

# The CARMENES search for exoplanets around M dwarfs

## A deep transfer learning method to determine $T_{\text{eff}}$ and $[M/H]$ of target stars<sup>★</sup>

A. Bello-García<sup>1</sup>, V. M. Passegger<sup>2,3,4,5</sup>, J. Ordieres-Meré<sup>6</sup>, A. Schweitzer<sup>4</sup>, J. A. Caballero<sup>7</sup>,  
A. González-Marcos<sup>8</sup>, I. Ribas<sup>9,10</sup>, A. Reiners<sup>11</sup>, A. Quirrenbach<sup>12</sup>, P. J. Amado<sup>13</sup>, V. J. S. Béjar<sup>2,3</sup>,  
C. Cifuentes<sup>7</sup>, Th. Henning<sup>14</sup>, A. Kaminski<sup>12</sup>, R. Luque<sup>13,15</sup>, D. Montes<sup>16</sup>, J. C. Morales<sup>9,10</sup>,  
S. Pedraz<sup>17</sup>, H. M. Tabernero<sup>18</sup>, and M. Zechmeister<sup>11</sup>

<sup>1</sup> Departamento de Construcción e Ingeniería de Fabricación, Universidad de Oviedo, c/ Pedro Puig Adam, Sede Departamental Oeste, Módulo 7, 1ª planta, 33203 Gijón, Spain  
e-mail: [abello@uniovi.es](mailto:abello@uniovi.es)

<sup>2</sup> Instituto de Astrofísica de Canarias, c/ Vía Láctea s/n, 38205 La Laguna, Tenerife, Spain

<sup>3</sup> Departamento de Astrofísica, Universidad de La Laguna, 38206 La Laguna, Tenerife, Spain

<sup>4</sup> Hamburger Sternwarte, Gojenbergsweg 112, 21029 Hamburg, Germany

<sup>5</sup> Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, 440 West Brooks Street, Norman, OK 73019, USA

<sup>6</sup> Departamento de Ingeniería de Organización, Administración de Empresas y Estadística, Universidad Politécnica de Madrid, c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

<sup>7</sup> Centro de Astrobiología (CSIC-INTA), ESAC, Camino bajo del castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain

<sup>8</sup> Departamento de Ingeniería Mecánica, Universidad de la Rioja, c/ San José de Calasanz 31, 26004 Logroño, La Rioja, Spain

<sup>9</sup> Institut de Ciències de l'Espai (CSIC-IEEC), Campus UAB, c/ de Can Magrans s/n, 08193 Bellaterra, Barcelona, Spain

<sup>10</sup> Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain

<sup>11</sup> Institut für Astrophysik und Geophysik, Georg-August-Universität, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

<sup>12</sup> Landessternwarte, Zentrum für Astronomie der Universität Heidelberg, Königstuhl 12, 69117 Heidelberg, Germany

<sup>13</sup> Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain

<sup>14</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

<sup>15</sup> Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA

<sup>16</sup> Departamento de Física de la Tierra y Astrofísica and IPARCOS-UCM (Instituto de Física de Partículas y del Cosmos de la UCM), Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

<sup>17</sup> Centro Astronómico Hispano en Andalucía (CAHA), Observatorio de Calar Alto, Sierra de los Filabres, 04550 Gérgal, Almería, Spain

<sup>18</sup> Centro de Astrobiología (CSIC-INTA), Carretera de Ajalvir km 4, Torrejón de Ardoz, 28850 Madrid, Spain

Received 3 May 2022 / Accepted 21 March 2023

### ABSTRACT

The large amounts of astrophysical data being provided by existing and future instrumentation require efficient and fast analysis tools. Transfer learning is a new technique promising higher accuracy in the derived data products, with information from one domain being transferred to improve the accuracy of a neural network model in another domain. In this work, we demonstrate the feasibility of applying the deep transfer learning (DTL) approach to high-resolution spectra in the framework of photospheric stellar parameter determination. To this end, we used 14 stars of the CARMENES survey sample with interferometric angular diameters to calculate the effective temperature, as well as six M dwarfs that are common proper motion companions to FGK-type primaries with known metallicity. After training a deep learning (DL) neural network model on synthetic PHOENIX-ACES spectra, we used the internal feature representations together with those 14+6 stars with independent parameter measurements as a new input for the transfer process. We compare the derived stellar parameters of a small sample of M dwarfs kept out of the training phase with results from other methods in the literature. Assuming that temperatures from bolometric luminosities and interferometric radii and metallicities from FGK+M binaries are sufficiently accurate, DTL provides a higher accuracy than our previous state-of-the-art DL method (mean absolute differences improve by 20 K for temperature and 0.2 dex for metallicity from DL to DTL when compared with reference values from interferometry and FGK+M binaries). Furthermore, the machine learning (internal) precision of DTL also improves as uncertainties are five times smaller on average. These results indicate that DTL is a robust tool for obtaining M-dwarf stellar parameters comparable to those obtained from independent estimations for well-known stars.

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

<sup>★</sup> Full Table A.1 is only available at the CDS via anonymous ftp to [cdsarc.cds.unistra.fr](https://cdsarc.cds.unistra.fr) (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/673/A105>

## 1. Introduction

The determination of photospheric stellar parameters in M dwarfs has always been challenging. M dwarfs are smaller, cooler, and fainter than Sun-like stars. Because of their faintness and their higher stellar activity, with sometimes stronger magnetic fields, stronger line blends, and the lack of true continuum, well-established photometric and spectroscopic methods are brought to their limits. In the literature, there are several methods to estimate M-dwarf photospheric parameters, such as effective temperature ( $T_{\text{eff}}$ ), surface gravity ( $\log g$ ), and metallicity ( $[M/H]$ ); for example, spectroscopic indices (see Rojas-Ayala et al. 2012; Gaidos & Mann 2014), photometric relations (see Dittmann et al. 2016; Houdebine et al. 2019), interferometry (see Boyajian et al. 2012; von Braun et al. 2014; Rabus et al. 2019), synthetic model fits (see Gaidos et al. 2014; Passegger et al. 2018; Marfil et al. 2021), and machine learning (ML; see Antoniadis-Karnavas et al. 2020; Passegger et al. 2020).

One method considered to be relatively precise is calibration with M dwarfs that have a late F, G, or early K common proper motion companion with known metallicity. Many of the relations mentioned below were calibrated using FGK+M multiple systems (e.g., Newton et al. 2014). As a representative example, Mann et al. (2013a) identified spectral features sensitive to metallicity in low-resolution optical and near-infrared (NIR) spectra of 112 late-K to mid-M dwarfs in multiple systems with earlier companions, from which they derived different metallicity calibrations. The same relations were used by Rodríguez Martínez et al. (2019) to determine metallicity from mid-resolution *K*-band spectra for 35 M dwarfs of the K2 mission. Other photometric calibrations using FGK+M binary systems were presented by Bonfils et al. (2005), Casagrande et al. (2008), Johnson & Apps (2009), Schlafman & Laughlin (2010), and Neves et al. (2012), among others, while several spectroscopic calibrations were explored by Rojas-Ayala et al. (2010), Dhital et al. (2012), Terrien et al. (2012), Mann et al. (2014, 2015), and, more recently, Montes et al. (2018).

