±0.010	(20)	4.8 - 4.9
convective blueshift CB^*	CB_\odot	~0.3	(15,16)	0.0 - 0.5	~0.0	(15,16)	0.0 - 0.5
differential rotation $d\Omega^{**}$	$d\Omega_\odot$	1.3	(3)	0.1 - 4	3.0	(21)	0.5 - 5
spot temperature difference ΔT	K	1080±670	(14)	250 - 1000	660±570	(14)	250 - 1000
minimum mass $M \sin i$	M_\oplus	210	(22)	...	20; 11	(23)	...
orbital period	day	2671	(22)	...	8.46; 18.86	(23)	...
RV semi-amplitude	m s ⁻¹	~11	~6; ~8

Notes. ⁽¹⁾ Gaia Collaboration (2020), ⁽²⁾ Janson et al. (2015), ⁽³⁾ Fröhlich (2007), ⁽⁴⁾ Plavchan et al. (2020), ⁽⁵⁾ Torres et al. (2006), ⁽⁶⁾ Keenan & McNeil (1989), ⁽⁷⁾ Baines & Armstrong (2012), ⁽⁸⁾ Kervella et al. (2019), ⁽⁹⁾ Gomes da Silva et al. (2021), ⁽¹⁰⁾ Heiter et al. (2015), ⁽¹¹⁾ Gonzalez et al. (2010), ⁽¹²⁾ Santos et al. (2004), ⁽¹³⁾ Quillen & Thorndike (2002) ⁽¹⁴⁾ following Herbst et al. (2021), ⁽¹⁵⁾ Liebing et al. (2021), ⁽¹⁶⁾ Meunier et al. (2016), ⁽¹⁷⁾ Keenan & McNeil (1989), ⁽¹⁸⁾ Gaidos et al. (2014), ⁽¹⁹⁾ Kalas et al. (2005) ⁽²⁰⁾ Krist et al. (2005), ⁽²¹⁾ Klein et al. (2021), ⁽²²⁾ Llop-Sayson et al. (2021), ⁽²³⁾ Zicher et al. (2022),

^(*) in units of the solar convective blueshift $CB_\odot \sim -325 \text{ m s}^{-1}$ (e.g., Löhner-Böttcher et al. 2018; Stief et al. 2019),

^(**) in units of the solar differential rotation $d\Omega_\odot = 3.15^\circ \text{ day}^{-1}$, assuming $P_{\odot,\text{min}} \sim 24.47 \text{ d}$, $P_{\odot,\text{max}} \sim 31.14 \text{ d}$, and $d\Omega_\odot = 360^\circ (P_{\odot,\text{min}}^{-1} - P_{\odot,\text{max}}^{-1})$ (Beck 2000).

1 Jan 2020 (BJD = 2458850 d) with a total baseline of 88 day¹. All observations within a night were consolidated in night average values to produce the 66 epochs that we will use in all that follows.

AU Mic, on the other hand, is a young star located at a distance of 9.71 pc (Plavchan et al. 2020). It rotates with a period of 4.9 d (Krist et al. 2005) and shows strong evidence of differential rotation ($d\Omega = 9.57 \pm 0.52^\circ \text{ day}^{-1}$, Klein et al. 2021). Its debris disk is seen edge-on (Kalas et al. 2005). Since planets are supposed to form in such disks, it could host transiting planets as well. Indeed, Plavchan et al. (2020) and follow-up publications reported the detection of (at least) two transiting planetary companions with orbital periods of 8.5 d, and 18.9 d, and minimum masses of about 20 and 11 M_\oplus , respectively, inducing possible RV semi-amplitudes of 6, and 8 m s⁻¹. In this study we use 75 CARMENES VIS spectra ($5200 \text{ \AA} < \lambda < 9600 \text{ \AA}$, $\mathcal{R} = 94600$) obtained for this star (available from CARMENES DR1, Ribas et al. 2023), which were observed quite regularly on 37 nights from 14 Jul 2019 (BJD = 2458679 d) to 9 Oct 2019 (BJD = 2458766 d) with a total baseline of 87 days. After consolidating the data in night averages, we reach an average time spacing of 2.5 days between consecutive observations.

Our early experiments were run using DRS instrument pipeline data-products for the real data, and `raccoon` generated measurements on synthetic data. The mismatches were very obvious in all cases and no correction close to what was observed could be obtained from them. This indicated that a high level

of consistency was needed in all the algorithms producing observables in the simulations and the real data-products. In order to be fully consistent, we use the `raccoon` code to retrieve RV, FWHM, BIS, and CON from the CCF using the binary masks as indicated in Table 1, and then adapted the codes in `starsim` to be fully consistent with them.

For the analyses that follow, it is important to separate what we consider random errors (photon noise, instrumental and other effects generating uncorrelated errors) from activity induced variability (which has time-scales of days, due to the rotation period and its harmonics), especially for the real datasets. To obtain an estimate of the random errors for all the real time-series, (RV, but also the activity indices), we perform a pre-whitening procedure using sequential fitting of sinusoids using Generalised Lomb-Scargle periodograms (GLS, Zechmeister & Kürster 2009). When there is no signal left below a 1% false alarm probability (FAP), we assume that the remaining variability are these random errors, although they might contain some correlated stellar jitter below the noise level. The standard deviation of these random errors in all the series will later be used to simulate synthetic datasets with *realistic* observational noise. Table 2 shows the residual RMS from this procedure, and the ratio between the original RMS and the residual RMS after the pre-whitening process, expressed as a percentage. This ratio between the residual RMS compared to the original RMS expressed as a percentile is what we call *Relative Residual Error* (or RRE), and it will be used in many instances throughout the paper for different time-series. We rather use this value because we compare our results to the relative RV change and not to its absolute true value (which would be the accuracy). A depiction of the time-series (RV and spectroscopic indices) as obtained from the ob-

¹ Programme IDs: 0104.C-0863, 072.C-0488, 072.C-0513, 073.C-0784, 074.C-0012, 076.C-0878, 077.C-0530, 078.C-0833, 079.C-0681, 192.C-0852, 60.A-9036 secured as part of the Red Dots ESO programme.

servations is shown in Figs. 1 (ε Eri) and 2 (AU Mic). The data in the figures can be found in Tables B.1 and B.2.

The signal of the stellar rotation period is present in all time series of ε Eri. Probably due to the differential rotation of the star, the rotation period as given in the literature manifests itself more strongly in its second harmonic at 5.6 days. By using the pre-whitening procedure described earlier, we achieve RREs between 17 to 25% in all indices.

For AU Mic we find the rotation period in all the data sets except the CON series. As also discussed later, the measurement of the CON series might not be useful for this kind of target and study. For the other indices the achieved RRE is about 30% to 60%.

Next we perform a quick analysis applying a Gaussian Process (GP) using the `george` package (Ambikasaran et al. 2015) and the quasi-periodic and cosine (QPC) kernel (Perger et al. 2021), and exploring the parameter space with a Monte-Carlo Markov Chain (MCMC) using `emcee` (Foreman-Mackey et al. 2013). We show in Fig. 3 the posterior distributions for the period hyperparameters $P_{\varepsilon\text{Eri}} = 11.50^{+0.30}_{-0.22}$ d and $P_{\text{AU Mic}} = 4.87 \pm 0.01$ d, and $\lambda = 44^{+19}_{-10}$ d and 130^{+41}_{-46} d, where λ is the characteristic decay time of the correlations (related to the average time it takes for a spot to change substantially in size or location, or even disappear) for the RVs of ε Eri and AU Mic, respectively.

2.3. Synthetic data

We use the `starsim` code (Herrero et al. 2016) to model time-series data for a spotted rotating star with the characteristics of our two test stars ε Eri and AU Mic. The current version of `starsim`² (Baroch et al. in prep.) is able to produce synthetic photometry for a predefined filter from low-resolution model spectra for our study, we produce light curves in Johnsons V , R , and I filters. CCFs at each observation epoch are also generated from high-resolution Phoenix spectra (Husser et al. 2013) of a predefined wavelength range, from which RVs, and the CCF activity indices are then measured. The calculations make use of important stellar parameters, geometric effects, and of the distribution of *magnetically active regions* (spots hereafter) on the surface of the star. One simulation consists of a list of epochs and their corresponding RV, FWHM, BIS, CON, and V , B and I photometric bands, all assumed to be simultaneous and without errors.

Preliminary experiments (not described here for brevity) showed that constraining the simulations to the known physical parameters of the star is rather important to improve the inversion capabilities of the algorithms used. This makes sense as NNs are known to be good tools for interpolation as long as the training set densely populates the region of the parameter space of interest. The priors of the most important stellar variables as adopted for the creation of the synthetic data sets of the two test stars are shown in Table 1. For mass, radius, effective temperature, inclination of the rotation axes, and the rotation period we use the published values in their error limits. In the case of ε Eri, we adapt the range for the latter parameter in order to include the larger periodicities found by our pre-whitening as shown in Table 2. We choose further fixed values for metallicity and surface gravity. In the algorithm, the convective blueshift is given in units of the solar shift of -325 m s^{-1} (e.g., Löhner-Böttcher et al. 2018; Stief et al. 2019), and the differential rotation in units of the solar value $d\Omega_{\odot} = 3.15^{\circ} \text{ day}^{-1}$ (following Beck 2000). Furthermore, the binary masks of the corresponding stellar type and

wavelength range were used to calculate the CCFs. In each simulation, the physical properties of the star are randomly initialised within these priors, thus implicitly accounting for the uncertainty in the basic properties of the star at simulation level (training and test sets).

`starsim` also requires to specify the distribution of the spots on the stellar surface (assumed circular) in what we call a *spot map*. The contribution of a spot in a simulation is taken into account when producing a CCF integrated over the stellar disk (spots have a different CCF than the photosphere because of lower temperature and different spectral features). For the spots we also use Phoenix spectra of lower temperature as given by the ΔT parameter in Table 1, and a modified bisector following what is observed on the Sun (Herrero et al. 2016). The priors of these parameter are specified by following empirical studies on a variety of stellar types in Herbst et al. (2021). In the input *spot map* each spot is described by five parameters: time of appearance, lifetime, longitude, co-latitude and radius. Taking into account that the contribution from stellar activity is clearly visible in all sets at all times (no flat regions in the observational data), and to match both absolute value and variation of each time series, we use randomly 25 to 40 dark spots to ensure that at least one to two spots are visible at any observing time. All spot map parameters are also generated randomly following a uniform distribution: radii (2.5 to 4°), lifetime τ (5 to 100 days), time of appearance ($t_0 - \tau$ to $t_0 + 100$ d), co-latitude (0 to 180°), and longitude (0 to 360°). The values adopted try to encompass the intervals of the most likely values from these parameters following the GP analysis from the previous section.

