ODUSSEAS: a machine learning tool to derive effective temperature and metallicity for M dwarf stars[★]

A. Antoniadis-Karnavas^{1,2}, S. G. Sousa¹, E. Delgado-Mena¹, N. C. Santos^{1,2}, G. D. C. Teixeira^{1,2}, and V. Neves^{3,4}

¹ Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
e-mail: alexandros.antoniadis@astro.up.pt

² Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal

³ Instituto Federal do Paraná, Campus Foz do Iguaçu, 85860000 Foz do Iguaçu-PR, Brazil

⁴ Casimiro Montenegro Filho Astronomy Center, Itaipu Technological Park, 85867-900 Foz do Iguaçu-PR, Brazil

Received 26 November 2019 / Accepted 19 February 2020

ABSTRACT

Aims. The derivation of spectroscopic parameters for M dwarf stars is very important in the fields of stellar and exoplanet characterization. The goal of this work is the creation of an automatic computational tool able to quickly and reliably derive the T_{eff} and $[\text{Fe}/\text{H}]$ of M dwarfs using optical spectra obtained by different spectrographs with different resolutions.

Methods. ODUSSEAS (Observing Dwarfs Using Stellar Spectroscopic Energy-Absorption Shapes) is based on the measurement of the pseudo equivalent widths for more than 4000 stellar absorption lines and on the use of the machine learning Python package “scikit-learn” for predicting the stellar parameters.

Results. We show that our tool is able to derive parameters accurately and with high precision, having precision errors of ~ 30 K for T_{eff} and ~ 0.04 dex for $[\text{Fe}/\text{H}]$. The results are consistent for spectra with resolutions of between 48 000 and 115 000 and a signal-to-noise ratio above 20.

Key words. stars: fundamental parameters – stars: atmospheres – stars: late-type – methods: data analysis – techniques: spectroscopic

1. Introduction

Spectra can be used to reveal the chemical composition of the stars, as well as important stellar atmospheric parameters, such as effective temperature (T_{eff}) and $[\text{Fe}/\text{H}]$. These parameters are crucial for the characterization of the stars and are therefore fundamental to understanding their formation and evolution. Furthermore, these stars influence the properties of the planets forming and orbiting around them (Everett et al. 2013). However, the spectroscopic analysis to derive these parameters has some difficulties to overcome. One of the main challenges is the correct determination of the spectral continuum, which is more problematic in cool and faint stars, such as M dwarfs. Their study is quite difficult and complicated compared to that of FGK stars, because in M dwarfs, molecules are the dominant sources of opacity. These molecules create thousands of spectral lines that are poorly known and moreover many of them blend with each other. Therefore, the position of the continuum is hardly identified in their spectra.

Methods that rely on the correct determination of the continuum only work well for the metal-poor and earliest types of M dwarfs (Woolf & Wallerstein 2005). Methods using spectral synthesis have not achieved as precise results as in FGK cases because of the poor knowledge of many molecular line strengths. Recently, spectral synthesis in the near-infrared has led to advances, as shown by several studies (Önehağ et al. 2012;

Lindgren et al. 2016; Rajpurohit et al. 2018; Passegger et al. 2019).

Regarding these limitations, most attempts at determining effective temperature and metallicity use photometric calibrations (Bonfils et al. 2005; Johnson & Apps 2009; Neves et al. 2012) or spectroscopic indices (Rojas-Ayala et al. 2010, 2012; Mann et al. 2013a). Uncertainties in metallicity range from 0.20 dex using photometric calibrations to 0.10 dex when using spectroscopic scales in the infrared (Rojas-Ayala et al. 2012). For T_{eff} , precisions of 100 K are reported, but significant uncertainties and systematic errors are still present, ranging from 150 to 300 K (Casagrande et al. 2008; Rojas-Ayala et al. 2012).

One of the most popular methods to derive atmospheric stellar parameters for FGK stars is by measuring the equivalent widths (EWs) of many metal lines of the spectrum. Neves et al. (2014) used the MCAL code to measure pseudo EWs in the optical part of the spectrum for 110 M dwarfs observed in the HARPS GTO M dwarf program by setting a pseudo continuum for each line. These latter authors proceeded to the derivation of T_{eff} and $[\text{Fe}/\text{H}]$ of these stars applying a calibration based on reference photometric T_{eff} and $[\text{Fe}/\text{H}]$ scales that exist for 65 of them from Casagrande et al. (2008) and Neves et al. (2012), respectively. In the first case, the reference T_{eff} is the average value of the $V-J$, $V-H$, and $V-K$ photometric scales as seen in Casagrande et al. (2008), while for $[\text{Fe}/\text{H}]$ the calculation of its reference values was done using stellar parallaxes, and V and K s magnitudes as described in Neves et al. (2012).

Machine learning is an increasingly popular tool and is used in several fields of science. It can be accurate in predicting

[★] ODUSSEAS can be tested by downloading the files from <https://github.com/AlexandrosAntoniadis/ODUSSEAS>, after reading the README instructions for clarifying the technical details.

outcomes without the need for the user to explicitly create a specific model to the problem at hand. The algorithms in machine learning receive input data and by applying statistical analysis, they predict an output value within a reasonable range. The interest for machine learning algorithms and automatic processes in astronomy is emerging from the increasing volume of survey data (Howard 2017). These tools can be applied to a wide range of studies, with the input attributes being, for example, the photometric properties of the sources (Das & Sanders 2019; Akras et al. 2019; Rau et al. 2019; Ucci et al. 2019).

In our work, we follow the pseudo-EW approach. We present our tool ODUSSEAS (Observing Dwarfs Using Stellar Spectroscopic Energy-Absorption Shapes), which makes use of the machine learning “scikit learn” package of Python. It offers a quick automatic derivation of T_{eff} and $[\text{Fe}/\text{H}]$ for M dwarf stars using their 1D spectra and resolutions as input. The main advantage of this tool when compared to other ones that derive stellar parameters, such as the MCAL code (Neves et al. 2014; which is limited to the range of HARPS spectrograph and needs manual adjustment of results for different resolutions), is that it can operate simultaneously in an automatic fashion for spectra of different resolutions and different wavelength ranges in the optical. It is based on a supervised machine learning algorithm, meaning that it is provided with both input and expected output and uses these to create a model. The input to the machine learning function are the values of the pseudo EWs for 65 HARPS spectra and the expected output are the values of their reference T_{eff} and $[\text{Fe}/\text{H}]$ from Casagrande et al. (2008) and Neves et al. (2012), respectively. After training with a part of these HARPS data, the algorithm produces a model and tests it on the rest of the data; it predicts their values and compares them with the reference ones given as expected output. Thus, it examines the accuracy and the precision of the model using several regression metrics described below. Finally, it applies the model to unknown spectra and estimates their stellar parameters.