Fundamental stellar parameters can also be derived from interferometric measurements. However, only a limited number of late-type dwarfs are accessible for such observations because they must be bright and nearby. Boyajian et al. (2012) presented interferometric angular diameters for 26 K and M dwarfs measured with the CHARA array and for 7 K and M dwarfs from the literature. With parallaxes and bolometric fluxes, these authors computed the absolute luminosity ( $L$ ), radii ( $R$ ), and  $T_{\text{eff}}$ . They also calculated empirical relations for K0 to M4 dwarfs to connect  $T_{\text{eff}}$ ,  $R$ , and  $L$  to a broadband color index and iron abundance  $[Fe/H]$ . On the other hand, Maldonado et al. (2015) estimated  $T_{\text{eff}}$  from pseudo-equivalent widths (pEWs) of temperature-sensitive lines calibrated with interferometric  $T_{\text{eff}}$  from Boyajian et al. (2012) and metallicities from pEWs calibrated with the relations of Neves et al. (2012). Maldonado et al. (2015) constructed a mass–radius relation using interferometric radii (Boyajian et al. 2012; von Braun et al. 2014) and masses from eclipsing binaries (Hartman et al. 2015). From this, they calculated  $\log g$ . Other studies that derived M-dwarf  $T_{\text{eff}}$  from angular diameters include, for example, those of Ségransan et al. (2003), Demory et al. (2009), von Braun et al. (2014), and Newton et al. (2015). Of these, Ségransan et al. (2003) also determined  $\log g$  from their measured masses and radii.

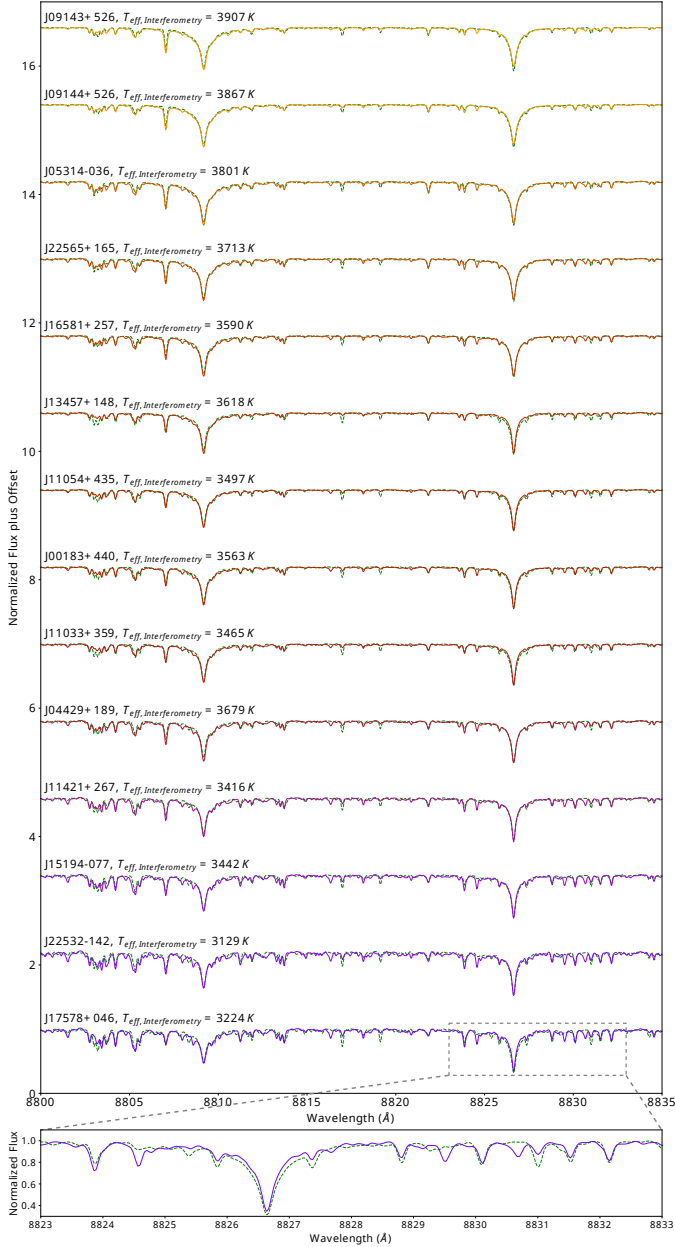
Different approaches have been taken to estimate photospheric stellar parameters for M dwarfs in general, mainly within the paradigm of comparing measured line fluxes with theoretical ones calculated from different sets of synthetic spectra.

Although different algorithms using  $\chi^2$ -minimization or principal component analysis have been employed, several artificial intelligence techniques have also been proposed. For these, the differences between observation and theory are not at the level of individual lines, but are based on whole spectral regions (Fabbro et al. 2018; KIELTY et al. 2018; BIALEK et al. 2020; MINGLEI et al. 2020; PASSEGGER et al. 2020). Indeed, some comparisons of techniques regarding the estimation of stellar parameters have already been carried out (Passegger et al. 2022). However, there are still several open questions related to the uncertainty of parameter estimation due to the signal-to-noise ratio (S/N) of the flux signal, and due to the “synthetic gap” (Fabbro et al. 2018; Tabernero et al. 2022), which is the difference in feature distribution between theoretical and observed spectra. The impact of the synthetic gap can be appreciated in Figs. 1 and 2, where we display a specific flux window showing one Mg I line at 8809 Å and one Fe I line at 8827 Å in the red optical for two sets of M dwarfs. Even the very latest synthetic models show slight differences with respect to high-S/N, high-resolution CARMENES<sup>1</sup> spectra, especially with faint lines for which parameters are not yet well constrained. In Figs. 1 and 2, the sequence of spectra is ordered according to  $T_{\text{eff}}$  and metallicity from Schweitzer et al. (2019), which are not identical to the parameters estimated from interferometry or binary companions, as shown in the figures. This mismatch is an effect caused by the synthetic gap. Another effect can be seen by the flux differences of observed and synthetic spectra in the bottom panels of Figs. 1 and 2 (zoomed-in spectra).

O’Brian et al. (2020) presented an interesting demonstration of the spectra transfer process, although their focus was rather different than ours. These authors showed that transferred spectra can reduce the synthetic gap from the pure physical models, which is further evidence of the value of transfer technologies.

The spatial dimension of features (i.e., the number of flux points within the wavelength window) depends on the size of the flux range, but it is usually very high (e.g., 3500 dimensions in the case of Figs. 1 and 2). Therefore, some specific techniques are needed to project a spectrum from such high-dimensional space into a lower one, while preserving inter-distances that help to better understand the topology. To this end, Passegger et al. (2020) introduced a technique to visualize the relative positions of a set of spectra in the 2D Euclidean space, the uniform manifold approximation and projection (UMAP; McInnes et al. 2018). The main purpose of these projections is to illustrate the difference in feature distribution between synthetic and observed spectra, that is, the so-called synthetic gap. In the left panel of Fig. 3, different theoretical PHOENIX spectra are projected along with high-S/N, high-resolution, telluric-subtracted spectra observed with CARMENES. To make the synthetic spectra comparable to the observed ones, before plotting we included continuum normalization and instrumental and rotational broadening. However, no noise was added, as it was shown in Fig. 4 of Passegger et al. (2020) that adding only noise has a negligible effect on the projection. In this representation, no stellar parameters are involved and the UMAP only depends on the flux values of every spectrum. The theoretical feature map only partially covers the CARMENES range (green circles), with a significant part of the spectra projected far away from them. Some patterns emerge when additional information, such as a color code for  $T_{\text{eff}}$ , is incorporated into the UMAP, as illustrated in both panels