To investigate the importance of sampling in the ability to correct for stellar activity, we generate simulations following five different time samplings with `starsim`

- **uniform** consisting of 100 data points uniformly distributed over 100 days;
- **random1** consisting of 100 data points randomly distributed over the same time baseline
- **gap1** consisting of 60 randomly distributed data points from time 0 d to 50 d; and 40 data points from time 70 d to 100 d, thereby creating an artificial visibility gap as we often encounter it in actual astronomical campaigns
- **random2** which uses 100 randomly chosen out of the 300 data points in *uniform*, *random1* and *gap1*
- **gap2** which has 40 data points in the first 30 days, and 60 in the last 45 days of the 300 data points, leaving open another gap.

where *random2* and *gap2* are generated from subsets of the previous samplings to save simulation time (most computationally intensive step as discussed later). The 100 day time interval was chosen to cover the rotation period of both test stars a sufficient number of times. We then generate 600 000 and 1 000 000 simulations for ε Eri and AU Mic, respectively, for each of these time-samplings.

We checked for possible linear relations between the synthetic RVs and the activity indices. Applying those relations, we reach RREs for V , I , BIS, FWHM, and CON indices of 99%, 99%, 79%, 98%, and 98% for ε Eri, and 99%, 99%, 73%, 86%, and 95% for AU Mic, respectively. If we apply a linear fit to all six datasets, we achieve RREs of 77% for ε Eri and 72% for AU Mic, respectively, showing the importance of a more sophisticated and RV-independent approach for our synthetic data.

² <https://github.com/dbarochlopez/starsim>

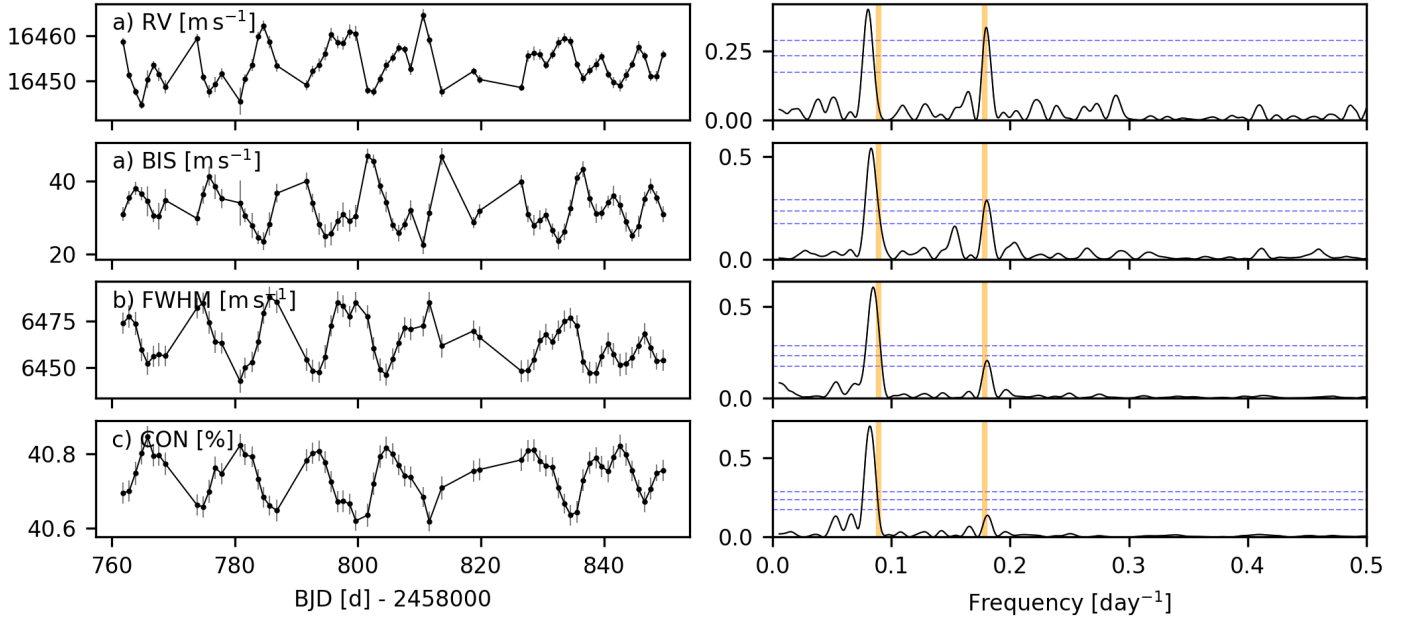


Fig. 1. Nightly binned time series (left panels) and their periodograms (right panels) of ϵ Eri extracted with *raccoon*. From top to bottom those are RV, BIS, FWHM, and CON. Dashed blue horizontal lines indicate the FAP levels of 0.1, 1, and 10%, and the vertical orange lines indicate the frequency of the rotation period of 11.2 d and its second harmonic.

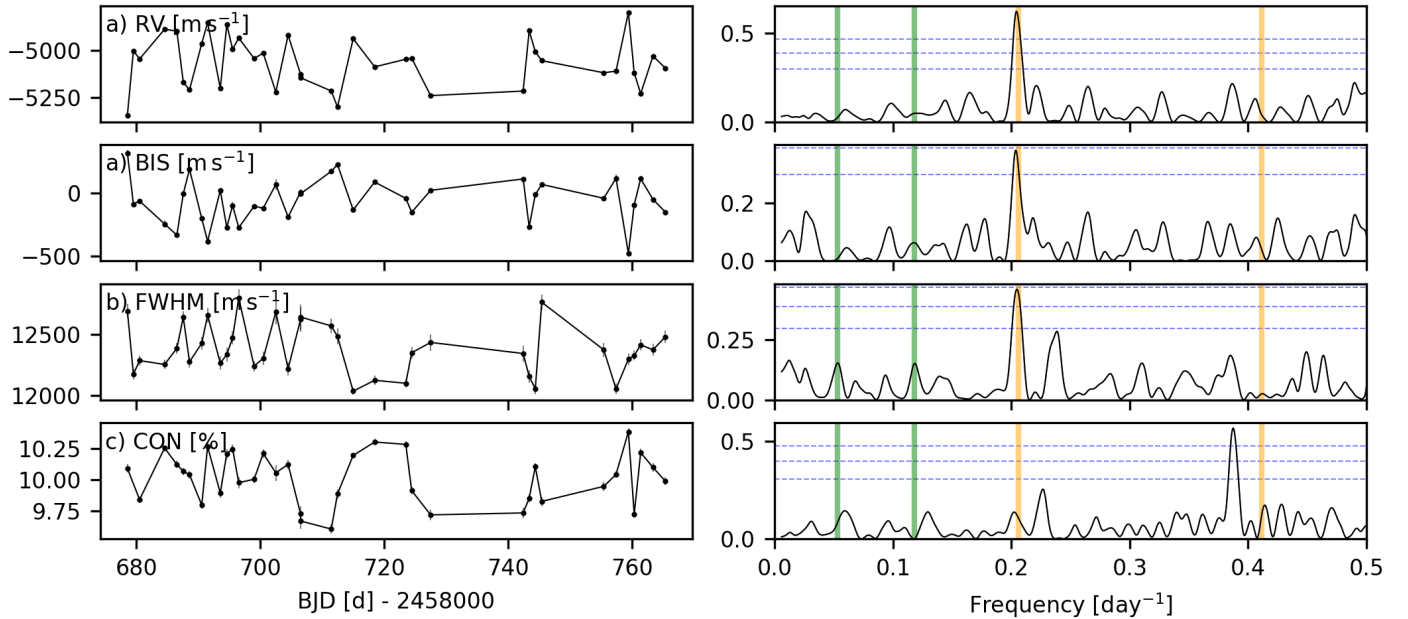


Fig. 2. Nightly binned time series (left panels) and their periodograms (right panels) of AU Mic extracted with *raccoon*. From top to bottom those are RV, BIS, FWHM, and CON. For the latter, one obvious outlier is not in this diagram. Dashed blue horizontal lines indicate the FAP levels of 0.1, 1, and 10%, and the vertical orange lines indicate the frequency of the rotation period of 4.9 d and its second harmonic. The vertical green lines mark the periods of the two planetary companions.

2.4. Neural Network Architectures

Finding the right format for the inputs, and defining an adequate NN architecture, are some of the major challenges when applying NN approaches to new types of data. Neural networks with several fully connected layers have a huge number of parameters, thus enabling a fit to almost anything. However, being so general, these are slow to train (and may not even converge at all) and often lead to over-fitting (i.e. the NN is able to reproduce the training set, but performs poorly on a test sample). For

an NN to be able to make non-linear predictions, more than two layers are needed. Since the data we are analysing contain strong correlations between neighboring points, and these correlations encode the physics, we propose an NN composed of a small number of convolutional layers first, followed by three dense fully connected layers. This architecture is similar to those used in pattern recognition in images, which also looks for patterns in the spatial domain to identify features in an image. We use

Table 2. Time-series data characteristics of 66 nightly binned data points of ε Eri using HARPS (top), and of 37 nightly binned data points from AU Mic using CARMENES VIS (bottom), both extracted by raccoon .

data	unit	mean [unit]	RMS [unit]	uncertainty [unit]	RRE [%]	significant periods [d]
RV	m s^{-1}	16453.68	4.36	1.19	39	12.44; 5.56
BIS	m s^{-1}	32.48	5.56	2.44	17	12.06; 5.56; 10.60
FWHM	m s^{-1}	6463.83	12.01	5.72	23	11.81; 5.54; 175.43; 12.81; 10.32
CON	%	40.739	0.058	0.029	25	12.18; 5.54; 175.43; 10.63; 12.62
RV	m s^{-1}	-5064.84	132.26	6.76	25	4.90; 2.59
BIS	m s^{-1}	-61.75	172.04	17.11	28	4.90; 37.72; 2.59; 5.74; 2.45
FWHM	m s^{-1}	12380.18	202.05	54.23	58	4.89; 4.19
CON*	%	10.013	0.204	0.035	71	2.58

Notes. Relative residual error (RRE) is the ratio between the original RMS versus the resulting one obtained after applying pre-whitening (aka. until no signal with FAP < 1% is present in the data)