In Sect. 2 we describe how ODUSSEAS computes the pseudo EWs. In Sect. 3 we provide a detailed description of ODUSSEAS and the flow of its process. We explain the characteristics of the machine learning function and its efficiency regarding different regression types, resolutions, and wavelength areas. In Sect. 4 we apply our tool to spectra obtained by several spectrographs of various resolutions and we examine the results. Finally, Sect. 5 summarizes the work presented in this paper.

2. Pseudo-EW measurements

Since the identification of the continuum is very difficult in the spectra of M dwarfs, we follow the way of setting a pseudo continuum in each absorption line. The method is based on measurements of the pseudo EWs of absorption lines and blended lines in the range between 530 and 690 nm. We have excluded the parts where the activity-sensitive Na doublet and $\text{H}\alpha$ lines and strong telluric lines reside. The line list consists of 4104 features, which are given in the form of left and right boundaries, between which these absorption features are supposed to be created. This method, based on pseudo EWs and the specific line list, was used by Neves et al. (2014).

We created our own Python version of the method to compute the pseudo EWs. Our code reads the line list and the 1D fits files of the stellar spectra. We set an option for radial-velocity correction of the input spectra by our code for cases where they are shifted. For each line, the code identifies the position of the minimum flux of the feature, which is the central absorption wavelength. Starting from this position, the code identifies

Fig. 1. Area fitting for the calculation of pseudo EW for a line with central $\lambda = 6531.4$ Å of the star G1176. The position of the pseudo continuum is adjusted accordingly. This pseudo EW is equal to 87 mÅ.

the maximum on each side of this absorption feature, after having cut this spectral area at the range defined by the respective boundaries provided in the line list. Eventually, the code fits the pseudo continuum along the edges of the absorption feature with a straight line and obtains the pseudo EW by calculating the area between the pseudo continuum and the flux. Mathematically, the pseudo EW is defined as follows, where F_{pp} is the value of the flux between the peaks of the feature (i.e., the pseudo continuum) and F_{λ} is the flux of the line at each integration step:

$$\text{pseudoEW} = \sum \frac{(F_{\text{pp}} - F_{\lambda})}{F_{\text{pp}}} \Delta\lambda. \quad (1)$$

We present an example in Fig. 1 where we use the star G1176 and an absorption line at the region around 6530 Å.

An evaluation of our pseudo-EW measurements, by comparing them with the ones obtained from MCAL code, is presented in Appendix B.

3. Machine learning on M dwarfs

We base our tool for the derivation of T_{eff} and $[\text{Fe}/\text{H}]$ on the machine learning concept. The user needs to run two codes. The “HARPS_dataset.py” creates the databases that contain pseudo-EW measurements in different resolutions and the reference stellar parameters. The “ODUSSEAS.py” measures the pseudo EWs of new stellar spectra and derives their unknown T_{eff} and $[\text{Fe}/\text{H}]$ via machine learning. Below, we explain the details of their structure, describing the input parameters and how to use the codes.

3.1. The HARPS dataset

Each time the code “HARPS_dataset.py” runs, the outcome is a file which is used later as input to the machine learning algorithm when running “ODUSSEAS.py” for training the machine and testing the generated model. It contains the names of 65 stars of the HARPS M dwarf sample, the central wavelengths of the 4104 absorption features from 530 to 690 nm, their pseudo-EW values according to the resolution at which we convolve the spectra, and their reference values of T_{eff} and $[\text{Fe}/\text{H}]$ from Casagrande et al. (2008) and Neves et al. (2012) respectively. All of these

Fig. 2. Distribution of reference T_{eff} and $[\text{Fe}/\text{H}]$ of the 65 stars used to train and test the machine learning models. The cross symbol represents the uncertainties of their photometric derivations, which are 100 K and 0.17 dex respectively.

65 spectra, presented in Table A.1, have a signal-to-noise ratio (S/N) of more than 100, as reported by Neves et al. (2014). The range of the reference stellar parameters is presented in Fig. 2. Their photometric derivations have uncertainties of 100 K for T_{eff} and 0.17 dex for $[\text{Fe}/\text{H}]$, as reported by Casagrande et al. (2008) and Neves et al. (2012) respectively.

The convolution function we use is the “instrBroadGauss-Fast” of “pyAstronomy”¹, which applies Gaussian instrumental broadening. The width of the kernel is determined by the resolution².

Since the HARPS spectra have a specific finite resolution, our code calculates the actual resolution to which they need to be convolved by the function in order to obtain spectra at the final resolution required. This calculation is done considering the following relation:

$$\sigma_{\text{conv}} = \sqrt{\sigma_{\text{final}}^2 - \sigma_{\text{orig}}^2}, \quad (2)$$

where σ_{conv} corresponds to the resolution to which we need to convolve a spectrum with an original resolution of σ_{orig} in order to obtain a final resolution of σ_{final} .

Two settings are input by the user. First, one chooses whether or not to convolve the reference HARPS spectra to the spectral resolution of the new data. We provided precomputed pseudo EWs for a range of spectral resolutions in widely used spectrographs. In that case there is no need to convolve the spectra again and recalculate the pseudo EWs. Second, the user must define the resolution of the data being analyzed. The “HARPS_dataset.py” is presented schematically in Fig. 3.

3.2. ODUSSEAS tool

ODUSSEAS.py makes use of two algorithms that we developed: the New_data.py algorithm for measuring the pseudo EWs of new spectra to analyze, and the MachineLearning.py algorithm for the derivation of their T_{eff} and $[\text{Fe}/\text{H}]$. The innovative aspect of this tool is the simultaneous predictions for spectra of different resolutions and wavelength ranges.

The user has the option to activate the automatic radial velocity correction for the spectra if they are shifted. In addition, the user can set the regression type to be used by the machine

learning process. The “ridge” is recommended, but also “ridgeCV” and “linear” work at a similar level of efficiency as well. Section 3.4 presents the efficiency of all the regression types used.

The workflow of “New_data.py” is similar to that of “HARPS_dataset.py”; it reads the files and resolutions of new spectra and, if needed, it calculates and corrects their radial velocity shift. In addition, if the original step of a spectrum is not 0.010, i.e., equal to that of the HARPS dataset, it is changed with linear interpolation to this value. Thus, the pseudo EWs are measured in a consistent way. The files containing the pseudo-EW measurements of each spectrum are then used during the operation of “MachineLearning.py”, which returns the values of T_{eff} and $[\text{Fe}/\text{H}]$ along with the regression metrics of the models that predicted them. The diagram of “ODUSSEAS.py” is presented in Fig. 4 showing its inputs, operations, and output in a concise manner.

3.3. Machine learning function

Here we present the machine learning function in more detail. The machine learning algorithm operates in a loop for each star separately, as each star may have a different wavelength range and resolution. For each star in the file list, the algorithm automatically loads two files: the HARPS dataset of the respective resolution, for training and testing the model, and the pseudo EWs of the star for which we want to derive T_{eff} and $[\text{Fe}/\text{H}]$, in order to apply the model and return the stellar parameters. Based on the wavelength range of each spectrum, a mask is applied to the HARPS dataset in order to consider the absorption lines in common between spectra.