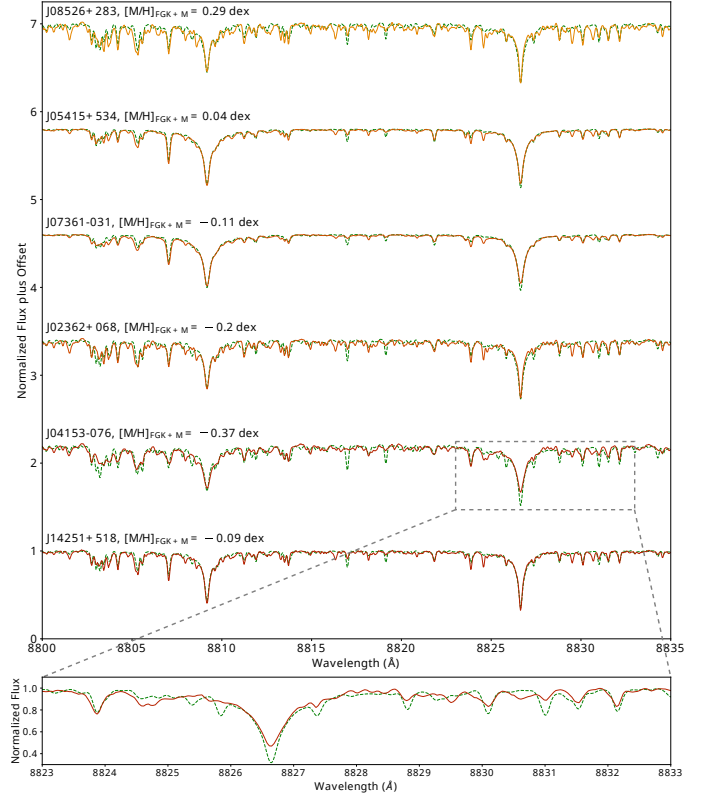
<sup>1</sup> Calar Alto high-Resolution search for M dwarfs with Exo-earths with Near-infrared and optical Échelle Spectrographs, <https://carmenes.caha.es>



**Fig. 1.** Observed vs. synthetic fluxes (8800–8835 Å) for stars with interferometric  $T_{\text{eff}}$ . Solid rainbow: CARMENES normalized template spectra. Dashed green: PHOENIX normalized synthetic spectra. Stars are sorted by decreasing  $T_{\text{eff}}$  from Schweitzer et al. (2019), with interferometrically derived  $T_{\text{eff}}$  indicated above them. Bottom panel: zoom-in detail of one representative spectrum.

of Fig. 3, independent of the source of  $T_{\text{eff}}$ : theoretical (left) or interferometric (right).

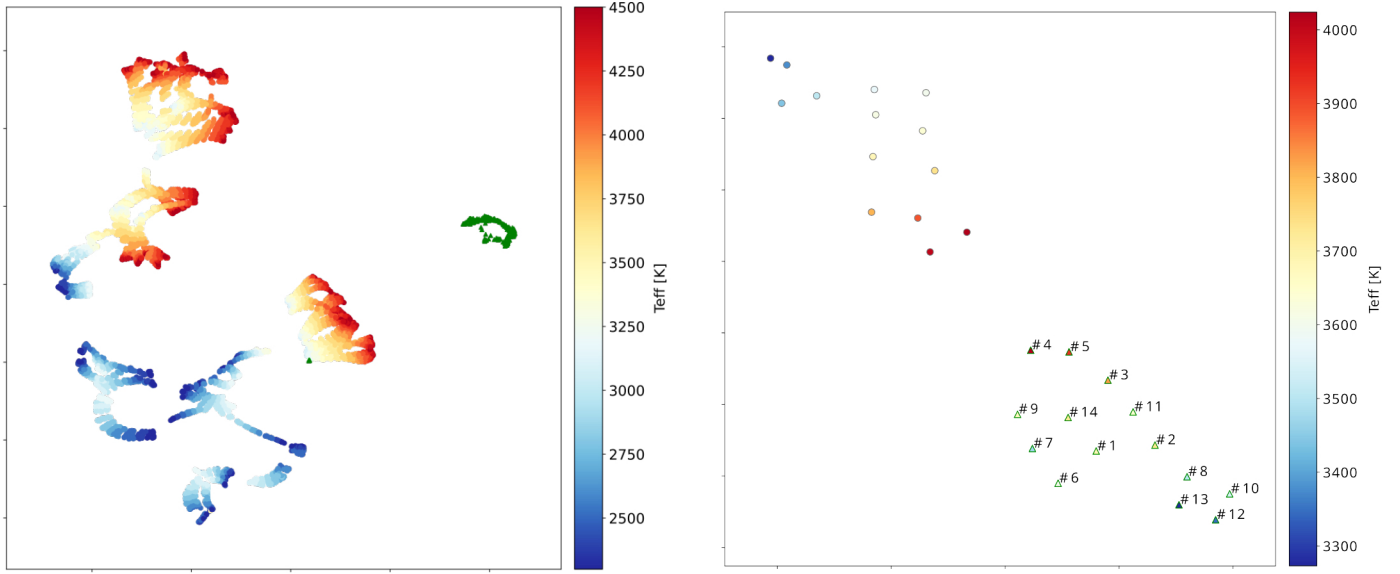
In this work, to reduce the uncertainties associated with the synthetic gap and therefore enable a more reliable estimation of stellar parameters, we propose a way to bridge the synthetic gap and transfer the knowledge from measured flux signals estimated by interferometry and FGK+M systems to the features derived from the theoretical models used with deep learning (DL). Such an approach is known as deep transfer learning (DTL; Tan et al. 2018a; Awang Iskandar et al. 2020; Wei et al. 2020). For  $T_{\text{eff}}$ , we transfer knowledge gained from interferometrically determined  $T_{\text{eff}}$  for a few stars to the rest of the CARMENES spectra, while



**Fig. 2.** Same as Fig. 1 but for M-dwarf common proper motion companions to FGK stars with well-determined metallicity, and sorted by decreasing metallicity from Schweitzer et al. (2019).

for metallicity we transfer knowledge gained from spectral synthesis of FGK stars. However, this DTL technique requires a significant amount of data, which is problematic because of the limited number of high-resolution spectra for stars fulfilling those conditions. This is even worse when data-based modeling techniques are used, as they require a methodology to assess the quality of the created model when applied to stars not used during the training phase. Despite these limitations, we show that the proposed technique is valid and its accuracy will increase as more stars with independent estimates of their parameters are incorporated.

In this paper, we use DTL to determine new  $T_{\text{eff}}$  and  $[M/H]$  for 286 M dwarfs from the CARMENES survey (Reiners et al. 2018; Quirrenbach et al. 2020), and compare our results with the literature. As our technique is based on our previous work on DL, we refer to Passegger et al. (2020) for further information. The basic workflow of the DTL can be summarized as follows: (1) train DL models on a large set of synthetic model spectra, (2) extract the internal feature representations (3) train DTL models based on the external knowledge about stellar parameters that was transferred to the neural network, (4) calculate stellar parameter estimations for the stars. In Sect. 2 we explain the DTL procedure and our artificial neural network (ANN) architecture. Section 3 describes the values obtained from the literature (interferometry and FGK+M systems) for each parameter that we used for training the ANN, the stellar sample, and the application of our ANN. The derived stellar parameters are presented in Sect. 4, together with a literature comparison and discussion. Finally, in Sect. 5 we provide a short summary.



**Fig. 3.** Representative two-dimensional UMAP projections of observed and synthetic spectra from the 8800–8835 Å window, with  $T_{\text{eff}}$  color coded. Left panel: PHOENIX-ACES set used for DL training and the 282 CARMENES spectra (green). Right panel: subsample of 14 CARMENES M dwarfs with interferometric  $T_{\text{eff}}$  values shown in Table 1 (green triangles and labeled). Colored circles represent their closest interpolated best-fit PHOENIX model, using Schweitzer et al. (2019) parameter estimations as a reference.