*: statistics without outlier

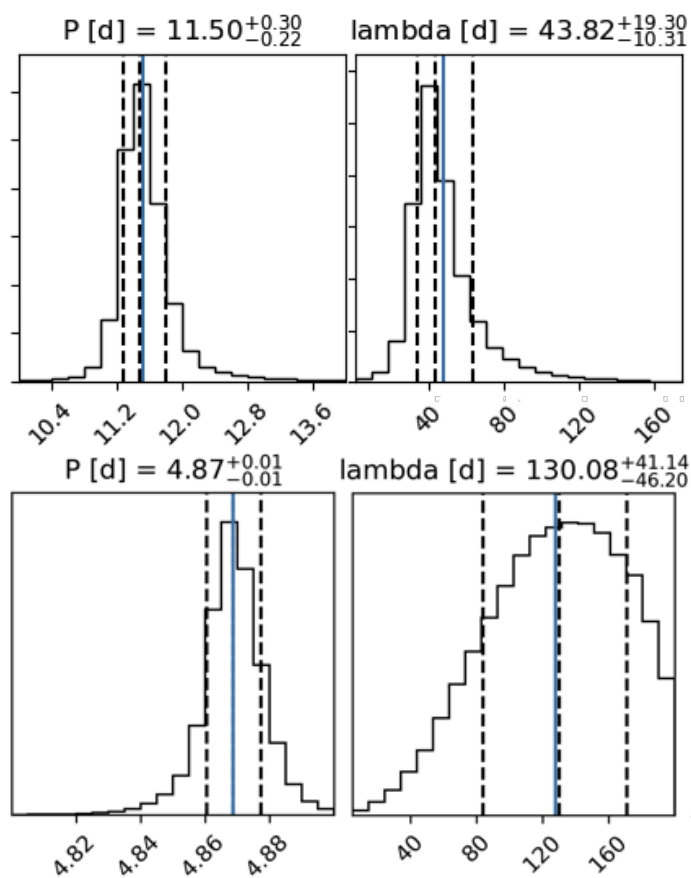


Fig. 3. Histogram of the MCMC posterior distributions for hyperparameters P (left panels) and λ (right panels) of Gaussian processes applied to ε Eri (top panels) and AU Mic (bottom panels) RV data using the QPC kernel.

the open-source python package PyTorch³. We also run experiments with Keras/Tensorflow⁴, which is the other most popular platform for NN development. Neither results nor performance depended much on the choice of the framework, so all results presented here were produced using PyTorch. In all cases, we installed the GPU enabled versions of the libraries to take full advantage of our hardware capabilities. The codes were devel-

oped in python. Both the starsim simulations and the training of the NN were executed on two identical dedicated machines (Intel Core i7-10700, which contains 8 CPU cores, and equipped with a GPU-NVIDIA GeForce RTX 2060 SUPER each).

An input sample consists of N_i time-series of indices (or channels) corresponding to the activity indices used for each test. The target (also called label) of each input sample is the corresponding RV time-series of the corresponding simulation. For example, for a synthetic set with 100 epochs, a NN network trained to produce RVs from FWHM and BIS data, would have two channels (FWHM and BIS) of 100 datapoints each as input, and one output list with 100 RV datapoints. All time-series are mean subtracted and normalised by the full ensemble standard deviation for that index (or label). Thereby we conserve the information of the relative strengths of the variations of the indices.

The input layer is then an array of $1 \times 100 \times N_i$ neurons. The first two layers are convolutional layers with small kernel sizes (typically 3 elements only), and a rather large number of channels (typically 32). This means that in the first convolutional layer, the information of each point and neighbors will be taken into account at the layer output. The second convolutional layer repeats the same operation, meaning that, effectively, the NN will take into account the value of each point and the ± 2 closest neighbors. Tests using kernels with larger sizes or the inclusion of more convolutional layers did not improve the performance of the NN significantly. The outputs of all neurons of the second convolutional layer are then arranged in a single list (so-called flatten operation), and this is then fed to three fully connected layers with 300 neurons each (also called dense layers). The output layer is a fully connected layer with 100 neurons. Except for the output layer (which has a linear function as its response), the response functions used in all other layers were the so-called ReLU (Rectified Linear Unit; $f(x) = \max(0, x)$) plus a constant offset (also called bias). We tested alternative functions (such as sigmoids), finding similar performances but at the cost of more training iterations and longer training time per iteration. As a cost function, we use the standard deviation of the difference between the simulated RVs (label) and the RV predictions by the NN (output). Training a NN consists of finding a combination of neuron parameters (weights and biases) that minimises this cost function. The optimization procedures and algorithms are already included in the aforementioned libraries (PyTorch and/or Tensorflow).

³ PyTorch is a trademark of Facebook, Inc.

⁴ Maintained by Google

For training, we use 90% of the synthetic sets, and 10% of the sets are reserved for testing and validation. On average, we found that these architectures would reach reasonable convergence (as measured by non-significant improvement of the cost function after 5 iterations) with about 50 training epochs. As an important check, we verified that the cost function applied to the test sets would always be equal (or slightly worse) than on the training sample, thus ensuring that we were not incurring in over-fitting. We show a visual representation of the NN architecture as described here in Fig. 4, as well as some computing-time benchmark experiments in Table 3.

3. Results

In this section we perform a series of experiments to benchmark the inversion performance of different activity indicators. Each experiment consists of taking a time sampling (uniform, random, gap) and using a combination of activity indicators as the input for the NN. Then, the RRE of the RVs is computed for the full ensemble of outputs and used as a figure of merit of the procedure. We first do it with simulated datasets without uncertainties in the input nor the output RVs (section 3.1), and then we replicate all the experiments for the uniform time sampling adding realistic noise (using the error value from Table 2) to the inputs to quantify the impact of uncertainties in the training set and the overall performances (section 3.2).

From the first experiments, it became clear that the most computationally intensive part of this study was the simulation of the synthetic data sets and not the training of the NN. With our setup, simulating one data set consisting of RV, FWHM, BIS, CON, V , B and I takes about one second per CPU. Assuming 8 cores, it takes about 1.5 days to complete 10^6 simulations. These simulations can be easily parallelised, the total computing time being proportional to the inverse of the number of CPUs employed. For illustrative purposes, Table 3 shows the computer time needed to perform one training iteration, showing that these are relatively small (the 50 epoch values that we use to train each NN can be reached within a few minutes). The table illustrates the advantage of using GPU enabled computers and choosing a suitable size of a parameter called batch-size, which is the number of samples used in a single optimization step within one training epoch. This is a key parameter to take full advantage of the parallel computing power of GPUs as abundantly discussed in NN literature. A large batch size uses many samples to adjust all the parameters of an NN (sends a large number of samples to the GPU to perform the back propagation adjustment of the NN), but may smooth out subtle effects contained in the simulations thus making it more difficult to converge to an optimal solution. We found that an optimal batch size of 512 samples per training step was sufficient to reproduce the convergence performance of smaller batch sizes while keeping a the computing time quite low.

3.1. Neural Network training without observational errors

Tables 4 and 5 show the RRE for each exercise, i.e. the relative residual RV error. The combinations of activity indices used in each exercise are shown in the left column. We show further the values averaged over all time samplings (*average*), and the results obtained for the *uniform* time sampling when we inject realistic noise (see Table 2) to the input indices (see Sect. 3.2). Figure 5 visually shows the same results. Additionally, the RRE of the test sets as a function of the number of epochs is shown in Fig. 6 and Fig. 7 for both stars. There, we used the NNs trained

Table 3. Numerical performance tests as a function of batch size.

Batch size	Computer time per epoch [sec]	Number of batches	GPU usage [%]	CPU usage [%]
CPU only				
32	120	28125	0	90
64	82	14063	0	90
128	81	7032	0	90
256	66	3516	0	92
512	60	1758	0	93
CPU + GPU				
32	92	28125	2	14
64	55	14063	5	17
128	28	7032	5	17
256	13	3516	6	10
512	8	1758	5	10
1024	6	879	5	10
2048	5	440	5	10

Notes. For this experiment in particular, we generated one million synthetic data sets following the time-sampling of ϵ Eri (66 observations). For each experiment we trained a convolutional neural network (three convolutional layers, followed by two fully connected ones). The input samples consisted of three spectroscopic indices (FWHM, BIS, and CON).

with the uniform time sampling and without additional errors, and applied it to the test sets without (left panels) and with (right panels) the additional errors for the test input sets, respectively.

As a further illustration, we show one random synthetic RV test set from the *uniform* time sampling and the output of the NNs with all the combinations of input data, i.e. index time series, in the left panels of Fig. 8. From the top, we see in the first panel the usage of only one index, then two indices, then the three indices from photometry and from the CCF, and then all six index time-series data. The model itself is shown by the black dots and line. In the right panels, the difference between the RV model and its corresponding prediction by the NN is shown.

In general, we find that even using a single index, there is always a reduction in RRE. But even for data without uncertainties, not all the activity induced RV variability can be removed from the observations. Let us note that this *inversion error* marks the floor of what is possible given the information in the corresponding index (these are intrinsic model degeneracies) and our NN implementation. Additional indices at the same epochs (additional spectroscopic observables) and/or additional non-simultaneous measurements (e.g. contemporaneous photometric observations) could improve these floor values, but one should expect worse performances when applied to real/noisy observations.

The most important take home messages from these experiments can be summarised as follows.

1. **The proposed neural network architecture is sufficient** to produce high quality corrections to RV data. The NNs converge after only a few training epochs irrespective of the fine details of the NN algorithm (response function, training strategy or other technical details). The convergence and non-overfitting can be easily verified by checking that the RRE of the test sample (which is the merit function also used during training) is similar or slightly larger than the RRE of the

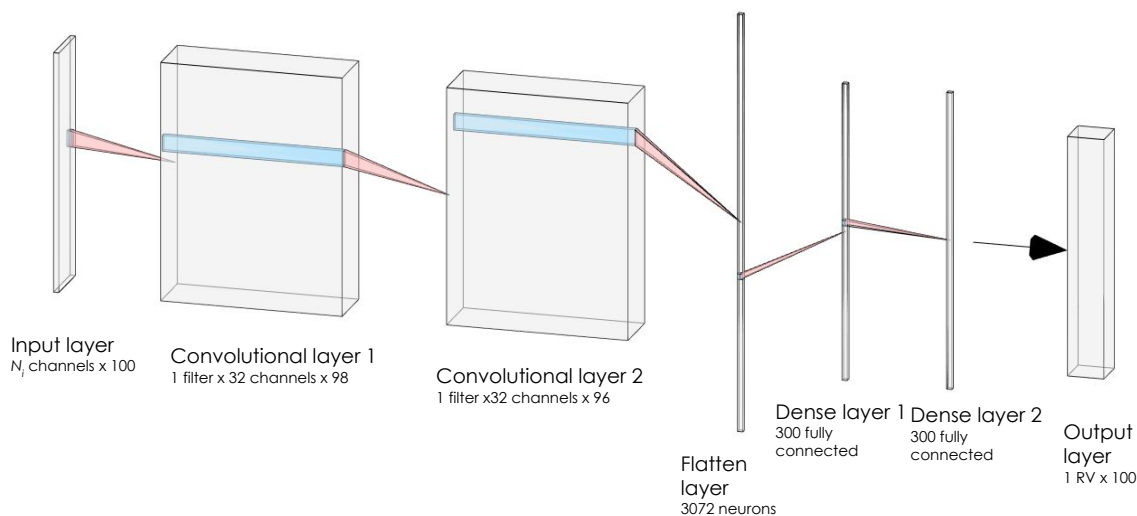


Fig. 4. Neural network architecture used throughout this study. The input layer consists of 100 measurements of N_i activity indicators, and the output layer (or label) are the corresponding 100 RVs. The first two hidden layers are convolutional layers, each with 32 kernels applied onto the three neighboring neurons. After the second convolutional layer, all the outputs for all the neurons are organised in a linear vector of neurons, and three subsequent fully connected layers are used to transform this vector into the final RV prediction. The response functions of all the neurons are the so-called Rectified Linear Unit functions (or ReLu), which clearly outperform other non-linear functions in terms of convergence speed and smaller computational time. As discussed in the text, small modifications of this architecture do not significantly change the obtained results.