The 65 HARPS stars are randomly split into training groups consisting of about 70% of the sample (45 stars) and testing groups consisting of the remaining 30% of the population (20 stars). With these numbers selected, the machine learning model can be both trained accurately and tested on a sufficient number of stars. This is done 100 times for each star for which we want to determine the parameters.

We provide the algorithm with different regression types to be used: the “linear”, the “ridge”, the “ridgeCV”, the “multi-task Lasso”, and the “multi-task Elastic Net”. All these kinds of models provide an output value by fitting a linear regression to the input values. The relation between the predicted value y (the stellar parameter), the input variables x (the pseudo EWs), and the coefficients w is expressed as

$$y(w, x) = w_o + w_1 x_1 + \dots + w_p x_p. \quad (3)$$

The mathematical details of each regression type are described in the official online documentation³.

The performance of machine learning is indicated by the following three kinds of returned regression metrics. The mean absolute error is computed when the model is applied to the test dataset. This corresponds to the expected value of the absolute error loss in the predictions. In addition, the “explained variance score” is calculated. The best possible value of this score is 1.0. Variance is the expectation of the squared deviation of a random variable from its mean. It measures how far a set of numbers spread out from their average value. Furthermore, the “r2 score” computes the coefficient of determination, defined as R^2 . The coefficient of determination is the proportion of the variance in the dependent variable that is predictable from the independent variables. This score provides a measure of how well future

¹ <https://github.com/sczesla/PyAstronomy>

² A description of it can be found at <https://www.hs.uni-hamburg.de/DE/Ins/Per/Czesla/PyA/PyA/pyasIDoc/aslDoc/broad.html>

³ https://scikit-learn.org/stable/modules/linear_model.html

Fig. 3. Workflow of `HARPS_dataset.py`. Fig. 4. Workflow of `ODUSSEAS.py`.

samples are likely to be predicted by the model. The best possible score is 1.0. A constant model that always predicts the expected value, disregarding the input features, would get a score of 0.0. In our case of multi-output, the resulting explained variance and r^2 scores are by default the averages with uniform weight of the respective scores for T_{eff} and $[\text{Fe}/\text{H}]$. The mathematical types of those regression metrics are described in the official online documentation⁴.

For each star, the tool makes 100 determinations by randomly splitting the training and testing groups each time. After these determinations, it returns the average values of T_{eff} and $[\text{Fe}/\text{H}]$, the average values of the mean absolute errors of the models,

the average scores of machine learning, and the dispersion of T_{eff} and $[\text{Fe}/\text{H}]$ (measured as the standard deviation). This iterative process minimizes the possible dependence of the resulting parameters on how the stars from the HARPS dataset are split for training and testing in one single measurement. As there are only 65 reference stars, the results of the measurements could be affected by the selection of stars that end up in the training set. This is why multiple runs are performed, shuffling and splitting the reference stars into different training and testing groups each time; we then calculate the average values and the dispersion. The final results are automatically saved in the file named “Parameter_Results.dat”. Moreover, the tool saves a group of plots with the reference and the predicted parameters of model testing, as well as their differences, as a visualization of the model accuracy. An example is presented in Fig. 5.

⁴ https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics

Fig. 5. Demonstration of predictions applying ridge regression. *Upper panel:* T_{eff} values expected (Ref.) and predicted (M.L.) on the test dataset, along with their differences. *Lower panel:* $[\text{Fe}/\text{H}]$ values expected (Ref.) and predicted (M.L.) on the test dataset, along with their differences.

3.4. Machine learning efficiency

First, we test the regression models mentioned above to find the best one. We use the original spectra of the HARPS dataset at their real resolution of 115 000. For 100 runs with each regression type, we measure the scores and the absolute mean errors of the stellar parameters on the test set. We report the average values around which each model tends to result in Table 1. The linear, ridge, and ridgeCV models work very well in general, giving r^2 and explained variance scores with average values of around 0.93 and 0.94, respectively. The range of these scores from the

Table 1. Average values of the scores and the mean absolute errors (M.A.E.) for T_{eff} and $[\text{Fe}/\text{H}]$ of the test dataset after 100 runs with each regression type.

Regression	r^2 score	E.V. score	M.A.E. T_{eff} [K]	M.A.E. $[\text{Fe}/\text{H}]$ [dex]
Ridge	0.93	0.94	27	0.037
RidgeCV	0.93	0.94	27	0.038
Linear	0.93	0.93	27	0.039
Multi-task Elastic Net	0.91	0.92	35	0.045
Multi-task Lasso	0.88	0.89	41	0.056

Table 2. Average values of the scores and the mean absolute errors (M.A.E.) for T_{eff} and $[\text{Fe}/\text{H}]$ of the test dataset after 100 runs at each resolution (using the “ridge” regression).

Resolution	r^2 score	E.V. score	M.A.E. T_{eff} [K]	M.A.E. $[\text{Fe}/\text{H}]$ [dex]
real 115 000	0.93	0.94	27	0.037
conv. 110 000	0.93	0.94	28	0.038
conv. 94 600	0.93	0.93	28	0.039
conv. 75 000	0.93	0.93	29	0.041
conv. 48 000	0.92	0.93	30	0.043

100 runs is usually from 0.87 to 0.99. The average uncertainties of those regression types are ~ 27 K for T_{eff} and ~ 0.04 for $[\text{Fe}/\text{H}]$. The ridge model leads to slightly greater scores than the linear one. The ridgeCV model, which has a built-in cross validation function that applies “leave-one-out” or “k-fold” strategies, does not seem to work better than the classic ridge one, at least for this sample of M dwarf measurements. Furthermore, multi-task Elastic Net and multi-task Lasso give considerably lower scores and higher mean absolute errors. Therefore, we advocate the use of ridge regression, as it operates best on the spectral values of the M dwarfs.

Second, we evaluate the explained variance and r^2 scores and the mean absolute errors of the algorithm for different resolutions of the spectra. We do this for the HARPS dataset at its actual resolution of 115 000 and we repeat this test for convolved datasets at resolutions of other broadly used spectrographs: 110 000 (UVES), 94 600 (CARMENES), 75 000 (SOPHIE), and 48 000 (FEROS). This is done to examine the level of machine learning precision towards lower resolutions. We present the average values from 100 measurements of each case in Table 2.

To further test the reliability of the method, we examine the efficiency of the machine learning in different wavelength ranges of the spectrum. We divide the line list, which is from 530 to 690 nm, in four spectral regions and we calculate the respective scores and mean absolute errors. We do this test to analyze whether machine learning works better when using the full range or a specific part of the wavelengths. For this test, we use the case of the convolved data at the resolution of 110 000. The machine learning operates at its best while using the full range of the initial line list. In addition, regarding the divided areas, we notice that using the bluer the part of the wavelength range leads to higher scores and lower mean absolute errors. In general, the results show that we can obtain highly precise predictions for stars observed at any part of the 530–690 nm spectrum. These results are presented in Table 3.