## 2. Methods

The aim of ML is to automatically discover rules that must be followed in order to efficiently map input data to a desired output. In this process, it is essential to create appropriate representations of the data. These representations are task-dependent and may vary according to the final task that the selected ML algorithm is going to perform. DL is a subfield of ML where a hierarchical representation of the data is created, and has received increasing attention in recent years in light of its successful application to numerous real-world problems (e.g., virtual assistants, visual recognition, fraud detection, machine translation, medical image analysis, photo descriptions; see Karpathy & Fei-Fei 2015 and many others). The higher levels of the hierarchy are formed by the composition of representations of the lower level (Passegger et al. 2020). More importantly, this hierarchy of representations is automatically learned from the data by completely automating the most crucial step in ML, namely feature engineering. Automatically learning features at multiple levels of abstraction allows a system to learn complex representations mapping the input to the output directly from the data, without completely depending on human-crafted features. The word “deep” refers to the multiple hidden layers used to obtain those representations. In this sense, DL can also be called hierarchical feature engineering (Sarkar et al. 2018).

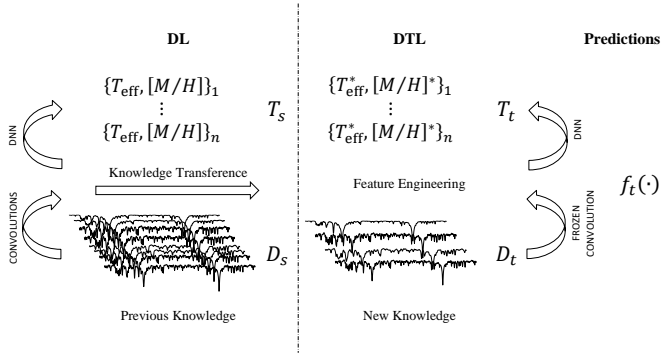
Data dependence is one of the most serious issues in DL, which is extremely dependent on massive training data sets when compared to traditional ML methods. Although the amount of data needed depends on the type of model, the required accuracy, and the complexity of the model, all these factors can lead to the requirement for large datasets. Therefore, an intrinsic and unavoidable problem has always been insufficient training data. Data collection is complex and expensive, making the generation of large-scale, high-quality annotated data sets extremely difficult. Therefore, techniques to work with data sets of limited size are of great value, as is the case for the DTL technique.

### 2.1. Deep transfer learning

It has become increasingly common in various domains, such as image recognition and natural language processing, to pre-train the entire model in a data-rich task (Kraus & Feuerriegel 2017; Gao & Mosalam 2018; Raffel et al. 2020; Han et al. 2021). Ideally, this pre-training process causes the model to develop general-purpose abilities and knowledge that can then be transferred to downstream tasks. Goodfellow et al. (2016) referred to transferred learning (TL) in the context of generalization. These latter authors defined TL as the situation where what has been learned in one setting is exploited to improve generalization in another setting. Therefore, TL provides a robust and practical solution to leverage information from one domain to improve the accuracy of a model built for a different domain (Vilalta 2018).

Pan & Yang (2010) proposed a more precise definition of TL, starting by defining a domain and a task, respectively. A domain can be represented by  $D = \chi, P(X)$ , which contains two parts: the feature space  $\chi$  and the marginal probability distribution  $P(X)$ , where  $X = \{x_1, \dots, x_n\} \in \chi$ . The task can be represented by  $T = \{y, f(x)\}$ , and consists of two parts: a label space  $y$  and a target-prediction function  $f(x)$ . This function  $f(x)$  can also be regarded as a conditional probability function  $P(y|x)$ . Then, given a learning task  $T_t$  based on  $D_t$  (where the subscript  $t$  refers to “transferred”), TL is designed to improve the performance of a predictive function  $f_T(\cdot)$  in learning the task  $T_t$  by discovering and transferring latent knowledge from another domain  $D_s$  and learning task  $T_s$  (where the subscript  $s$  refers to “source”, which in our case is the PHOENIX-ACES synthetic models), where  $D_s \neq D_t$  and/or  $T_s \neq T_t$ . Usually, the size of the source domain  $D_s$  is much larger than the size of the transferred domain  $D_t$  (i.e.,  $N_s \gg N_t$ ). Based on the previous definitions, a DTL task is defined by  $\langle D_s, T_s, D_t, T_t, f_T(\cdot) \rangle$ , where  $f_T(\cdot)$  is a nonlinear function involving a deep ANN. TL relaxes the hypothesis that the training data must be independent





**Fig. 4.** Representation of the domains and tasks that the DTL process can be applied to.

of and identically distributed with the test data, which motivates the use of TL for the problem of insufficient training data.

The popularity of DL has led to many different DTL methods, and several authors have proposed a classification of them (Tan et al. 2018b; Zhao et al. 2021). Common categories involve instance-, mapping-, network-, and adversarial-based TL. Each of these categories has its particular applicability, considering the specific context and characteristics of domains and tasks  $\langle D_s, T_s, D_t, T_t, f_T(\cdot) \rangle$ . In our case, due to the characteristics of the problem, we selected the network category to implement our DTL. The relationship between domains and tasks is illustrated by Fig. 4.

## 2.2. Artificial neural network architectures

Network-based DTL refers to reusing the part of the pre-trained network in the source domain, including its network structure and connection parameters, and transferring it to be a part of the deep neural network that is used in the target domain. The main assumption is connected with the idea that the features identified in the source domain will be valid in the transfer domain, whereas  $f_T(\cdot)$  requires adaptation. As a result, we kept the original distribution of fully connected layers that we applied for the DL analysis by Passegger et al. (2020). Moreover, Passegger et al. (2020) constructed several neural network models for different spectral regions, finding that the results for all regions are comparable, but that the region 8800–8835 Å gives the smallest validation error. Therefore, we also adopted this strategy and only use that region. In this sense, we rely on the DL models already trained in Passegger et al. (2020) and extract the internal feature representations for each star in order to use it here as a new input for the DTL process. The DL models were trained on a grid of 449 806 synthetic PHOENIX-ACES spectra, after removing unphysical stellar parameter combinations not corresponding to main sequence stars using the PARSEC v1.2S evolutionary models (Bressan et al. 2012; Chen et al. 2014, 2015; Tang et al. 2014). The feature representation was taken from the flattened layer of the DL model, as represented in Fig. 5.