Table 4. Relative residual error (RRE) of the RVs in % after 50 training runs for the neural networks using 540 000 starsim models for different time samplings and for different kinds and numbers of activity indices as input for ϵ Eri .

Input index	<i>uniform</i>	<i>random1</i>	<i>gap1</i>	<i>random2</i>	<i>gap2</i>	<i>average</i>	<i>with error</i>
<i>V</i>	19.9	20.7	18.7	21.3	20.3	20.2±0.4	23.8
<i>R</i>	22.4	22.9	21.4	23.7	23.0	22.7±0.3	26.0
<i>I</i>	23.9	24.7	24.2	25.5	24.7	24.6±0.2	27.5
<i>V-I</i>	8.6	9.5	8.6	9.9	8.8	9.1±0.2	22.4
<i>V-R-I</i>	8.7	9.6	8.5	10.0	8.8	9.1±0.2	22.0
BIS	17.0	19.3	18.8	19.7	18.2	18.6±0.4	19.1
FWHM	23.1	29.1	27.0	28.5	24.9	26.5±1.0	27.1
CON	12.6	12.4	10.4	13.1	12.0	12.1±0.4	20.6
BIS-FWHM	11.9	13.9	13.4	14.0	13.0	13.2±0.3	14.5
BIS-CON	6.5	7.8	6.1	7.8	7.2	7.1±0.3	9.1
FWHM-CON	7.5	7.9	6.6	8.7	7.8	7.7±0.3	10.8
BIS-FWHM-CON	5.1	5.8	5.5	6.5	5.9	5.8±0.2	7.5
all 6	4.7	5.7	4.9	5.6	5.3	5.2±0.2	6.1
time-BIS	17.0	19.2	18.9	19.7	18.2	18.6±0.4	19.1

Notes. Column *average* shows the average of all the values of the other time samplings and their standard deviation, *with error* shows the RV RRE of the *uniform* time sampling, where the additional relative residual errors (see Table 2) are added to each index of the NN input data.

- training sample, which is always the case in all the experiments.
- Line-profile indices contain enough information** to successfully train the algorithms, and produce a significant correction of the activity signals in the RV time-series. Despite the existence of correlations between RV and indices has been demonstrated in the past, this is the first time (to our knowledge) where it is shown that there is enough information in them to reproduce the RV activity signal.
- The best performing single index was found to be CON, rather than BIS**, which would have been expected according to Lafarga et al. (2020). This indicates that data/model driven approaches rather than guesswork or theoretical expectations are preferable in future activity/RV noise investigations. It is possible that other indices provide even better performances than those discussed here.
- Combining multiple indices is the most effective way** to achieve a higher RV reduction, given that these measurements can be obtained from the same spectra used to obtain

Table 5. Relative residual error (RRE) of the RVs in % after 50 training runs for the neural networks using 900 000 *starsim* models for different time samplings and for different kinds and numbers of activity indices as input for AU Mic .

	<i>uniform</i>	<i>random1</i>	<i>gap1</i>	<i>random2</i>	<i>gap2</i>	<i>average</i>	with error
<i>V</i>	55.8	53.5	52.2	54.1	51.9	53.5±0.6	57.4
<i>R</i>	56.2	54.4	53.0	54.9	52.7	54.2±0.6	57.8
<i>I</i>	57.7	57.2	55.9	57.5	55.6	56.8±0.4	59.0
<i>V-I</i>	48.9	47.7	46.6	48.1	46.6	47.6±0.4	55.0
<i>V-R-I</i>	48.9	47.5	46.3	48.1	46.5	47.4±0.4	54.6
BIS	48.2	47.9	47.8	48.0	47.6	47.9±0.1	51.7
FWHM	19.2	16.0	14.9	16.6	15.0	16.3±0.7	44.0
CON	21.7	14.3	11.8	16.4	14.0	15.6±1.5	44.7
BIS-FWHM	4.7	4.6	4.4	4.7	4.2	4.5±0.1	10.4
BIS-CON	7.4	7.3	6.2	7.6	6.3	7.0±0.3	17.8
FWHM-CON	7.9	6.5	5.6	7.1	5.7	6.6±0.4	26.8
BIS-FWHM-CON	2.4	2.5	2.3	2.5	2.1	2.4±0.1	8.0
all 6	2.0	2.0	1.9	2.1	1.7	1.9±0.1	5.1

Notes. See notes on Table 4 for further details.

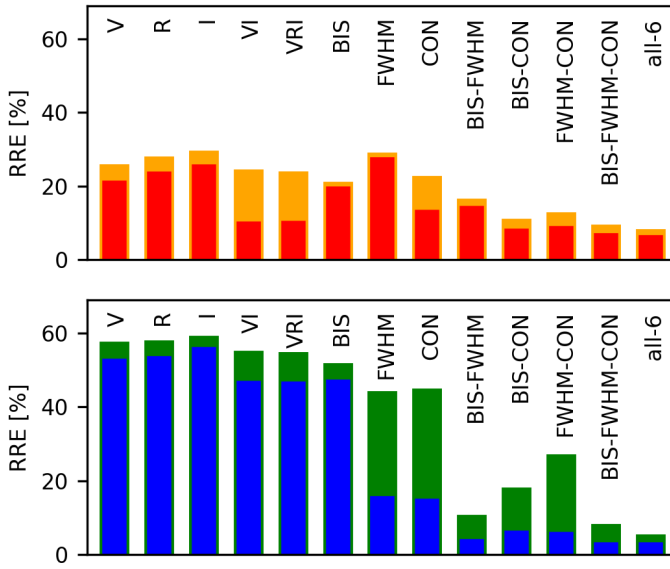


Fig. 5. Relative residual error (RRE) after the application of the trained NNs on the 540 000 and 900 000 *starsim* models in percentage for ε Eri (top) and AU Mic (bottom), respectively. We mark the different activity-index combinations used for the neural network training. The average RV reduction for the different time samplings *average* is given for ε Eri in red and for AU Mic in blue. The yellow and green bars indicate the value, if the relative additional error from Table 2 was added to the input data.

the RV measurement. The actual reduction, and the most relevant index is star dependent (as has been suspected for some time now) and, in general, one should aim at using a combination of all possible indices that show variability consistent with stellar activity. In particular, using all the available CCF indices, the simulations indicate that we can reach a RRE of 5.1, and 2.4 % for ε Eri and AU Mic, respectively. This is better than half the value as for any single index for ε Eri, but is remarkably lower (eight times) for AU Mic .

- 5. Simultaneous photometric time-series seem to contain the least information of the tested indices,** worsening for the redder filters. Even when combined, they only match the performance of a single CCF index as input. Especially for

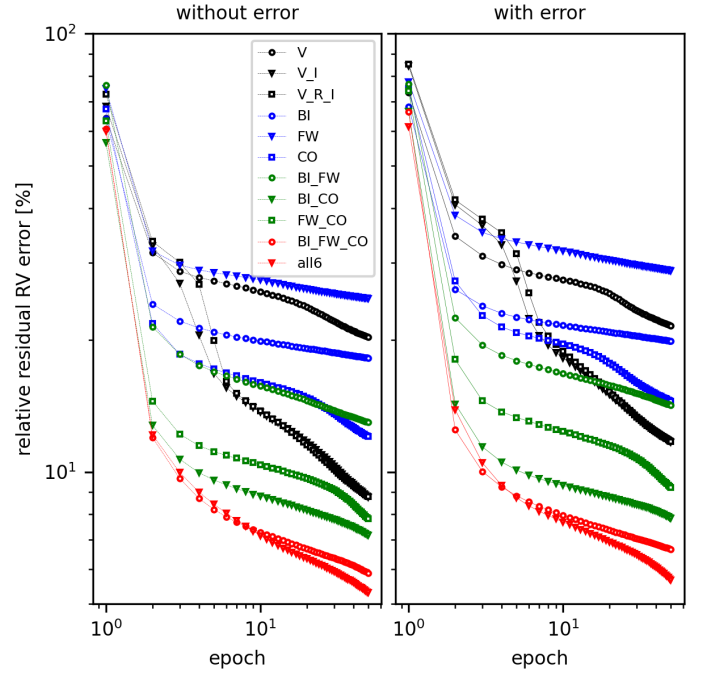


Fig. 6. Evolution of the relative residual error (RRE) of the 60 000 synthetic ε Eri test set RVs for the 50 epochs for the neural networks using all different combinations of input data, i.e. activity index time series as indicated, the uniform time sampling, and no additional error on the training input sets. The curves are shown for the test sets without additional uncertainties as calculated by *starsim* (left panel, dashed lines), and for the case, where we included the error value from Table 2 for each activity index into the test sets (right panel, full drawn lines).