Table 3. Average values of the scores and the mean absolute errors (M.A.E.) for T_{eff} and $[\text{Fe}/\text{H}]$ of the convolved-to-110 000 dataset after 100 runs with different wavelength ranges of the line list.

Wavelength range (nm)	Number of lines	r2 score	E.V. score	M.A.E. T_{eff} [K]	M.A.E. $[\text{Fe}/\text{H}]$ [dex]
530–690	4104	0.93	0.94	28	0.038
530–580	1300	0.92	0.93	31	0.039
580–630	1300	0.91	0.91	48	0.044
630–690	1504	0.89	0.90	56	0.048

4. Derivation of stellar parameters

We apply our tool to spectra obtained using five widely used instruments of different resolutions: HARPS of 115 000, UVES of 110 000, CARMENES of 94 600, SOPHIE of 75 000, and FEROS of 48 000. The spectra were taken from the respective public data archives. To test the efficiency of our tool on spectra obtained using instruments other than HARPS, we use spectra from stars in common with the HARPS dataset, so that we can compare their results with the reference parameters of the respective HARPS spectra. To further validate the accuracy of our tool, we proceed to determinations and comparisons for a different selection of stars. Finally, we discuss possible future improvements of our determinations.

4.1. Resolution and spectral shape

We examine the spectral change of M dwarfs according to convolution at different resolutions. The shapes of M dwarf spectra are different when obtained at lower resolutions. In general, the lower the resolution, the shallower the absorption lines. This is illustrated in Fig. 6 where three lines of Gl176 are shown in detail, both for the original HARPS spectrum and the convolved ones to several resolutions. We also measure these lines and report their pseudo-EW values in Table 4 to show their differences. The relative differences can vary, as not only is the depth different but so is the location of the pseudo continuum in each case. All the differences confirm that the lower resolution always leads to lower pseudo-EW values. This is why we need to convolve the HARPS spectra to the respective resolutions of the new spectra. Consequently, machine learning compares the pseudo EWs of the same resolution and accurately predicts the stellar parameters. Figure 7 shows the spectral shapes of Gl674 for three different cases: the original HARPS spectrum with resolution 115 000, the convolved HARPS spectrum to the resolution of FEROS (48 000), and the original FEROS spectrum, which is the lowest resolution we examine. We notice that the convolved HARPS spectrum follows the shape of the FEROS one in a consistent way. Figure 8 shows the comparison of the pseudo EWs of the SOPHIE spectrum for Gl908 and the spectrum of the same star by HARPS before and after its convolution. The SOPHIE spectrum, which is of a lower resolution, has consistently lower pseudo-EW values than the HARPS one, as expected. After the convolution of the HARPS spectrum to the resolution of SOPHIE, the overall trend of their pseudo-EW values become highly compatible.

4.2. Measurements on different spectrographs

We examine the performance of our tool in new spectra. We show the accuracy of the stellar parameters predicted and the

Fig. 6. Shape of the HARPS original spectrum for Gl176 and convolved in different resolutions. The lower the resolution the shallower the absorption lines.

precision for each resolution by presenting the mean absolute errors of the models and the dispersion of the results, as calculated after 100 determinations for each spectrum.

For the case of HARPS, we use a HARPS spectrum of Gl643 with $S/N = 83$, which is not part of the HARPS dataset used in the machine learning. As reference values for this star, we consider its parameters reported by [Neves et al. \(2014\)](#). For the cases of the other instruments, we use a UVES spectrum of Gl846 with $S/N = 149$, a CARMENES spectrum of Gl514 with $S/N = 191$, a SOPHIE spectrum of Gl908 with $S/N = 90$ and a FEROS spectrum of Gl674 with $S/N = 61$. As reference values to those spectra, we consider the values of the respective HARPS ones in the dataset.

The results of T_{eff} and $[\text{Fe}/\text{H}]$ are presented in Table 5. We notice that the parameters of the new spectra are very close to the respective reference values. The differences in T_{eff} vary up to ~ 50 K and the differences in $[\text{Fe}/\text{H}]$ vary up to 0.03 dex. The mean absolute errors of models and the dispersions of values increase slightly with decreasing resolution.

4.3. Measurements at different S/N

Here we examine the possible variation of the results regarding different S/N for a given spectrum. We take the spectrum Gl514 of CARMENES, which has the highest S/N of those we examine (equal to 191, as reported in the CARMENES data archive) and we inject amounts of noise which correspond to lower S/N values that we set. Since the final noise is obtained by the quadratic sum of the initial noise and the injected noise, the final S/N values are

Table 4. Pseudo EWs of three absorption lines for HARPS spectrum G1176 at different resolutions.

Resolution	p.EW of λ 6536.67 [mÅ]	p.EW of λ 6537.08 [mÅ]	p.EW of λ 6537.64 [mÅ]
original 115 000	14.19	29.08	27.31
convolved 110 000	14.04	28.89	26.98
convolved 94 600	13.39	28.27	26.35
convolved 75 000	11.99	27.45	25.08
convolved 48 000	7.50	22.50	20.04

Notes. The lower the resolution, the smaller the pseudo EW.

Fig. 7. Shape of spectra for G1674 at the original HARPS resolution (blue), at the FEROS resolution (green), and at the HARPS convolved to FEROS resolution (orange).

calculated using the relation below.

$$\left(\frac{1}{S/N}\right)_{\text{final}}^2 = \left(\frac{1}{S/N}\right)_{\text{initial}}^2 + \left(\frac{1}{S/N}\right)_{\text{injected}}^2. \quad (4)$$

We create new spectra with final S/N values ranging from 100 to 9. For each spectrum, we measure the stellar parameters and their dispersion. Figure 9 illustrates the measurements of the CARMENES spectrum while degrading its S/N. Overall, the results are similar to those for the original spectrum and the differences are kept roughly constant with respect to the reference values. For S/N values down to 20, we notice that the dispersions are between 17 and 27 K for T_{eff} and between 0.03 and 0.04 dex for [Fe/H], i.e., at similar levels to those for the original spectrum. For S/N values below 20, the dispersions start to increase up to ~ 50 K and up to ~ 0.07 dex respectively. Moreover, it seems that there is a slight decrease of the order of 20 K in T_{eff} and a slight increase of the order of 0.02 dex in [Fe/H] for the spectra with S/N below 20. However, these results are within the uncertainties of the tool. Therefore, we conclude that our tool works consistently for spectra with S/N above 20. Below this S/N, the errors increase significantly.