## 2.3. DTL training and testing

To build the transfer domain  $D_t$  for  $T_{\text{eff}}$ , we started with 14 stars of the CARMENES survey with interferometric angular diameters  $\theta_{\text{LD}}$  measured by Boyajian et al. (2012), von Braun et al. (2014), and references therein. We did not use the derived  $T_{\text{eff}}$  from these publications. Instead, we used updated bolometric

fluxes  $S$  at Earth based on the most recent photometry (in particular, *Gaia* photometry for the optical passbands) collected by Cifuentes et al. (2020) to calculate our reference  $T_{\text{eff}}$  with the distance-independent form of the Stefan-Boltzmann law:

$$T_{\text{eff}} = \theta_{\text{LD}}^{-1/2} \left( \frac{4S}{\sigma} \right)^{1/4}. \quad (1)$$

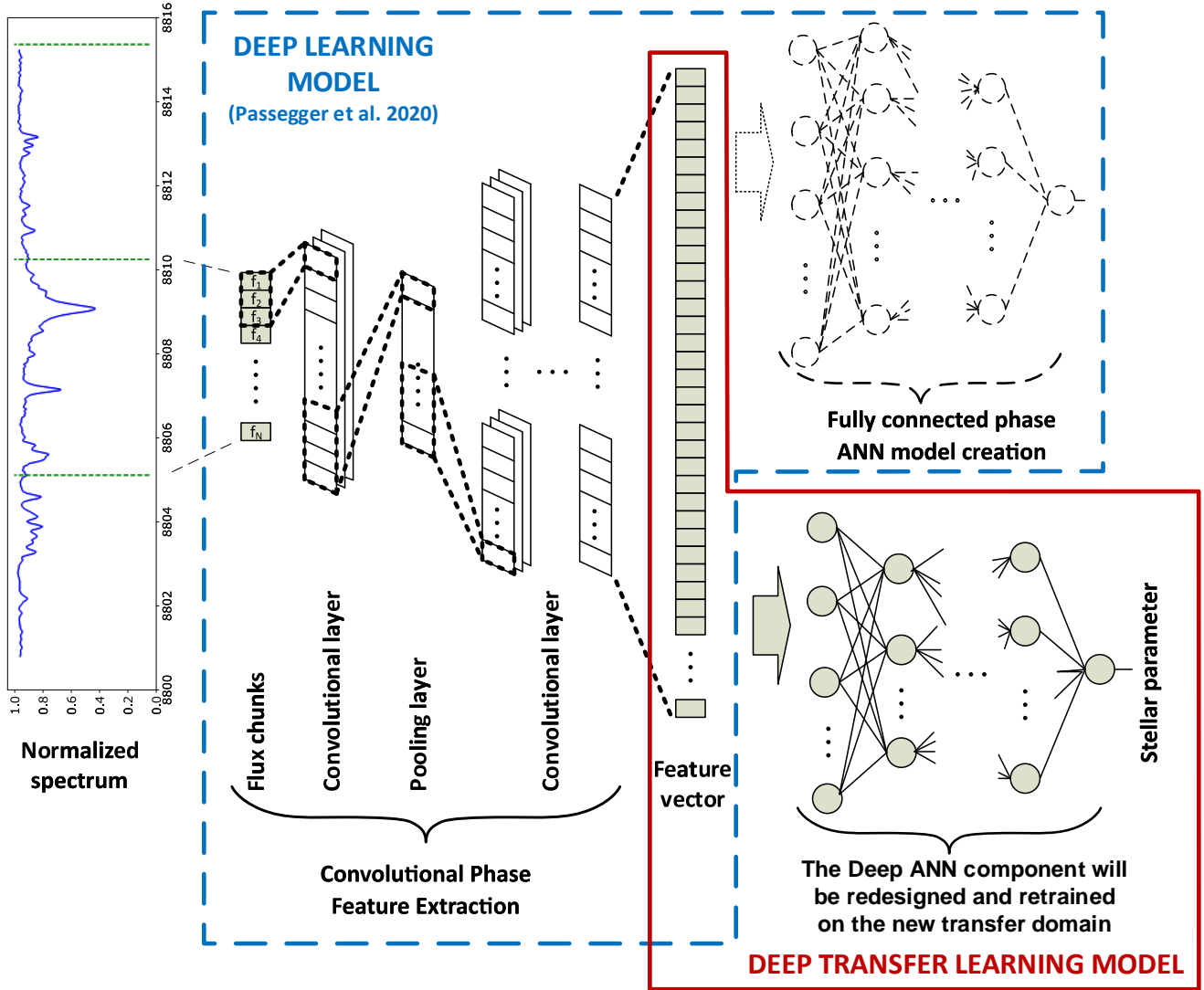
As Cifuentes et al. (2020) did not actually tabulate the bolometric flux at Earth  $S$ , we used the tabulated bolometric luminosities  $L$  and the distances  $d$ , which were used in measuring  $L$ , and calculated  $S$  via  $S = L/(4\pi d^2)$ . All used and derived values ( $L$ ,  $d$ ,  $S$ ,  $\theta_{\text{LD}}$ , and  $T_{\text{eff}}$ ) are listed in Table 1.

As for G1 15A our derived  $T_{\text{eff}}$  from Eq. (1) was in severe disagreement with  $T_{\text{eff}}$  listed in Boyajian et al. (2012), we realized that for this bright star, as well as for the bright star G1 411, the luminosities in Cifuentes et al. (2020) were missing reliable  $J$  band photometry. Therefore, we revise the luminosity determinations of Cifuentes et al. (2020) by adding classical Johnson  $J$  band photometry as well as the *Gaia* Early Data Release 3 (EDR3) data. G1 205 was not included in Cifuentes et al. (2020), but its  $L$  is derived in the same fashion. These three stars are marked in Table 1 with ‘This work’.

As the 14 stars of the interferometric sample include only two M dwarfs with  $T_{\text{eff}} < 3440$  K, we supplemented them with five mid-to-late M dwarfs listed in Table 2, for which a good  $T_{\text{eff}}$  estimation is available in the literature, as recommended by Passegger et al. (2022). This was done in order to obtain training and validation sets with regularly spaced parameters. To achieve such regular distribution, we binned the to-be-transferred data set of 19 M dwarfs into a variable number of bins. The goal was to have as many as 75% nonempty bins. For each of those bins, we selected a representative element. If two or more stars were included in one bin, we picked the closest to the midpoint of the bin.

For the metallicity,  $[M/H]$ , the adopted strategy followed the same structure as for  $T_{\text{eff}}$ . We reverted to metallicities measured for FGK stars with a proper motion M-dwarf companion. Due to their formation from the same cloud, it is assumed that both components share the same metallicity (Desidera et al. 2006; Andrews et al. 2018). Table 3 presents five M dwarfs with CARMENES spectra with an FGK primary with known metallicity obtained from Montes et al. (2018), and one from Tabernero et al. (2022), namely J14251+518 ( $\theta$  Boo B). Due to the low number of multiple systems, our transfer strategy is to use the list of 18 stars from Passegger et al. (2022) in Table 4 combined with those listed in Table 3 as  $D_t$ . Although the values from Passegger et al. (2022) are not as accurate as those from binaries, they do not depend strongly on a specific model because they were calculated as medians from several literature values.

To avoid the potential lack of generalization linked to accepting models based solely on their performance over the validation set, we propose using a more robust methodology, which is designed to create two different and separate groups of samples: one set for training and validation, and the rest of the stars as a test group to measure the quality of the models. Indeed, because of the relatively significant influence that a single sample can have on the model performance, which is due to the low number of samples in the training or validation subsets, we used the cross-validation training approach. Cross-validation is a data-resampling method to assess the generalization ability of predictive models and to prevent overfitting (Refaeilzadeh et al. 2009). Briefly, the data are usually divided into two segments: one used for training a neural network model and one used for



**Fig. 5.** Representation of the adopted strategy for the network-based approach, where the ANN component has been redesigned with respect to the DL model derived from PHOENIX-ACES spectra (to be compared with Fig. 1 of [Passegger et al. 2020](#)).