AU Mic, the obtained RREs are only at the 55 % level, which is a rather poor performance. Using all the available filters, we can reach RREs of 9 and 49 % for ε Eri and AU Mic, respectively. This is likely caused by intrinsic degeneracies when mapping photometric signals into putative spots (and corresponding RV signal), as often happens in light curve-only inversion cases. For example, a small equatorial spot can produce the same photometric signal than a small polar one, but its RV counterpart can be rather different. In any case, adding photometry always helps, and the putative use

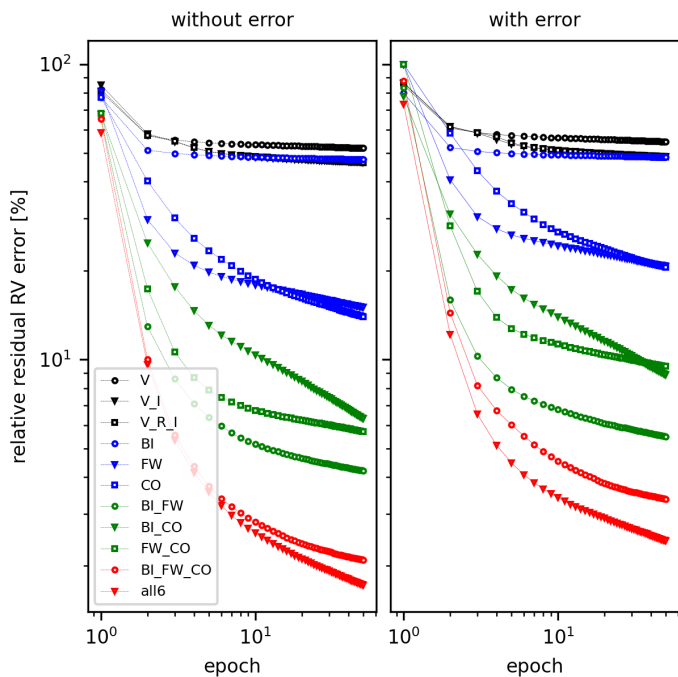


Fig. 7. Evolution of the relative residual error (RRE) of the 100 000 synthetic AU Mic test set RVs for the 50 epochs for the neural networks using all different combinations of input data. See Fig. 6 for more details.

of photometric time samplings denser than the RV sampling (as often happens when using space-based photometry) remains to be investigated.

6. **A larger number of data-points seem to improve overall performance.** Probably due to the smaller number of data points (37 versus 66) and the later spectral type, the RRE of the RVs across all input data combinations is more than double in the AU Mic data compared to the ε Eri sets, and this is despite the larger training data sample (540 000 vs. 900 000).
7. **Injecting noise in the test sets does not improve performance** when applied to noiseless sets. That is, when trained with synthetic data *without errors*, the performance on test data *with errors* we observed that the RV RRE is not as good, and comparable to training *without errors* (as discussed in the next section). This is illustrated in Figs. 6 and 7.
8. **The precise time sampling does not have a strong influence** on the outcome, even with seasonal gaps. We remark, however, that all the test samplings used here cover the rotation period of both stars quite well (several cycles, more than one point per cycle). More tests would be needed to re-analyse old survey data, which typically spans several years and with observations clustered in seasons of a few months each year.
9. **The time-tags of the observations were NOT used** in the experiments presented here (Tables 4 & 5). Experiments ran with time-tags and all combinations of indices converged much more slowly than those without them. This is something we did not expect that requires further investigation. The outcome of a test including the time tags as inputs together with the BIS index for ε Eri is shown in Table 4.

3.2. Neural Network training with realistic observational errors

Since noise is an intrinsic characteristic of real data that shall also be learned, we considered whether we should *train* our NN with noisy synthetic observations. In the following list we summarise the conclusions drawn by running test with noisy training sets and presented in Tables 4 and 5, and Figs. 5:

1. **The performance of CCF indices** becomes quite affected when using noisy input indices (and their combinations) when their RRE exceeds 50% (FWHM and CON for AU Mic). Uncertainties in activity indices are not as well understood as for RV measurements, so this calls for a quick examination of possible periodic signals in all the indices to be used before deciding whether an index shall be used in the training and later de-trending of real time series. A real case example is discussed later in the section devoted to AU Mic real observations (Sec. 3.3.2), where we found that using the CON index (which does not show the same periodicities as the other indices) makes things worse.
2. **Along the same line, uncertainties in the photometry at few % level (which are quite easy to obtain, even from the ground), do not affect the inversion power of the photometry as an input.** To illustrate this in the tests presented here, the precision of the photometric time series was arbitrarily chosen to be at 10% level, which is rather high (see Table 2). Despite this high value, this does not influence the outcome very much confirming that the less inversion power of the photometry comes from intrinsic physical degeneracies.
3. **The RRE of an activity index should be 30–40% smaller than the contribution to the RV RRE caused by stellar activity variability.** If larger (FWHM and CON for AU Mic), the introduction of those errors degrades performances very much (RV RREs form 16 to 44 %, and 16 to 45 % in these example cases).

3.3. Application to observational data

3.3.1. Results on ε Eridani

Before jumping into generating a final set of synthetic training sets, we perform an exact inversion using the `starsim` code in order to verify that the priors set in our simulations matched what would be obtained if attempting a more classic full stellar fitting approach. This process is explained in detail in Appendix A. Most importantly, the `starsim` inversion resulted in differences in the stellar parameters compared to the literature ones, especially the stellar radius. With this new parameters, we then produce 100 000 `starsim` simulations of RV, BIS, FWHM, and CON time series, each using the same original time-sampling consisting in 66 observations. For the spot map, we use the random procedure as used in previous sections to avoid producing simulations too specific to the data. To train the NN, we add uncertainties and errors to the inputs and outputs as described in previous sections since in the case for the real observational data the NN performance is significantly enhanced.

We then train our NNs again on 50 training epochs achieving an RV RRE down to 10% on our 100 000 training sets, which is slightly larger but comparable to the experiments (7.5%) in the previous section.

When applied to the real dataset, we obtain the correction shown in Fig. 9. The black curve shows the observed RVs and the red curve shows the outputs of a trained NN with inputs of BIS,

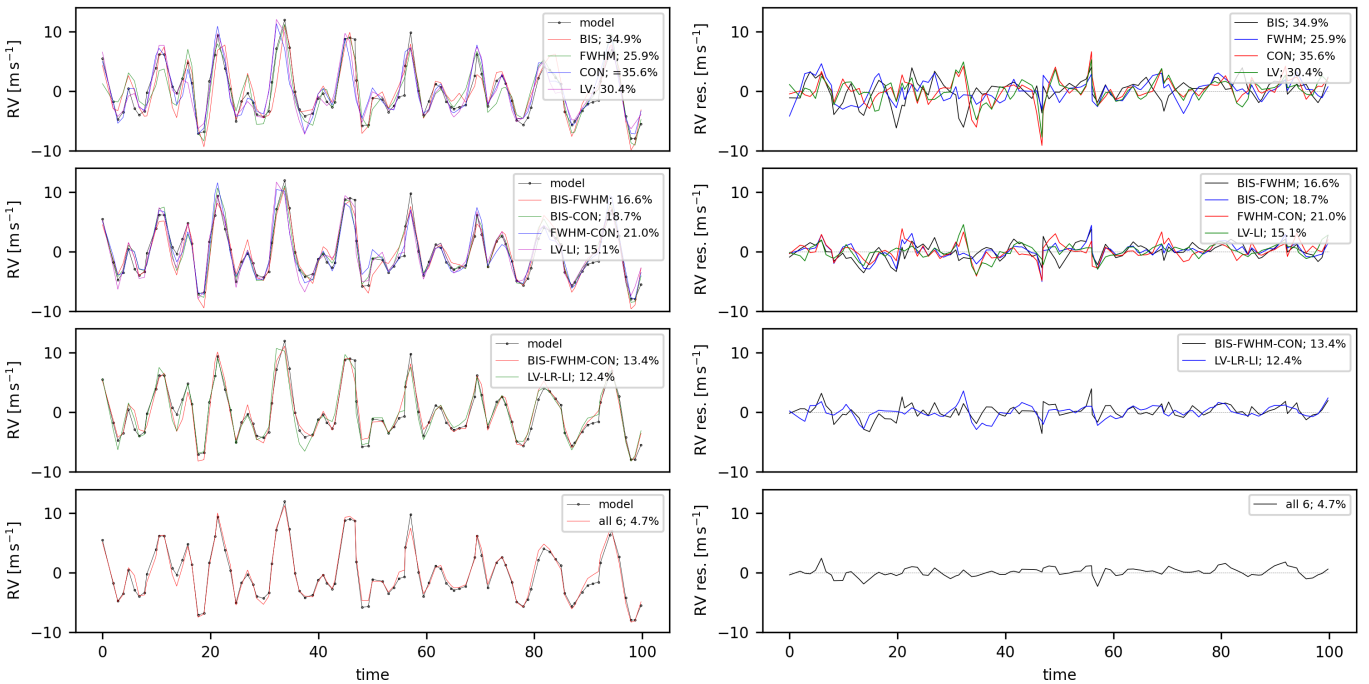


Fig. 8. Illustration of the quality of the RV correction as a function of the input indices used (random synthetic set chosen) for the ε Eri data and the *uniform* time-sampling (black lines). The output predictions of the neural networks with all the combinations of input data, i.e. index time series are also illustrated with different colours. The outcome of using only one index is shown in the top panel, further down two indices, then combinations of three indices (photometry and CCF), and then all six index time-series data. RVs are shown as modeled by *starsim*. The legends contain RV RRE of this particular set for each combination of index inputs. The residuals (differences between RV predictions and the simulated RV) are shown in the right panels.

FWHM, and CON from the observations. The periodograms of the real and predicted RVs are also shown in the panel below. The corresponding RV RRE is 45% (from 4.4 m s^{-1} to 2.0 m s^{-1}) which is quite far from the nominal 10% inferred from the training and test samples. We argue that at the level below few m s^{-1} , the physics of the activity induced effects are not yet properly included in the simulations and, possibly, some instrumental systematic errors are also affecting the result (e.g. uncertainty in the wavelength solution of each night of observation). More work on the models is needed to identify all possible sources of such mismatches. In any case, the reduction of the RMS below 2 m s^{-1} already places ε Eri as a high quality target (similar “jitter” as the Sun) for exoplanet searches (enough sensitivity to detect close-in super-Earths).

3.3.2. Results on AU Microscopii

We apply the same procedure to the case of AU Mic, but without performing the initial inversion with *starsim*. Instead, we use the stellar parameters as introduced in Table 1, the original time sampling containing 37 data points, and randomised spot maps as introduced in Sect. 2.3.

For this experiment, we create 400 000 new *starsim* simulations calculating RV, BIS, and FWHM, and we include again the observational errors (see Table 2) into the time series since it has been shown that it results in lower RV RREs. We then train our NN for 50 training epochs as before and reach an RV RRE of 10% for the test set. This is essentially the same value as in the previous test. We show the RV observations and the NN output in the top panel of Fig. 10. In the bottom panel, we show the periodograms of those two time series and add in green the periodogram of the difference.