4.4. Comparison of results between our tool and Neves et al. (2014)

Now, we make an overall comparison of our results on a group of HARPS spectra with the ones presented by Neves et al. (2014). For this purpose, we measure 30 HARPS spectra from the initial

Fig. 8. Upper panel: pseudo-EW values of the original G1908 spectra for HARPS and SOPHIE. Lower panel: pseudo-EW values of G1908 after the convolution of HARPS to the resolution of SOPHIE. The units of pseudo EWs are mÅ. After the convolution, there is agreement between the identity line (solid green) and the slope (dashed red), with the intersection being close to zero.

GTO sample, for which we do not know their parameters from photometry. These HARPS spectra are not part of the machine learning dataset we use. Based on the information from Neves et al. (2014), we excluded very active stars and stars with S/N lower than 25, below which the method of these latter authors does not apply. Both methods were tested and shown to not work properly for very active or young stars, since the pseudo EWs of such spectra are affected and their parameters cannot be determined accurately with the pseudo-EW approach we follow. Subsequently, we compare the results we obtain using our tool with the results presented by Neves et al. (2014).

The errors of the stellar parameters derived using our tool are 27 K for T_{eff} and 0.04 dex for [Fe/H], as the mean absolute errors are measured when the machine learning model is applied on the

Fig. 9. Average differences and the dispersion of T_{eff} (upper panel) and $[\text{Fe}/\text{H}]$ (lower panel) for CARMENES Gl514 spectrum with different values of S/N.

test dataset. The errors of the calibration by [Neves et al. \(2014\)](#), which are quantified from the root mean squared error (RMSE) in that work, are equal to 91 K and 0.08 dex respectively. We highlight the fact that both methods are tied to the same initial systematic uncertainties of the reference parameters used, which are 100 K for T_{eff} and 0.17 dex for $[\text{Fe}/\text{H}]$. The results and their differences are presented in Table 6 and Fig. 10. The mean and median difference in T_{eff} is 11 and 22 K respectively, with a standard deviation of 101 K. Regarding $[\text{Fe}/\text{H}]$, the mean and median difference is -0.04 dex, with a standard deviation of 0.06 dex.

Work by [Neves et al. \(2014\)](#) follows a traditional approach, using a least-squares weighted fit to determine parameters. The regression of our tool reduces those errors of T_{eff} and $[\text{Fe}/\text{H}]$ from 91 to 27 K and from 0.08 to 0.04 dex respectively. Therefore, our machine learning approach significantly increases the precision of parameter determinations. In terms of speed, the determination of T_{eff} and $[\text{Fe}/\text{H}]$ for a star using machine learning, including multiple runs, shuffling and splitting the training and test samples each time, requires only a few seconds.

4.5. Estimating total uncertainties

Intrinsic uncertainties exist in the T_{eff} and $[\text{Fe}/\text{H}]$ reference values of the HARPS dataset, because their initial photometric

Fig. 10. T_{eff} comparison (upper panel) and $[\text{Fe}/\text{H}]$ comparison (lower panel) between this work and [Neves et al. \(2014\)](#).

derivations have average uncertainties of 100 K and 0.17 dex respectively. Since these parameters are used as the training values for the machine learning process, we decided to inject these uncertainties by perturbing their values accordingly in order to see how the final results of the predictions vary.

To this end, we create Gaussian distributions on the parameters for each HARPS training dataset, increasing the dispersion of distribution of the reference parameters each time in steps of 10 K and 0.02 dex until reaching uncertainties of 100 K and 0.17 dex. This adds different training values to the machine learning algorithm each time. For each step, we created 100 Gaussian-distributed training datasets. After these runs of machine learning, we calculated the average values of predicted parameters and their dispersion.

In Fig. 11, we present the variations for spectra from the highest resolution (HARPS), the lowest resolution (FEROS), and an intermediate resolution (CARMENES). The datapoints represent the average difference between the resulting parameters after being calculated with the 100 different datasets and the reference values. The error bars show the dispersion of the difference. We notice that the average differences from the reference values are almost the same among different Gaussian distributions, regardless of the amount of uncertainty injected to the Gaussian distribution.

The average results of T_{eff} and $[\text{Fe}/\text{H}]$ for the spectra from all the instruments are presented in Table 7. We report their maximum errors after considering the maximum Gaussian distribution with 100 K and 0.17 dex. Overall, the average values of the parameters remain roughly the same as those calculated with no Gaussian distribution at all. The mean absolute errors of the machine learning models have grown to values of between

Table 5. Machine learning (M.L.) results of T_{eff} and $[\text{Fe}/\text{H}]$, their dispersion (Disp.), the mean absolute errors (M.A.E.) of the models, and the reference values (Ref.) for comparison.

Star	Spec.	Res.	Ref. T_{eff} [K]	M.L. T_{eff} [K]	M.A.E. T_{eff} [K]	Disp. T_{eff} [K]	Ref. [Fe/H] [dex]	M.L. [Fe/H] [dex]	M.A.E. [Fe/H] [dex]	Disp. [Fe/H] [dex]
Gl643	HARPS	115 000	3102	3113	27	10	-0.26	-0.28	0.04	0.01
Gl846	UVES	110 000	3682	3691	28	13	-0.08	-0.05	0.04	0.02
Gl514	CARMENES	94 600	3574	3547	28	17	-0.13	-0.13	0.04	0.03
Gl908	SOPHIE	75 000	3587	3580	28	18	-0.38	-0.35	0.04	0.03
Gl674	FEROS	48 000	3284	3338	30	24	-0.18	-0.16	0.04	0.03

Table 6. Stellar parameters of 30 HARPS spectra as calculated by Neves et al. (2014) and using our tool (AA) and their difference.

Star	T_{eff} (AA) [± 27 K]	T_{eff} (Ne14) [± 91 K]	T_{eff} diff. [K]	[Fe/H] (AA) [± 0.04 dex]	[Fe/H] (Ne14) [± 0.08 dex]	[Fe/H] diff. [dex]
CD-44-836A	3104	3032	72	-0.07	-0.07	0.00
G108-21	3214	3186	28	-0.02	-0.02	0.00
GJ1057	2926	2916	10	-0.11	-0.10	-0.01
GJ1061	2772	2882	-110	-0.25	-0.09	-0.16
GJ1065	3106	3082	24	-0.32	-0.23	-0.09
GJ1123	2971	2779	192	-0.02	0.15	-0.17
GJ1129	3037	3017	20	-0.02	0.05	-0.07
GJ1236	3225	3280	-55	-0.44	-0.47	0.03
GJ1256	2964	2853	111	-0.02	0.06	-0.08
GJ1265	3020	2941	79	-0.28	-0.20	-0.08
G112	3245	3239	6	-0.31	-0.29	-0.02
G1145	3297	3270	27	-0.27	-0.28	0.01
G1203	3174	3138	36	-0.31	-0.22	-0.09
G1299	3078	3373	-295	-0.53	-0.53	0.00
G1402	3052	2943	109	0.00	0.03	-0.03
G1480.1	3214	3211	3	-0.48	-0.48	0.00
G1486	3096	2941	155	-0.02	0.03	-0.05
G1643	3113	3102	11	-0.29	-0.26	-0.03
G1754	2988	3005	-17	-0.23	-0.14	-0.09
L707-74	3250	3353	-103	-0.39	-0.38	-0.01
LHS1134	3007	2950	57	-0.20	-0.13	-0.07
LHS1481	3342	3510	-168	-0.66	-0.76	0.10
LHS1723	3031	3167	-136	-0.29	-0.24	-0.05
LHS1731	3229	3273	-44	-0.22	-0.19	0.03
LHS1935	3222	3181	41	-0.20	-0.22	0.02
LHS337	3003	3007	-4	-0.33	-0.27	-0.06
LHS3583	3205	3236	-31	-0.13	-0.22	0.09
LHS3746	3111	3013	98	-0.17	-0.13	-0.04
LHS543	3042	2872	170	0.17	0.23	-0.06
LP816-60	3030	2960	70	-0.11	-0.07	-0.04