**Table 1.** Interferometrically derived  $T_{\text{eff}}$  values transferred.

| #  | Karman     | Gliese | Name           | $L$<br>[ $10^{-4}L_{\odot}$ ] | Ref. $L$  | $d$<br>[pc]           | Ref. $d$ | $S^{(a)}$<br>[ $10^{-11}\text{W m}^{-2}$ ] | $\theta_{\text{LD}}$<br>[mas] | Ref. $\theta_{\text{LD}}$ | $T_{\text{eff}}$<br>[K] |
|----|------------|--------|----------------|-------------------------------|-----------|-----------------------|----------|--|-------------------------------|---------------------------|-------------------------|
| 1  | J00183+440 | 15A    | GX And         | $239.1 \pm 9.2$               | This work | $3.56244 \pm 0.00026$ | EDR3     | $6.03 \pm 0.23$                            | $1.005 \pm 0.005$             | Boy12                     | $3658 \pm 36$           |
| 2  | J04429+189 | 176    | HD 285968      | $358.1 \pm 8.9$               | Cif20     | $9.4730 \pm 0.0063$   | DR2      | $1.28 \pm 0.03$                            | $0.448 \pm 0.021$             | vBr14                     | $3717 \pm 90$           |
| 3  | J05314-036 | 205    | HD 36395       | $657.0 \pm 3.9$               | This work | $5.70408 \pm 0.00066$ | EDR3     | $6.46 \pm 0.04$                            | $0.943 \pm 0.004$             | Boy12                     | $3843 \pm 10$           |
| 4  | J09143+526 | 338A   | HD 79210       | $811 \pm 30$                  | Cif20     | $6.3339 \pm 0.0015$   | DR2      | $6.47 \pm 0.24$                            | $0.834 \pm 0.014$             | Boy12                     | $4087 \pm 51$           |
| 5  | J09144+526 | 338B   | HD 79211       | $759 \pm 15$                  | Cif20     | $6.3337 \pm 0.0017$   | DR2      | $6.05 \pm 0.12$                            | $0.856 \pm 0.016$             | Boy12                     | $3968 \pm 42$           |
| 6  | J11033+359 | 411    | Lalande 21185  | $225.2 \pm 8.9$               | This work | $2.54613 \pm 0.00021$ | EDR3     | $11.11 \pm 0.44$                           | $1.432 \pm 0.013$             | Boy12                     | $3571 \pm 39$           |
| 7  | J11054+435 | 412A   | BD+44 2051A    | $193.4 \pm 4.3$               | Cif20     | $4.848 \pm 0.024$     | HIP2     | $2.63 \pm 0.06$                            | $0.764 \pm 0.017$             | Boy12                     | $3411 \pm 43$           |
| 8  | J11421+267 | 436    | Ross 905       | $243.1 \pm 3.4$               | Cif20     | $9.7560 \pm 0.0089$   | DR2      | $0.82 \pm 0.01$                            | $0.417 \pm 0.013$             | vBr12                     | $3446 \pm 55$           |
| 9  | J13457+148 | 526    | HD 119850      | $377 \pm 13$                  | Cif20     | $5.4353 \pm 0.0015$   | DR2      | $4.09 \pm 0.14$                            | $0.835 \pm 0.014$             | Boy12                     | $3642 \pm 43$           |
| 10 | J15194-077 | 581    | HO Lib         | $123.5 \pm 2.6$               | Cif20     | $6.2992 \pm 0.0021$   | DR2      | $1.00 \pm 0.02$                            | $0.446 \pm 0.014$             | vBr11                     | $3501 \pm 58$           |
| 11 | J16581+257 | 649    | BD+25 3173     | $447 \pm 20$                  | Cif20     | $10.3827 \pm 0.0034$  | DR2      | $1.33 \pm 0.06$                            | $0.484 \pm 0.012$             | vBr14                     | $3610 \pm 61$           |
| 12 | J17578+046 | 699    | Barnard's star | $35.23 \pm 0.91$              | Cif20     | $1.82665 \pm 0.00097$ | DR2      | $3.38 \pm 0.09$                            | $0.952 \pm 0.005$             | Boy12                     | $3252 \pm 23$           |
| 13 | J22532-142 | 876    | IL Aqr         | $127.1 \pm 2.5$               | Cif20     | $4.6758 \pm 0.0017$   | DR2      | $1.86 \pm 0.04$                            | $0.746 \pm 0.009$             | vBr14                     | $3165 \pm 25$           |
| 14 | J22565+165 | 880    | HD 216899      | $516.8 \pm 8.9$               | Cif20     | $6.8676 \pm 0.0018$   | DR2      | $3.51 \pm 0.06$                            | $0.744 \pm 0.004$             | Boy12                     | $3713 \pm 19$           |

**Notes.** <sup>(a)</sup>  $S$  is the flux at Earth calculated from the tabulated values  $L$  and  $d$ .  $T_{\text{eff}}$  is calculated using Eq. (1).

**References.** HIP2: [van Leeuwen \(2007\)](#); vBr11: [von Braun et al. \(2011\)](#); Boy12: [Boyajian et al. \(2012\)](#); vBr12: [von Braun et al. \(2012\)](#); vBr14: [von Braun et al. \(2014\)](#); DR2: [Gaia Collaboration \(2018\)](#); Cif20: [Cifuentes et al. \(2020\)](#); EDR3: [Gaia Collaboration \(2021\)](#).

**Table 2.** Supplementing literature  $T_{\text{eff}}$  values transferred from [Passegger et al. \(2022\)](#).

| Karmn      | Gliese | Name     | $T_{\text{eff}}$ [K] |
|------------|--------|----------|----------------------|
| J00067–075 | 1002   | G 158-27 | $2844 \pm 149$       |
| J07558+883 | 1101   | G 253-6  | $3217 \pm 68$        |
| J10508+068 | 402    | Wolf 358 | $3169 \pm 93$        |
| J13005+056 | 493.1  | Wolf 461 | $3090 \pm 107$       |
| J23419+441 | 905    | HH And   | $3058 \pm 88$        |

validating the trained model. The basic form of cross-validation is a  $k$ -fold cross-validation, where the data are divided into  $k$  folds (in our case, we set  $k = 4$ ) before  $k$  iterations of training and validation are conducted, where in each iteration a different fold is kept aside for validation, while the remaining  $k - 1$  folds are used for training.

DTL model creation requires a quality criterion to assess the learning progress of the ANN. The quality criterion that is often adopted is a threshold in the loss error during the validation process (the loss function is widely used in mathematical optimization and decision theory). Looking to obtain a sufficient variety of models due to randomness in the selection of samples and the optimization starting point, several repetitions of the model-creation process were accomplished. As a four-fold cross-validation strategy was adopted, four potentially valid models were created per repetition of the model-creation process, and another four validation loss errors were measured at the end of the training processes. The trained models were deemed of sufficient quality when the validation error was lower than 0.01. In order to have a significant set of models in the case of metallicity, 80 repetitions of the model-creation process were adopted, which means  $80 \times 4 = 320$  potential models. Only 121 of these reached convergence under the adopted criterion, and they were later used for predictions. In the case of  $T_{\text{eff}}$ , 20 repetitions of the model-creation process were adopted and all the  $20 \times 4 = 80$  potential models reached convergence. This behavior indirectly shows that, as expected, the  $T_{\text{eff}}$  parameter has more power than metallicity in terms of impact on the spectra.

It is straightforward to build a model that is perfectly adapted to the data set at hand but then unable to generalize to new and unseen data. Therefore, the value of measuring model quality over an independent data set becomes evident. In this way, the model performance can be externally assessed using the estimation provided by the information gathered during the validation step ([Vabalas et al. 2019](#)).