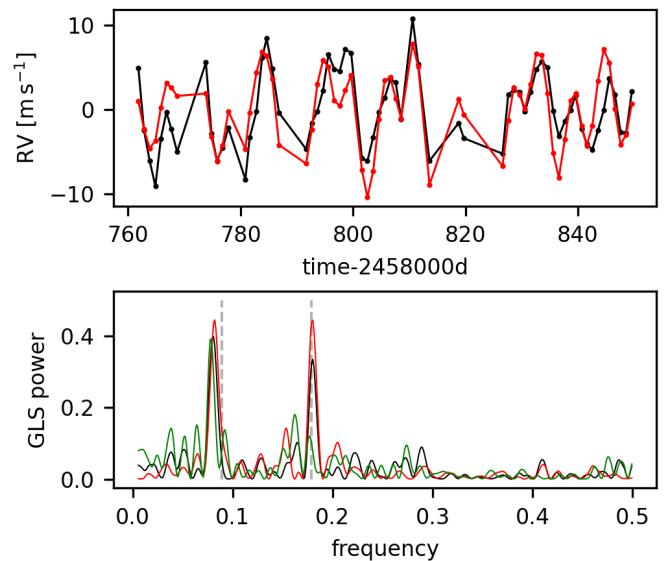


Fig. 9. RVs (top panel) and periodogram (bottom panel) of observational ε Eri data in black. In red, we show the prediction produced by our neural network using BIS, FWHM, and CON indices as input. In green we show the periodogram of the *cleaned* data, that is the difference between the observed and modeled RVs. The rotation period and its second harmonic are indicated as gray vertical dashed lines. The level of RMS reduction is about 45%.

As opposed to the expectations from all previous experiments, the inclusion of the CON index produced very poor performances when applied to the real data. After a close examination of the CON measurements and corresponding periodograms

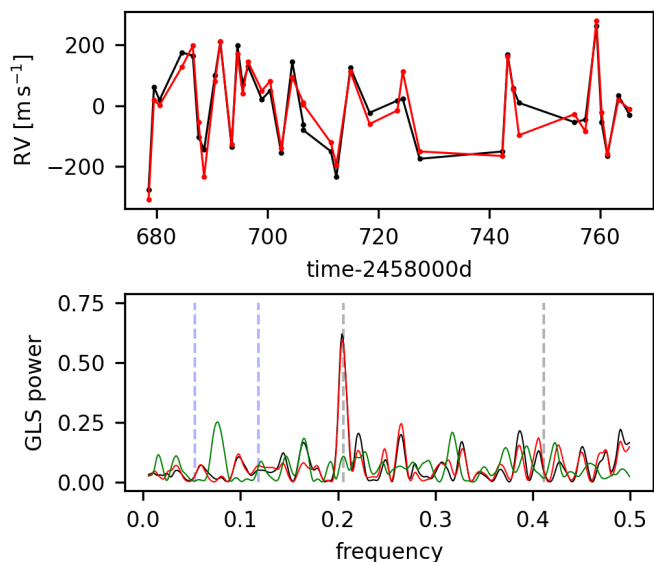


Fig. 10. top panel: RVs from AU Mic as observed (black) and as modeled by the trained neural network using BIS, and FWHM indices with their respective relative residual errors. bottom panel: periodogram of the observed RVs (black), the neural network output (red), and the difference (green). Black dashed lines indicate the rotation period and its second harmonic, blue dashed lines indicate the orbits of the two planetary companions of AU Mic. We reach down to RREs of 10%.

(see Fig. 2) it is clear that they do not contain the same characteristic periodicities as the other indices, or the RV. We attribute this to the fact that CON measurements from CARMENES are affected by additional noise sources not included in the simulations, including instrumental and astrophysical effects. For example, CON measurements are very sensitive to the continuum determination (and corresponding line-depths of the lines). Therefore, the source of this systematic error might be instrumental (e.g. change in the instrumental setup such as using simultaneous calibration lamps, sky background contribution, change in gain in the detector), but it could also be associated to astrophysical phenomena causing changes in the spectral continuum. For these reasons, we only use the BIS and FWHM indices for the performance analysis on real observations of AU Mic.

The correction obtained on the RVs is rather good, reaching to a RV RRE down to 10%, the same as for the *starsim* training sets. The cleaned RVs (subtraction of the computed correction to the observed ones), do not contain any trace of the rotation period. Moreover, the resulting scatter is consistent with the presence of the Doppler signals of the planets (although there are not enough observations to confirm them with only these 37 epochs). As opposed to ϵ Eri, most of the activity induced variability is clearly related to the dark spot effect on a fast rotating star (peak to peak variability over 100 m s^{-1}), which is one of the better understood RV noise phenomena. Although the relative improvement is rather good, we also note that the *residual* RV variability is around 13 m s^{-1} , which is substantially poorer than the one from ϵ Eri. This is again consistent with the models lacking relevant physical realism at the few m s^{-1} level. Using other de-trending techniques and evidence of periodic transits from photometric observations (Cale et al. 2021), the possible amplitude of two low mass companions have been inferred at the level of $\sim 10 \text{ m s}^{-1}$, which could explain part of this variability as well (Keplerian orbits have not been included in the training samples in any case). We show the residual RVs in Fig. 11,

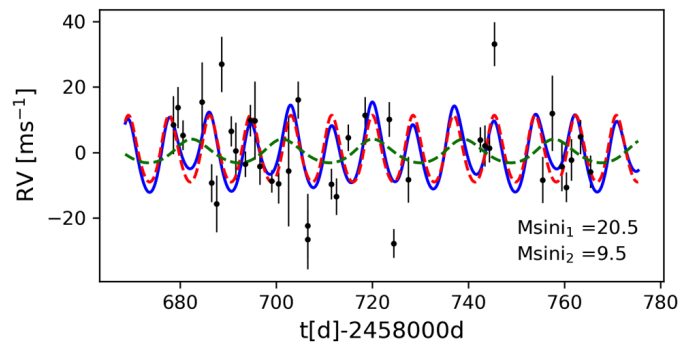


Fig. 11. Residual RVs of the 37 AU Mic observations after applying the neural network using BIS and FWHM indices (black dots). We further show the curves introduced to the RVs by the two detected planets of AU Mic (red and green dashed lines) and their combination (blue curve).

and illustrate the possible contribution of the signals of the two published planetary companions to the residual RV noise (red and green dashed curves for the individual planets, and the blue curve for their combination).

4. Conclusions

We developed a model-driven machine learning approach using convolutional deep neural networks to produce corrections to radial velocity time series contaminated by stellar activity. It is *model driven* in the sense that synthetic data sets were used to train the algorithms. Stellar activity manifests itself in other observables than RVs and this additional information was already used previously to mitigate the influence of activity on the RVs. However, it was not entirely clear whether the information contained in the individual and combined observables was sufficient to invert the problem and produce a correction in a more general and independent way, not using the information contained in the RVs. This work applied to the test data sets demonstrates that the inversion is possible and that it is solvable, without requiring explicit expressions for all the relations of the indices with the RV variability.

We identified neural network architectures that are well suited to the task (a few convolutional layers followed by a few dense layers) and that they can be trained with quite modest computational resources. Actually, producing realistic simulations is the most computer intensive part of the processing.

We assessed the inversion power of the classic activity indices used in exoplanet detection papers, and found that combinations of them are sufficient to provide corrections below the photon noise of state-of-the-art spectrometers when we consider the effects of spots combined with rotation and convective blueshift only ($\sim 10 \text{ cm s}^{-1}$, the goal of ESPRESSO). Photometric information obtained simultaneously with the observations did not have that much inversion power, but on its own can still produce significant reduction of RV noise. Non-simultaneous (space-based) multi-band photometry is likely to perform better but it will require additional work and, possibly, slightly different neural network architectural choices.

We applied our technique to two real datasets achieving a significant reduction of the activity induced RV variability, but the results are not yet at the level of the photon noise of the observations. This is likely caused by astrophysical phenomena yet to be included in the simulations, and due to systematic (instru-

mental and algorithmic) errors in the measurements of both the activity indicators and the RVs.

As a future step, we plan to test the Machine Learning framework with more models that include additional activity phenomena. In particular, models will have to include more complex phenomena such as non-uniform convective flows (granulation, supergranulation, meridional flows), magnetic distortion of spectral lines (Zeeman splitting), different sensitivities of some spectral lines to stellar activity, and the effect of bright and chromospherically active regions to apparent Doppler shifts (plages, faculae or others).

Acknowledgements. CARMENES is an instrument at the Centro Astronómico Hispano en Andalucía (CAHA) at Calar Alto (Almería, Spain), operated jointly by the Junta de Andalucía and the Instituto de Astrofísica de Andalucía (CSIC). CARMENES was funded by the Max-Planck-Gesellschaft (MPG), the Consejo Superior de Investigaciones Científicas (CSIC), the Ministerio de Economía y Competitividad (MINECO) and the European Regional Development Fund (ERDF) through projects FICTS-2011-02, ICTS-2017-07-CAHA-4, and CAHA16-CE-3978, and the members of the CARMENES Consortium (Max-Planck-Institut für Astronomie, Instituto de Astrofísica de Andalucía, Landessternwarte Königstuhl, Institut de Ciències de l’Espai, Institut für Astrophysik Göttingen, Universidad Complutense de Madrid, Thüringer Landessternwarte Tautenburg, Instituto de Astrofísica de Canarias, Hamburger Sternwarte, Centro de Astrobiología and Centro Astronómico Hispano-Alemán), with additional contributions by the MINECO, the Deutsche Forschungsgemeinschaft through the Major Research Instrumentation Programme and Research Unit FOR2544 “Blue Planets around Red Stars”, the Klaus Tschira Stiftung, the states of Baden-Württemberg and Niedersachsen, and by the Junta de Andalucía. We acknowledge financial support from the Agencia Estatal de Investigación of the Ministerio de Ciencia e Innovación MCIN/AEI/10.13039/501100011033 and the ERDF “A way of making Europe” through projects PID2020-120375GB-I00, PID2019-109522GB-C5[1:4], PGC2018-098153-B-C33, and the Centre of Excellence “Severo Ochoa” and “María de Maeztu” awards to the Institut de Ciències de l’Espai (CEX2020-001058-M) Instituto de Astrofísica de Canarias (CEX2019-000920-S), Instituto de Astrofísica de Andalucía (SEV-2017-0709), and Centro de Astrobiología (MDM-2017-0737); the Generalitat de Catalunya/CERCA programme; and the DFG through priority program SPP 1992 “Exploring the Diversity of Extrasolar Planets” (JE 701/5-1). Based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO programme 0104.C-0863(A); also called Red Dots. We also thank the Red Dots members Carole Haswell, Mario Damasso, Fabio del Sordo and Stefan Dreizler for their participation in the preparation of the Red Dots ESO observations with HARPS, their comments and insightful discussions.