Notes. The S/Ns of those stars are between 28 and 97, as reported by Neves et al. (2014).

65 and 80 K for T_{eff} and between 0.10 and 0.13 dex for $[\text{Fe}/\text{H}]$, depending on the resolution of the HARPS dataset. The dispersion of the derived parameters grows as the resolution of the spectra decreases. Specifically, the dispersion is smaller than the injected uncertainties for the HARPS spectrum (~ 60 K and ~ 0.10 dex), while for the spectra from other instruments, it is slightly higher than the uncertainties injected (~ 110 to ~ 130 K and ~ 0.18 to ~ 0.22 dex respectively).

However, in all cases the resulting average values of stellar parameters are very close to their expected values. Differences in

T_{eff} are up to ~ 40 K and differences in $[\text{Fe}/\text{H}]$ are up to 0.03 dex, with respect to the expected values.

4.6. Validation of $[\text{Fe}/\text{H}]$ determinations by measuring binary systems

Here, we measure $[\text{Fe}/\text{H}]$ in binary systems containing M dwarfs which are not part of the reference sample used for machine learning. Thus, we validate our method of $[\text{Fe}/\text{H}]$ prediction in an independent way. We determine $[\text{Fe}/\text{H}]$ both in FGK+M

Fig. 11. Average differences and the dispersions of the results for several amounts of Gaussian distribution injected to the reference parameters of the training HARPS datasets. The result of each step is the average outcome from 100 different distributed datasets.

and in M+M systems for an even more intrinsic test of $[\text{Fe}/\text{H}]$ agreement.

The $[\text{Fe}/\text{H}]$ determinations of eight FGK+M binary systems from spectra obtained by UVES and FEROS spectrographs are presented in Table 8. Regarding the FGK stars, their $[\text{Fe}/\text{H}]$ and respective uncertainties were derived using the methodology described in Sousa et al. (2008) and Santos et al. (2013). The method measures the equivalent widths of FeI and FeII lines

and assumes ionization and excitation equilibrium. It makes use of the radiative transfer code MOOG (Snedden 1973) and a grid of Kurucz model atmospheres (Kurucz 1993). The $[\text{Fe}/\text{H}]$ values of the respective M dwarf secondaries, derived by ODUSSEAS, are presented along with the total uncertainties of our tool at the resolutions of UVES (0.10 dex) and FEROS (0.13 dex). All binaries have differences that are within the uncertainties of the methods.

Table 7. Machine learning (M.L.) results of T_{eff} and $[\text{Fe}/\text{H}]$ after injecting uncertainties with Gaussian distributions of 100 K and 0.17 dex in the parameters of the training HARPS datasets, their dispersion (Disp.), the mean absolute errors (M.A.E.) of the models, and the reference values (Ref.) for comparison.

Star	Spec.	Res.	Ref. T_{eff} [K]	M.L. T_{eff} [K]	M.A.E. T_{eff} [K]	Disp. T_{eff} [K]	Ref. [Fe/H] [dex]	M.L. [Fe/H] [dex]	M.A.E. [Fe/H] [dex]	Disp. [Fe/H] [dex]
Gl643	HARPS	115 000	3102	3126	65	60	-0.26	-0.29	0.10	0.10
Gl846	UVES	110 000	3682	3678	68	109	-0.08	-0.06	0.10	0.18
Gl514	CARMENES	94 600	3574	3545	77	113	-0.13	-0.13	0.12	0.19
Gl908	SOPHIE	75 000	3587	3585	78	120	-0.38	-0.36	0.13	0.21
Gl674	FEROS	48 000	3284	3324	80	138	-0.18	-0.15	0.13	0.22

Table 8. $[\text{Fe}/\text{H}]$ difference between members of FGK+M binary systems.

Primary	[Fe/H] [dex]	$\sigma[\text{Fe}/\text{H}]$ [dex]	Secondary	[Fe/H] [dex]	$\sigma[\text{Fe}/\text{H}]$ [dex]	[Fe/H] difference [dex]
Gl100A	-0.29	0.05	Gl100C	-0.29	0.13	0.00
Gl118.1A	0.02	0.05	Gl118.1B	0.05	0.10	0.03
Gl173.1A	-0.37	0.03	Gl173.1B	-0.29	0.10	0.08
Gl157A	-0.08	0.04	Gl157B	0.02	0.13	0.10
NLTT19073	0.08	0.03	NLTT19072	-0.07	0.13	-0.15
NLTT29534	0.00	0.03	NLTT29540	-0.03	0.10	-0.03
NLTT34137	-0.12	0.05	NLTT34150	-0.13	0.10	-0.01
NLTT34353	-0.10	0.03	NLTT34357	-0.11	0.10	-0.01

Table 9. $[\text{Fe}/\text{H}]$ difference between members of M+M binary systems.

Primary	[Fe/H] [dex]	Disp. [dex]	Secondary	[Fe/H] [dex]	Disp. [dex]	[Fe/H] difference [dex]
Gl553	-0.07	0.06	Gl553.1	-0.10	0.07	0.03
Gl875	-0.15	0.05	Gl875.1	-0.17	0.06	0.02
Gl617A	0.04	0.04	Gl617B	-0.08	0.08	0.12
Gl745A	-0.48	0.04	Gl745B	-0.53	0.06	0.05
Gl752A	0.01	0.03	Gl752B	-0.07	0.08	0.08

Furthermore, we proceed to $[\text{Fe}/\text{H}]$ determinations of stars in five M+M binary systems, measuring their available spectra from the CARMENES public archive. In Table 9, we present these results along with their own dispersions, since both are estimated by our tool based on the same reference values with the same initial uncertainties. We notice agreement between the respective members of all the M+M binaries, within the dispersions of their $[\text{Fe}/\text{H}]$ determinations. This is a validation that our tool predicts $[\text{Fe}/\text{H}]$ in a consistent and accurate way.