### 3. Analysis

#### 3.1. Observational sample

To test our DTL method, we used the same template spectra as in [Passegger et al. \(2019\)](#) and applied it to all 282 M dwarfs listed in their Table B.1 plus four more stars coming from the independent interferometric sample used for the learning process. We focus here on a small sample of the galactic stellar population, namely M dwarfs of spectral type M0 to M6. To verify whether or not our method is able to generalize beyond this parameter range, it is necessary to apply it to a much larger stellar sample, such as APOGEE or *Gaia*, which shall be part of a subsequent study. The stars were observed with CARMENES on the Zeiss 3.5 m telescope at the Observatorio de Calar Alto,

Spain. CARMENES combines two highly stable fiber-fed spectrographs covering a spectral range from 520 to 960 nm in the optical (VIS) and from 960 to 1710 nm in the NIR, with spectral resolutions of  $R \approx 94\,600$  and  $80\,400$ , respectively ([Quirrenbach et al. 2018](#); [Reiners et al. 2018](#)). The primary goal of this instrument is to search for Earth-sized planets in the habitable zones of M dwarfs (e.g., [Zechmeister et al. 2019](#)).

For a detailed description of our data-reduction procedure, we refer to [Zechmeister et al. \(2014\)](#), [Caballero et al. \(2016\)](#), and [Passegger et al. \(2019\)](#). As in the latter, we used the high-S/N template (co-added) spectrum for each star. These templates are a byproduct of the CARMENES radial-velocity pipeline *serval* (SpEctrum Radial Velocity AnaLyser; [Zechmeister et al. 2018](#)). In the standard data flow, the code constructs a template for every target star from at least five individual spectra to derive the radial velocities of a single spectrum by least-square fitting to the template. For our sample, the average S/N of the order, in which our investigated wavelength window of 8800–8835 Å lies in the beginning of the order, amounts to  $258 \pm 158$ .

Before creating the templates, the NIR spectra were corrected for telluric lines. We did not use the telluric correction for the VIS spectra because the telluric features are negligible in the investigated range. The telluric correction was explained in detail by [Nagel et al. \(2020\)](#).

For normalization of our spectra, we used the same method and routine as in [Passegger et al. \(2020\)](#), the Gaussian Inflection Spline Interpolation Continuum (GISIC<sup>2</sup>, developed by D.D. Whitten and designed for spectra with strong molecular features). After the spectrum was smoothed with a Gaussian, and continuum points were selected, the pseudo-continuum was normalized with a cubic spline interpolation. We applied the same procedure to both observed and synthetic spectra within the spectral window 8800–8835 Å by adding 5 Å on each side to avoid possible edge effects. The observed spectra have been corrected for radial velocity to match the rest frame of the synthetic spectra using the cross-correlation (crosscorrRV from PyAstronomy, [Czesla et al. 2019](#)) between a PHOENIX model spectrum and the observed spectrum. To obtain a universal wavelength grid, which is necessary for applying the DL method, the wavelength grid of the observed spectra was linearly interpolated with the grid of the synthetic spectra.

#### 3.2. Transferred knowledge

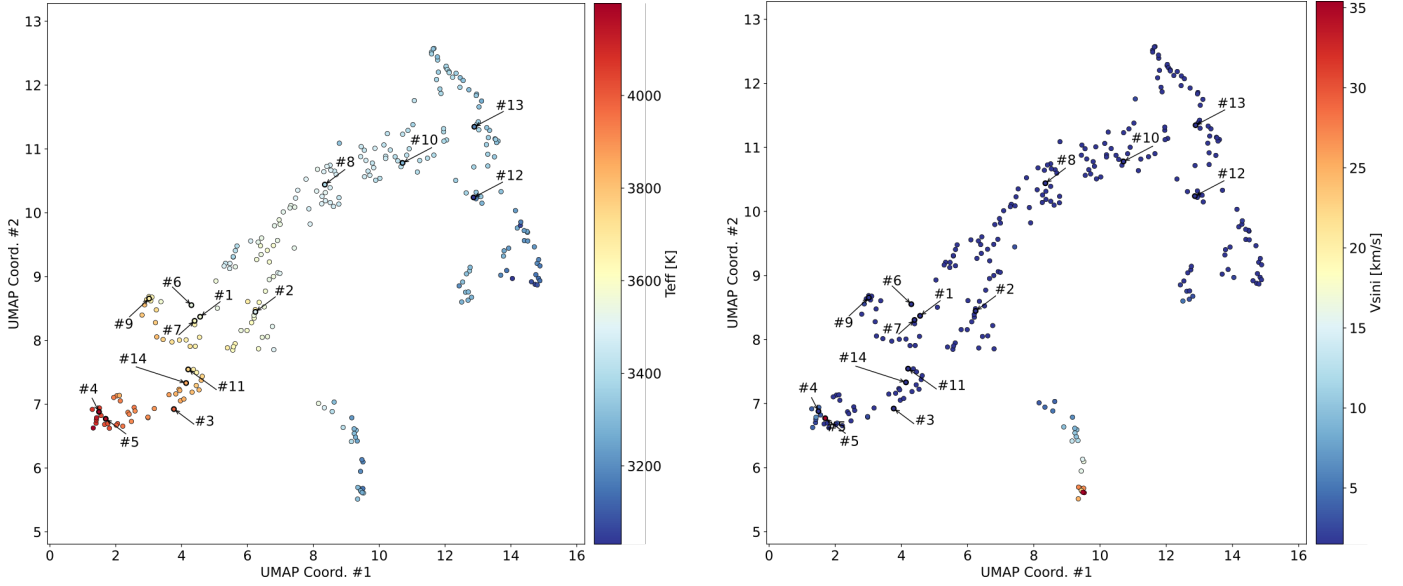
In our particular case, where the distance between domains is significant (see Fig. 3) and the sample density in  $D_t$  is limited (see Tables 1 and 3), the network-based approach was selected as the adequate DTL method. In Fig. 6, the stars with interferometric  $T_{\text{eff}}$  are regularly distributed in  $T_{\text{eff}}$  along the whole CARMENES data set. This is needed and must be checked before transferring any knowledge from another study, as DL and DTL techniques are not very good at extrapolating information because of the limited stellar parameter range of our transferred and training sets.

In terms of the terminology introduced in Sect. 2.1, our  $D_s$  domain was built over the PHOENIX-ACES spectra library ([Husser et al. 2013](#)) with a flux window of 35 Å between 8800 and 8835 Å in the VIS channel for consistency with previous work ([Passegger et al. 2020, 2022](#)). Furthermore,  $T_s$  is the DL model that minimizes the error on a test set of unused PHOENIX-ACES spectra. In other words, the DL model selected for transferring the feature space is the best one of those trained

<sup>2</sup> <https://pypi.org/project/GISIC/>

**Table 3.** Transferred spectroscopically determined  $[M/H]$  values from FGK+M systems.

| Karmn      | Gliese | Name              | $[M/H]$<br>[dex] | FGK primary       |          | Reference               |
|------------|--------|-------------------|------------------|-------------------|----------|-------------------------|
|            |        |                   |                  | Name              | Sp. type |                         |
| J02362+068 | 105B   | BX Cet            | $-0.20 \pm 0.02$ | HD 16160 A        | K0 V+    | Montes et al. (2018)    |
| J04153-076 | 166C   | $\rho^{02}$ Eri C | $-0.37 \pm 0.02$ | $\rho^{02}$ Eri A | K0.5 V   | Montes et al. (2018)    |
| J05415+534 | 212    | HD 233153         | $+0.04 \pm 0.02$ | V538 Aur          | K1 V     | Montes et al. (2018)    |
| J07361-031 | 282C   | BD-02 2198        | $-0.11 \pm 0.03$ | V869 Mon          | K2 V     | Montes et al. (2018)    |
| J08526+283 | 324B   | $\rho^{01}$ Cnc B | $+0.29 \pm 0.04$ | $\rho^{01}$ Cnc A | G8 V     | Montes et al. (2018)    |
| J14251+518 | 549B   | $\theta$ Boo B    | $-0.09 \pm 0.01$ | $\theta$ Boo A    | F7 V     | Tabernero et al. (2022) |


**Fig. 6.** UMAP of CARMENES color-coded according to  $T_{\text{eff}}$  (left) and  $v \sin i$  (right), built with the flux values from the 8800–8835 Å window. The M dwarfs with interferometric values are labeled.  $T_{\text{eff}}$  values in the left panel are taken from Schweitzer et al. (2019) and Table 1, while rotational velocities in the right panel are taken from Passegger et al. (2019).