References

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2015, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 252

Anglada-Escudé, G. & Butler, R. P. 2012, *ApJS*, 200, 15

Baines, E. K. & Armstrong, J. T. 2012, *ApJ*, 744, 138

Beck, J. G. 2000, *Sol. Phys.*, 191, 47

Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, 327, 977

Cale, B. L., Reefe, M., Plavchan, P., et al. 2021, *AJ*, 162, 295

Campbell, B., Walker, G. A. H., & Yang, S. 1988, *ApJ*, 331, 902

Collier Cameron, A., Mortier, A., Phillips, D., et al. 2019, *MNRAS*, 487, 1082

Cosentino, R., Lovis, C., Pepe, F., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84461V

Dattilo, A., Vanderburg, A., Shallue, C. J., et al. 2019, *AJ*, 157, 169

de Beurs, Z. L., Vanderburg, A., Shallue, C. J., et al. 2020, *arXiv e-prints*, arXiv:2011.00003

Dumusque, X., Boisse, I., & Santos, N. C. 2014, *ApJ*, 796, 132

Dumusque, X., Pepe, F., Lovis, C., et al. 2012, *Nature*, 491, 207

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306

Fröhlich, H. E. 2007, *Astronomische Nachrichten*, 328, 1037

Gaia Collaboration. 2020, *VizieR Online Data Catalog*, I/350

Gaidos, E., Mann, A. W., Lépine, S., et al. 2014, *MNRAS*, 443, 2561

Gomes da Silva, J., Santos, N. C., Adibekyan, V., et al. 2021, *A&A*, 646, A77

Gonzalez, G., Carlson, M. K., & Tobin, R. W. 2010, *MNRAS*, 403, 1368

Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, *A&A*, 582, A49

Herbst, K., Papaioannou, A., Airapetian, V. S., & Atri, D. 2021, *ApJ*, 907, 89

Herrero, E., Ribas, I., Jordi, C., et al. 2016, *A&A*, 586, A131

Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, *A&A*, 553, A6

Janson, M., Quanz, S. P., Carson, J. C., et al. 2015, *A&A*, 574, A120

Kalas, P., Graham, J. R., & Clampin, M. 2005, *Nature*, 435, 1067

Keenan, P. C. & McNeil, R. C. 1989, *ApJS*, 71, 245

Kervella, P., Arenou, F., Mignard, F., & Thévenin, F. 2019, *A&A*, 623, A72

Klein, B., Donati, J.-F., Moutou, C., et al. 2021, *MNRAS*, 502, 188

Krist, J. E., Ardila, D. R., Golimowski, D. A., et al. 2005, *AJ*, 129, 1008

Lafarga, M., Ribas, I., Lovis, C., et al. 2020, *A&A*, 636, A36

Liebing, F., Jeffers, S. V., Reiners, A., & Zechmeister, M. 2021, *A&A*, 654, A168

Llop-Sayson, J., Wang, J. J., Ruffio, J.-B., et al. 2021, *AJ*, 162, 181

Löhner-Böttcher, J., Schmidt, W., Stief, F., Steinmetz, T., & Holzwarth, R. 2018, *A&A*, 611, A4

Mayor, M., Pepe, F., Queloz, D., et al. 2003, *The Messenger*, 114, 20

Meunier, N., Lagrange, A.-M., Mbemba Kabuiku, L., et al. 2016, in *19th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun (CS19)*, 104

Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, *A&A*, 645, A96

Perger, M., Anglada-Escudé, G., Ribas, I., et al. 2021, *A&A*, 645, A58

Perger, M., García-Piquer, A., Ribas, I., et al. 2017, *A&A*, 598, A26

Perger, M., Scandariato, G., Ribas, I., et al. 2019, *A&A*, 624, A123

Plavchan, P., Barclay, T., Gagné, J., et al. 2020, *Nature*, 582, 497

Quillen, A. C. & Thorndike, S. 2002, *ApJ*, 578, L149

Quirrenbach, A., CARMENES Consortium, Amado, P. J., et al. 2020, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 11447, 114473C

Ribas, I., Reiners, A., Zechmeister, M., et al. 2023, *A&A*, in press.

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 014003

Rosich, A., Herrero, E., Mallonn, M., et al. 2020, *A&A*, 641, A82

Santos, N. C., Israelian, G., & Mayor, M. 2004, *A&A*, 415, 1153

Shallue, C. J. & Vanderburg, A. 2018, *AJ*, 155, 94

Stief, F., Löhner-Böttcher, J., Schmidt, W., Steinmetz, T., & Holzwarth, R. 2019, *A&A*, 622, A34

Suárez Mascareño, A., Rebolo, R., González Hernández, J. I., & Esposito, M. 2015, *MNRAS*, 452, 2745

Torres, C. A. O., Quast, G. R., da Silva, L., et al. 2006, *A&A*, 460, 695

Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 2198, Instrumentation in Astronomy VIII, ed. D. L. Crawford & E. R. Craine, 362

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

Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, *A&A*, 609, A12

Zicher, N., Barragán, O., Klein, B., et al. 2022, *MNRAS*, 512, 3060

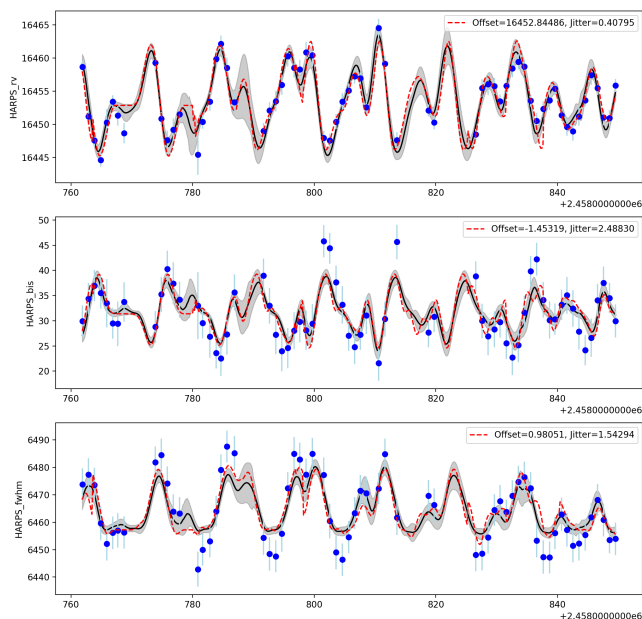


Fig. A.1. Results of the *starsim* inversions for ε Eri using RV (top panel), FWHM (middle panel), and BIS (bottom panel). Blue dots indicate the measurements as extracted from *raccoon* from the observed spectra. The black curve shows the average fit and the gray area the uncertainty of the different solutions. The dashed red curve shows the solution with the largest likelihood. On the top right, the offset for each time series is indicated as well as its additional jitter both showing the shortcomings of the modeling of time-series data with *starsim*.

Appendix A: *starsim* data inversion for ε Eridani

With the current version of the *starsim* code we calculate the inversion of the algorithm, i.e. using the observational time-series data of ε Eri to fit for stellar parameters and an average distribution of stellar spots. Differently from our initial approach, we use a fixed number of 10 different dark spots up to 15° in radius, appearing at any point of the calculated time interval randomly and stellar parameters floating around the measurements as given in Table 1. As a result, we achieve a best fit on observational RV, BIS, and FWHM time series as shown in Fig. A.1. This is achieved for the absolute offsets for the three data sets as given in the top right of each panel. In Table A.1, we show the differences for the stellar parameters from the literature and from our best data inversion. Especially R_* does differ significantly. Note, that ΔCB , i , and $d\Omega$ were fixed for the inversion test. A histogram of the posterior distributions for all the inversion iterations of the five parameters of the 10 spots are shown in Fig. A.2. We can see a clear clustering around longitudes 50 to 150 and 275 to 325° , as well as co-latitudes ranging between 20 and 40° . Spots show mostly 2 to 4° radius resulting in absolute spot filling factors of 2.5 to 3.5%. Spot lifetimes seem to be mostly in-between 40 to 60 d coinciding thereby with the results from the GP as shown in Fig. 3.

Appendix B: Tables of observational data

Table A.1. Differences of important stellar parameters from the literature and from our data inversion with *starsim*

ε Eri	inv.	lit.
spot temperature difference ΔT [K]	500	1080
convective blueshift difference CB [CB_\odot]	0	~ 0.3
rotation period P_{rot} [day]	11.33	11.2
stellar radius R_* [R_\odot]	0.85	0.74
stellar mass M_* [M_\odot]	0.85	0.82
inclination i [deg]	90	60
differential rotation $d\Omega$ [$d\Omega_\odot$]	0	1.3

Notes. See Table 1 for details.

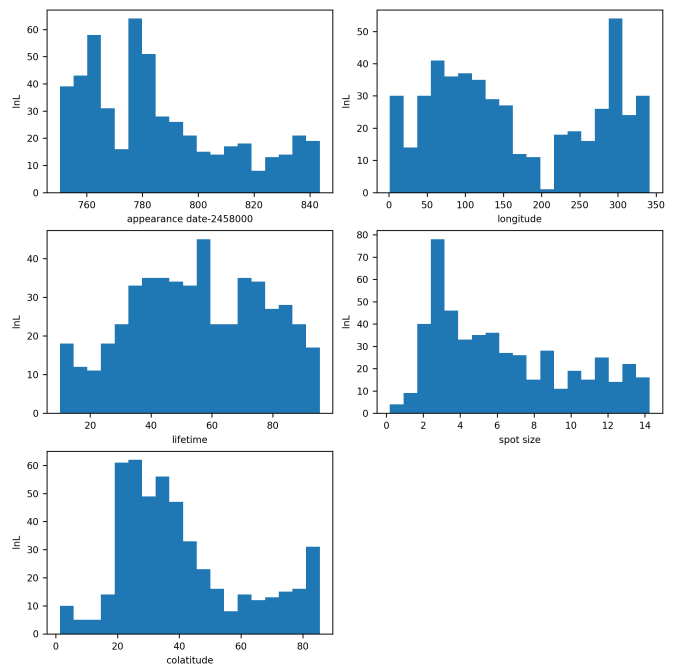


Fig. A.2. Histogram of the posterior distribution of the five parameters of the spot maps containing 10 spots of the *starsim* inversion applied to observational ε Eri data RV, BIS, and FWHM.

Table B.1. 66 nightly binned time-series data points extracted with *raccoon* from 204 HARPS spectra of ϵ Eri.