4.7. Discussion on the reference parameter scales

Since supervised machine learning determines the parameters based on reference values given to it, their systematic errors will also apply to the results of new stars. In this work, we used the reference T_{eff} and $[\text{Fe}/\text{H}]$ photometric scales of Casagrande et al. (2008) and Neves et al. (2012) respectively, as they are derived in a homogeneous way for a sufficiently large number of spectra available to us. It is important to make a comparison between the reference values we use and values for the same stars derived in other recent studies that may be subject to different systematic errors, such as Mann et al. (2015), with which we share 26 of

the 65 stars that we use as our reference dataset. In Table 10, we compare our reference parameters with determinations by Mann et al. (2015) and report the differences. These differences are illustrated in Fig. 12. Regarding T_{eff} , we notice that our reference values have an average underestimation of 178 K with a standard deviation of 73 K. This systematic difference originates from the different methods of derivation used. Work by Casagrande et al. (2008) is based on the multiple optical-infrared technique (MOITE) for M dwarfs, which is an extension of the infrared flux method (IRFM) as described in Casagrande et al. (2006). On the other hand, determinations by Mann et al. (2015) were made by comparing the optical spectra with the CFIST suite of the BT-SETTL version of the PHOENIX atmosphere models (Allard et al. 2013). A detailed description of this method can be found in Mann et al. (2013b). Regarding $[\text{Fe}/\text{H}]$, we notice no significant systematic difference in parameter values resulting from the methods of calibration by Neves et al. (2012) and Mann et al. (2015). The average difference is 0.06 dex with a standard deviation of 0.11 dex for the sample of stars in common.

As a potential future improvement of our determinations, we consider the possibility of replacing our reference dataset. As

Table 10. Stellar parameters of 26 stars in common with Mann et al. (2015) and their difference.

Star	T_{eff} (Ref.) [± 100 K]	T_{eff} (Mann15) [± 60 K]	T_{eff} diff. [K]	[Fe/H] (Ref.) [± 0.17 dex]	[Fe/H] (Mann15) [± 0.08 dex]	[Fe/H] diff. [dex]
G154.1	2091	3056	-65	-0.40	-0.26	-0.14
G187	3565	3638	-73	-0.30	-0.36	0.06
G1105B	3054	3284	-230	-0.14	-0.12	-0.02
G1176	3369	3680	-311	0.02	0.14	-0.12
G1205	3497	3801	-304	0.17	0.49	-0.32
G213	3026	3250	-224	-0.19	-0.22	-0.03
G1250B	3369	3481	-112	-0.09	0.14	-0.25
G1273	3107	3317	-210	-0.05	-0.11	0.06
G1382	3429	3623	-194	0.04	0.13	-0.09
G1393	3396	3548	-154	-0.13	-0.18	0.05
G1436	3277	3479	-202	0.01	0.01	0.00
G1447	2952	3192	-240	-0.23	-0.02	-0.21
G1514	3574	3727	-153	-0.13	-0.09	-0.04
G1526	3545	3649	-104	-0.18	-0.31	0.13
G1555	2987	3211	-224	0.13	0.17	-0.04
G1581	3203	3395	-192	-0.18	-0.15	-0.03
G1686	3542	3657	-115	-0.29	-0.25	-0.04
G1699	3094	3228	-134	-0.59	-0.40	-0.19
G1701	3535	3614	-79	-0.20	-0.22	0.02
G1752A	3336	3558	-222	0.04	0.10	-0.06
G1846	3682	3848	-166	-0.08	0.02	-0.10
G1849	3200	3530	-330	0.24	0.37	-0.13
G1876	3059	3247	-188	0.14	0.17	-0.03
G1880	3488	3720	-232	0.05	0.21	-0.16
G1887	3560	3688	-128	-0.20	-0.06	-0.14
G1908	3587	3646	-59	-0.38	-0.45	-0.07

Fig. 12. T_{eff} comparison (upper panel) and $[\text{Fe}/\text{H}]$ comparison (lower panel) between the reference values we use and Mann et al. (2015) for 26 common stars.

new techniques for parameter determination become more accurate and precise and as more spectra become available to us, their homogeneously derived parameters can be correlated with their pseudo EWs. Therefore, we take into account the creation of an improved reference dataset for our machine learning tool, that could consist of more reference stars and of reference stellar parameters derived with greater accuracy and precision.

5. Summary

We present our machine learning tool ODUSSEAS for the derivation of T_{eff} and $[\text{Fe}/\text{H}]$ in M dwarf stars, whose spectra are of varying resolution at wavelengths ranging between 530 and 690 nm. We provide a detailed explanation of how ODUSSEAS was built and how it works. We present the results of the tests we perform and we examine their accuracy and precision on spectra with resolutions that vary from 115 000 down to 48 000. Our tool seems to be reliable, as it operates with high machine learning scores of around 0.94 and achieves excellent predictions of significantly high precision with mean absolute errors of ~ 30 K for T_{eff} and ~ 0.04 dex for $[\text{Fe}/\text{H}]$. Taking into consideration the intrinsic uncertainties of the reference parameters and perturbing them accordingly, our models have maximum uncertainties of ~ 80 K for T_{eff} and ~ 0.13 dex for $[\text{Fe}/\text{H}]$, which are within the typical uncertainties for M dwarfs. Our parameters for spectra from different spectrographs, obtained from the average of 100 determinations, have consistent values with differences within ~ 50 K and ~ 0.03 dex from the expected ones. Spectra are required to have a S/N of above 20 for optimal predictions. Our tool is valid for M dwarfs in the intervals 2800–4000 K for T_{eff} and -0.83 to 0.26 dex for $[\text{Fe}/\text{H}]$, except from very active or young stars.

Acknowledgements. This work was supported by FCT/MCTES through national funds and by FEDER – Fundo Europeu de Desenvolvimento Regional through COMPETE2020 – Programa Operacional Competitividade e Internacionalização by these grants: UID/FIS/04434/2019, UIDB/04434/2020 and UIDP/04434/2020; PTDC/FIS-AST/32113/2017 and POCI-01-0145-FEDER-032113; PTDC/FIS-AST/28953/2017 and POCI-01-0145-FEDER-028953. A.A.K., S.G.S., E.D.M. and G.D.C.T. acknowledge the support from FCT in the form of the exploratory projects with references IF/00028/2014/CP1215/CT0002, IF/00849/2015/CP1273/CT0003 and IF/00956/2015/CP1273/CT0002. S.G.S. and E.D.M. further acknowledge the support from FCT through the Investigador FCT contracts IF/00028/2014/CP1215/CT0002, IF/00849/2015/CP1273/CT0003 and POCH/FSE (EC). G.D.C.T. further acknowledges the support from an FCT/Portugal PhD grant with reference PD/BD/113478/2015.