BJD (d)	RV (m s^{-1})	BIS (m s^{-1})	FWHM (m s^{-1})	CON (%)
2458761.86	16458.65±0.87	30.9±1.9	6473.8±5.7	40.694±0.029
2458762.86	16451.20±0.79	35.4±1.8	6477.3±5.9	40.700±0.029
2458763.87	16447.57±0.89	38.0±1.9	6473.5±6.1	40.747±0.031
2458764.88	16444.59±0.87	36.5±1.8	6459.6±6.0	40.802±0.030
2458765.85	16450.26±1.97	34.5±4.1	6452.2±6.0	40.846±0.030
2458766.85	16453.40±0.90	30.5±1.7	6456.1±5.7	40.794±0.029
2458767.71	16451.34±1.48	30.4±3.7	6457.0±6.1	40.796±0.029
2458768.84	16448.66±1.47	34.7±3.0	6456.3±5.6	40.773±0.030
2458773.86	16459.30±1.01	29.8±1.9	6481.9±5.5	40.662±0.029
2458774.85	16450.86±0.90	36.3±2.4	6484.5±5.8	40.657±0.029
2458775.83	16447.62±1.35	41.3±2.6	6474.1±6.1	40.698±0.031
2458776.85	16449.16±1.71	38.4±3.3	6463.9±6.1	40.762±0.030
2458777.86	16451.50±1.20	35.2±2.6	6463.2±5.8	40.746±0.029
2458780.88	16445.40±3.00	34.0±6.2	6442.8±6.2	40.823±0.030
2458781.64	16450.38±0.99	30.6±1.9	6449.9±5.6	40.798±0.028
2458782.84	16453.43±1.32	27.8±3.4	6453.0±5.6	40.792±0.028
2458783.85	16459.83±1.04	24.6±1.8	6464.0±5.5	40.732±0.028
2458784.66	16462.15±1.18	23.5±2.4	6479.0±5.5	40.683±0.027
2458785.65	16458.54±1.45	28.3±3.1	6487.6±5.5	40.661±0.027
2458786.86	16453.31±1.36	36.6±2.6	6485.2±6.0	40.647±0.030
2458791.67	16448.99±1.12	39.9±2.3	6454.3±6.0	40.782±0.030
2458792.66	16452.07±1.19	34.0±2.5	6448.4±5.9	40.801±0.029
2458793.67	16453.47±1.50	28.2±3.0	6447.6±5.6	40.808±0.028
2458794.68	16455.97±1.52	25.0±3.2	6455.8±6.4	40.777±0.029
2458795.65	16460.26±1.50	25.6±3.1	6472.3±5.7	40.724±0.029
2458796.68	16458.50±1.41	29.0±2.7	6484.9±6.0	40.671±0.029
2458797.64	16458.26±1.37	30.8±3.3	6482.9±5.9	40.672±0.029
2458798.64	16460.87±1.50	29.1±3.0	6477.4±5.6	40.665±0.028
2458799.68	16460.42±1.65	30.4±3.0	6484.9±5.5	40.619±0.028
2458801.59	16447.92±0.98	46.8±2.0	6477.2±6.1	40.635±0.031
2458802.57	16447.56±0.97	45.4±1.8	6460.4±6.1	40.719±0.031
2458803.58	16450.37±1.04	38.6±2.3	6449.0±5.9	40.792±0.030
2458804.63	16453.41±1.23	34.2±2.8	6446.3±5.8	40.816±0.030
2458805.64	16455.08±0.96	28.0±1.7	6454.5±5.7	40.799±0.028
2458806.63	16457.23±1.02	25.8±2.3	6463.3±5.6	40.770±0.029
2458807.60	16456.94±0.81	28.2±1.6	6471.5±5.6	40.740±0.028
2458808.56	16452.57±1.35	32.1±2.7	6470.5±5.8	40.737±0.029
2458810.60	16464.52±1.33	22.6±2.4	6472.3±5.5	40.683±0.027
2458811.61	16459.12±1.16	31.3±2.4	6484.8±5.5	40.617±0.028
2458813.60	16447.59±1.12	46.6±2.3	6461.7±6.2	40.708±0.031
2458818.79	16452.10±0.82	28.7±1.6	6469.7±5.5	40.754±0.028
2458819.72	16450.28±0.96	31.8±1.8	6466.3±5.8	40.757±0.029
2458826.59	16448.45±0.72	39.8±1.7	6448.1±5.8	40.784±0.030
2458827.61	16455.48±1.24	31.0±2.3	6448.5±5.8	40.808±0.029
2458828.59	16456.04±1.56	27.9±2.8	6454.4±5.7	40.810±0.029
2458829.61	16455.77±1.15	29.3±2.6	6464.5±5.6	40.781±0.029
2458830.58	16453.45±0.91	30.7±1.8	6467.7±5.7	40.768±0.029
2458831.59	16455.78±1.06	26.5±1.8	6463.8±4.2	40.765±0.026
2458832.57	16458.44±1.26	23.7±2.5	6469.6±5.4	40.708±0.028
2458833.58	16459.41±1.07	26.1±2.3	6474.7±5.6	40.665±0.027
2458834.58	16458.73±1.02	32.6±2.2	6476.5±5.4	40.635±0.027
2458835.58	16453.54±0.78	40.8±1.7	6472.4±5.7	40.643±0.029
2458836.60	16450.49±1.14	43.2±2.1	6453.3±6.0	40.729±0.030
2458837.67	16452.31±1.09	35.2±2.5	6447.3±5.9	40.775±0.030
2458838.68	16453.60±1.39	31.1±2.7	6447.1±5.6	40.788±0.029
2458839.59	16455.40±0.93	31.3±1.8	6456.1±5.6	40.766±0.028
2458840.69	16451.39±0.83	34.2±1.7	6462.9±5.9	40.753±0.030
2458841.55	16449.60±1.25	36.0±2.7	6457.2±5.9	40.791±0.030
2458842.55	16448.93±1.35	33.4±2.7	6451.5±5.9	40.822±0.030
2458843.55	16451.20±1.34	28.9±2.6	6452.2±6.0	40.798±0.030
2458844.58	16453.62±0.84	25.2±1.7	6455.4±5.6	40.754±0.028
2458845.58	16457.42±1.40	27.6±2.9	6461.9±4.7	40.705±0.024
2458846.58	16455.45±0.95	35.1±2.0	6468.1±5.6	40.670±0.028
2458847.57	16451.02±1.11	38.5±2.1	6460.8±5.9	40.706±0.030
2458848.57	16450.93±1.14	35.5±1.7	6453.5±4.5	40.748±0.023
2458849.57	16455.84±0.90	30.9±2.1	6454.0±5.7	40.755±0.029

Table B.2. 37 nightly binned time-series data extracted with raccoon from 75 CARMENES VIS spectra for AU Microscopii.

BJD (d)	RV (m s ⁻¹)	BIS (m s ⁻¹)	FWHM (m s ⁻¹)	CON (%)
2458678.57	-5342.0±8.9	315.5±24.6	12687.3±55.4	10.092±0.035
2458679.54*	-5003.1±6.2	-87.2±15.6	12175.3±43.6	7.370±0.021
2458680.54	-5046.2±4.6	-63.0±11.6	12285.4±39.7	9.843±0.025
2458684.57	-4889.9±12.0	-244.6±30.3	12253.9±37.4	10.254±0.025
2458686.55	-4899.8±5.8	-332.9±14.4	12385.5±44.7	10.124±0.029
2458687.58	-5168.7±8.7	-6.0±22.0	12638.0±46.1	10.068±0.029
2458688.58	-5209.1±8.4	187.5±21.8	12275.3±46.7	10.042±0.030
2458690.56	-4965.3±4.6	-201.1±11.3	12425.4±51.3	9.799±0.032
2458691.51	-4854.7±8.8	-380.8±21.6	12653.4±58.2	10.266±0.037
2458693.55	-5200.6±3.9	21.0±9.8	12264.1±54.0	9.895±0.035
2458694.60	-4865.5±4.9	-272.3±12.0	12334.8±55.4	10.209±0.036
2458695.54	-4994.6±12.0	-98.0±30.6	12471.0±54.3	10.247±0.035
2458696.52	-4935.8±5.7	-273.6±14.1	12794.4±72.0	9.978±0.044
2458699.00	-5043.8±3.4	-104.1±8.5	12240.9±43.4	10.005±0.028
2458700.48	-5015.3±6.0	-119.6±14.8	12303.3±54.1	10.211±0.036
2458702.50	-5220.2±16.9	67.3±42.6	12679.6±100.9	10.057±0.063
2458704.49	-4920.8±5.6	-189.4±14.0	12219.0±54.7	10.124±0.036
2458706.49	-5126.3±9.7	7.2±24.9	12618.9±101.1	9.729±0.062
2458706.50	-5145.3±9.0	-6.1±23.0	12638.4±104.0	9.672±0.063
2458711.45	-5214.7±4.7	171.6±12.5	12567.0±61.0	9.608±0.037
2458712.45	-5299.4±5.6	225.9±15.1	12481.2±63.8	9.888±0.040
2458714.98	-4939.9±4.1	-133.8±10.2	12037.4±29.1	10.195±0.020
2458718.45	-5087.8±5.6	86.4±14.1	12125.5±39.6	10.303±0.027
2458723.46	-5047.8±5.3	-41.7±13.2	12101.2±35.5	10.283±0.024
2458724.42	-5042.3±4.4	-151.8±10.9	12349.0±46.3	9.914±0.030
2458727.47	-5239.3±6.9	21.5±17.3	12432.7±67.3	9.721±0.042
2458742.39	-5215.3±3.8	111.3±9.9	12342.7±63.9	9.736±0.040
2458743.37	-4897.4±6.3	-267.0±15.4	12156.9±50.3	9.854±0.032
2458744.35	-5008.4±4.4	-13.0±11.3	12055.4±41.0	10.105±0.027
2458745.39	-5055.2±6.7	68.0±17.3	12758.8±60.9	9.828±0.037
2458755.40	-5119.0±7.0	-41.5±17.6	12372.5±56.2	9.948±0.036
2458757.38	-5110.3±11.5	116.6±29.7	12055.5±42.2	10.044±0.028
2458759.37	-4801.7±7.5	-477.6±18.4	12297.0±45.6	10.384±0.031
2458760.31	-5119.5±4.5	-92.3±11.2	12326.5±40.9	9.727±0.026
2458761.35	-5229.1±6.4	112.6±16.3	12412.7±45.4	10.218±0.030
2458763.34	-5031.1±5.2	-50.2±13.0	12372.9±49.9	10.100±0.032
2458765.33	-5094.2±5.0	-150.0±12.3	12478.2±51.0	9.990±0.032

Notes. * marks the obvious data outlier for the CON measurement.