References

- Akras, S., Leal-Ferreira, M. L., Guzman-Ramirez, L., & Ramos-Larios, G. 2019, *MNRAS*, **483**, 5077
- Allard, F., Homeier, D., Freytag, B., et al. 2013, *Mem. Soc. Astron. It. Suppl.*, **24**, 128
- Bonfils, X., Delfosse, X., Udry, S., et al. 2005, *A&A*, **442**, 635
- Bonfils, X., Delfosse, X., Udry, S., et al. 2013, *A&A*, **549**, A109
- Casagrande, L., Portinari, L., & Flynn, C. 2006, *MNRAS*, **373**, 13
- Casagrande, L., Flynn, C., & Bessell, M. 2008, *MNRAS*, **389**, 535
- Das, P., & Sanders, J. L. 2019, *MNRAS*, **484**, 294
- Everett, M. E., Howell, S. B., Silva, D. R., et al. 2013, *ApJ*, **771**, 107
- Howard, E. M. 2017, *Astron. Data Anal. Softw. Syst.* **XXV**, 512, 245
- Johnson, J. A., & Apps, K. 2009, *ApJ*, **699**, 933
- Kurucz, R. 1993, *ATLAS9 Stellar Atmosphere Programs and 2 km/s grid. Kurucz CD-ROM No. 13*. (Smithsonian Astrophysical Observatory: Cambridge), 13
- Lindgren, S., Heiter, U., & Seifahrt, A. 2016, *A&A*, **586**, A100
- Mann, A. W., Brewer, J. M., Gaidos, E., Lépine, S., & Hilton, E. J. 2013a, *AJ*, **145**, 52
- Mann, A. W., Gaidos, E., & Ansdell, M. 2013b, *ApJ*, **779**, 188
- Mann, A. W., Feiden, G. A., Gaidos, E., et al. 2015, *ApJ*, **804**, 64
- Neves, V., Bonfils, X., Santos, N. C., et al. 2012, *A&A*, **538**, A25
- Neves, V., Bonfils, X., Santos, N. C., et al. 2014, *A&A*, **568**, A121
- Önehag, A., Heiter, U., Gustafsson, B., et al. 2012, *A&A*, **542**, A33
- Passegger, V. M., Schweitzer, A., Shulyak, D., et al. 2019, *A&A*, **627**, A161
- Rajpurohit, A. S., Allard, F., Rajpurohit, S., et al. 2018, *A&A*, **620**, A180
- Rau, M. M., Koposov, S. E., Trac, H., & Mandelbaum, R. 2019, *MNRAS*, **484**, 409
- Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2010, *ApJ*, **720**, L113
- Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2012, *ApJ*, **748**, 93
- Santos, N. C., Sousa, S. G., Mortier, A., et al. 2013, *A&A*, **556**, A150
- Snedden, C. A. 1973, *PhD thesis*, The University of Texas at Austin, Texas, USA
- Sousa, S. G., Santos, N. C., Mayor, M., et al. 2008, *A&A*, **487**, 373
- Ucci, G., Ferrara, A., Gallerani, S., et al. 2019, *MNRAS*, **483**, 1295
- Woolf, V. M., & Wallerstein, G. 2005, *MNRAS*, **356**, 963

Appendix A: Additional table

Table A.1. Reference values of the HARPS spectra used for the machine learning.

Star	[Fe/H] [dex]	T_{eff} [K]
GI1	-0.40	3528
GI54.1	-0.40	2901
GI87	-0.30	3565
GI105B	-0.14	3054
HIP 12961	-0.12	3904
LP771-95A	-0.51	3393
GJ163	0.00	3223
GI176	0.02	3369
GJ179	0.14	3076
GI191	-0.79	3679
GI205	0.17	3497
GI213	-0.19	3026
GI229	-0.04	3586
HIP 31293	-0.04	3312
HIP 31292	-0.11	3158
GI250B	-0.09	3369
GI273	-0.05	3107
GI300	0.09	2965
GJ2066	-0.09	3388
GJ317	0.22	3130
GI341	-0.14	3633
GJ1125	-0.15	3162
GI357	-0.30	3335
GI358	0.01	3240
GI367	-0.09	3452
GI382	0.04	3429
GI393	-0.13	3396
GJ3634	-0.02	3332
GI413.1	-0.06	3373
GI433	-0.13	3450
GI436	0.01	3277
GI438	-0.31	3536
GI447	-0.23	2952
GI465	-0.54	3382
GI479	0.05	3238
GI514	-0.13	3574
GI526	-0.18	3545
GI536	-0.13	3546
GI555	0.13	2987
GI569A	0.16	3235
GI581	-0.18	3203
GI588	0.07	3284
GI618A	-0.05	3242
GI628	-0.05	3107
GJ1214	0.03	2856
GI667C	-0.47	3431
GI674	-0.18	3284
GJ676A	0.10	3734
GI678.1A	-0.10	3611
GI680	-0.07	3395
GI682	0.09	3002
GI686	-0.29	3542
GI693	-0.28	3188
GI699	-0.59	3094
GI701	-0.20	3535
GI752A	0.04	3336
GI832	-0.17	3450
GI846	-0.08	3682
GI849	0.24	3200
GI876	0.14	3059
GI877	-0.01	3266
GI880	0.05	3488
GI887	-0.20	3560
GI908	-0.38	3587
LTT9759	0.17	3316

Notes. T_{eff} and [Fe/H] were derived photometrically by Casagrande et al. (2008) and Neves et al. (2012), respectively.

Appendix B: Evaluation of our pseudo-EW measurements

Fig. B.1. Upper panel: comparison between the pseudo-EW values measured by our tool and the MCAL code by Neves et al. (2014). Lower panel: percentage difference of the pseudo-EW values plotted against our values.

We calculated the pseudo EWs of 4104 lines for the 110 stars of the total HARPS sample. Here we compare our values with the ones obtained by the MCAL code. In the upper panel of Fig. B.1, we present the comparison of all the pseudo-EW values for the star GI176 as an example. The units of pseudo EWs are mÅ. Inside the plots, AA stands for our measurements and VN stands for the measurements by Neves et al. (2014). The slope in the diagrams of most stars is almost identical with the identity line, with only a few pseudo EWs having considerably different values. In the lower panel of Fig. B.1 we show the relative difference (the percentage) of the values against our values. A scatter appears for the pseudo-EW values smaller than 30 mÅ, which is normal, as the relative difference for these narrow lines is greater. In contrast, nearly all lines broader than 50 mÅ are measured with significant agreement. The quality test for measuring the pseudo EWs comes from the following comparison. We plot the mean differences and mean relative differences between our method and the code by Neves et al. (2014) for each line averaged over all the stars in order to see how all the lines are measured. The result is very good as can be seen in Fig. B.2, with only 184 lines out of 4104 showing a mean relative difference greater than $\pm 15\%$.

Fig. B.2. *Upper panel:* difference of each line as averaged over the measurements of all stars. *Lower panel:* percentage difference of each line as averaged over the measurements of all stars.