**Table 4.** Supplementary literature  $[M/H]$  values transferred from Passegger et al. (2022).

| Karmn      | Gliese | Name           | $[M/H]$          |
|------------|--------|----------------|------------------|
| J00067-075 | 1002   | G 158-27       | $-0.19 \pm 0.16$ |
| J00183+440 | 15A    | GX And         | $-0.26 \pm 0.09$ |
| J04429+189 | 176    | HD 285968      | $+0.08 \pm 0.10$ |
| J05314-036 | 205    | HD 36395       | $+0.39 \pm 0.11$ |
| J07558+833 | 1101   | G 253-6        | $+0.00 \pm 0.10$ |
| J09143+526 | 338A   | HD 79210       | $-0.13 \pm 0.12$ |
| J09144+526 | 338B   | HD 79211       | $-0.11 \pm 0.14$ |
| J10508+068 | 402    | EE Leo         | $+0.16 \pm 0.11$ |
| J11033+359 | 411    | Lalande 21185  | $-0.31 \pm 0.11$ |
| J11054+435 | 412A   | BD+44 2051A    | $-0.38 \pm 0.12$ |
| J11421+267 | 436    | Ross 905       | $+0.00 \pm 0.11$ |
| J13005+056 | 493.1  | FN Vir         | $+0.09 \pm 0.10$ |
| J13457+148 | 526    | HD 119850      | $-0.22 \pm 0.11$ |
| J15194-077 | 581    | HO Lib         | $-0.13 \pm 0.11$ |
| J16581+257 | 649    | BD+25 3173     | $-0.04 \pm 0.12$ |
| J17578+046 | 699    | Barnard's star | $-0.39 \pm 0.12$ |
| J22565+165 | 880    | HD 216899      | $+0.18 \pm 0.09$ |
| J23419+441 | 905    | HH And         | $+0.25 \pm 0.11$ |

on synthetic PHOENIX-ACES spectra. The definition of  $D_t$  for the two stellar parameters is introduced in Sect. 2.1.

From this, up to 80 different transference models were created. If their training process reaches convergence, they were selected as contributing to the prediction of the stellar parameters of the stars in the test set. Proposals from different transferred models are collected and integrated using the kernel density estimate (KDE) technique. This technique allows us to establish the most frequent value for the stellar parameter, but also its uncertainty, which depends on the star and the flux window. This KDE estimation can be seen as the predictive function  $f_T(\cdot)$ .

### 3.3. Implementation

As already explained in Sect. 2.1, transfer learning is an approach in DL (and ML) where knowledge is transferred from one model to another. This means that a properly trained DL model is the first step, as presented in Fig. 5. For the DTL network-oriented approach, we kept the features selected by the DL model, which means freezing the convolutional transformers and providing a new deep ANN configuration. This enabled us to train the weights of the connections for stellar parameters according to the interferometric measurements, binary companion estimations, or



**Table 5.** Convolutional architecture used in the DL model.

| Operation | Input size | Kernel size | Number of variables |
|-----------|------------|-------------|---------------------|
| Input     | 3501       | 1           | ...                 |
| Conv 1D   | 3501       | 32          | 128                 |
| MaxPool   | 1750       | ...         | ...                 |
| Conv 1D   | 1750       | 16          | 1552                |
| MaxPool   | 875        | ...         | ...                 |
| Conv 1D   | 875        | 8           | 392                 |
| MaxPool   | 437        | ...         | ...                 |
| Conv 1D   | 437        | 4           | 100                 |
| MaxPool   | 218        | ...         | ...                 |
| Flatten   | 872        | ...         | ...                 |

**Table 6.** Neural network architecture used in the DTL model.

| Operation | Input size | Variables |
|-----------|------------|-----------|
| Input     | 872        | 0         |
| Dense     | 16         | 13 698    |
| Dense     | 8          | 136       |
| Dense     | 4          | 36        |
| Dense     | 1          | 5         |

both. The architectures used for convolutional layers and the deep neural network are presented in Tables 5 and 6, respectively.

Once the configuration has been defined, the training process takes place. Due to the fact that convolutional layers are frozen, the evolution of models is due to the update of variable weights, which requires a larger number of iterations than for training the whole convolutional neural network (CNN) at once. Indeed, specific attention must be paid to the cross-validation strategy used to avoid over-fitting, which means that the same number of models as folders is required. Different numbers of folders were tested in the cross-validation process (from three to five), as well as different bases for features derived from different DL models, which required a significant number of repetitions for the training procedure before being able to propose a set of DTL models. In the present case, and because of the limited number of available samples, the best method for measuring the global quality of the whole data set was using four cross-validation folders, which were then adopted for the implementation. Therefore, the number of computing operations in the training step is expected to be large, and advanced computing capabilities are targeted to keep the effort bounded. As the operations are tensor-based computations, the use of existing frameworks can save a lot of time.

The adoption of the TensorFlow framework (Abadi et al. 2016) for the creation of DL models enables the use of accelerated hardware based on Nvidia general-purpose graphics processing unit (GPU) cards, which outperform the central processing unit (CPU) in terms of computation time by around a factor of 20 (Mittal & Vaishay 2019). As the base DL models were selected from those performed in Passegger et al. (2020) and were able to identify the best features, the same framework was retained for the current implementation of DTL. Current features were extracted from the aforementioned models, and a complete new deep ANN was configured and trained over the new set of spectra with better stellar parameter estimation. In this

**Table 7.**  $T_{\text{eff}}$  comparison for the stars not used in the training phase.

| Karmn      | $T_{\text{eff,interf}}$<br>[K] | $T_{\text{eff,DL}}$<br>[K] | $T_{\text{eff,DTL}}$<br>[K] | $T_{\text{eff,Schw19}}$<br>[K] |
|------------|--------------------------------|----------------------------|-----------------------------|--------------------------------|
| J09144+526 | $3968 \pm 42$                  | $4033 \pm 60$              | $3877 \pm 16$               | $4005 \pm 51$                  |
| J15194-077 | $3501 \pm 58$                  | $3391 \pm 47$              | $3259 \pm 24$               | $3441 \pm 51$                  |
| J16581+257 | $3610 \pm 61$                  | $3745 \pm 43$              | $3732 \pm 11$               | $3734 \pm 51$                  |
| J17578+046 | $3252 \pm 23$                  | $3352 \pm 55$              | $3212 \pm 48$               | $3273 \pm 51$                  |
| J22532-142 | $3165 \pm 25$                  | $3313 \pm 55$              | $3125 \pm 37$               | $3377 \pm 51$                  |

**Notes.** Observed ('interf'), DL-estimated ('DL'), DTL-predicted ('DTL'), and Schw19 values.

particular case, because training involved only adjustment of the deep ANN weights, several thousand epochs were required to produce and estimate the adapted function.

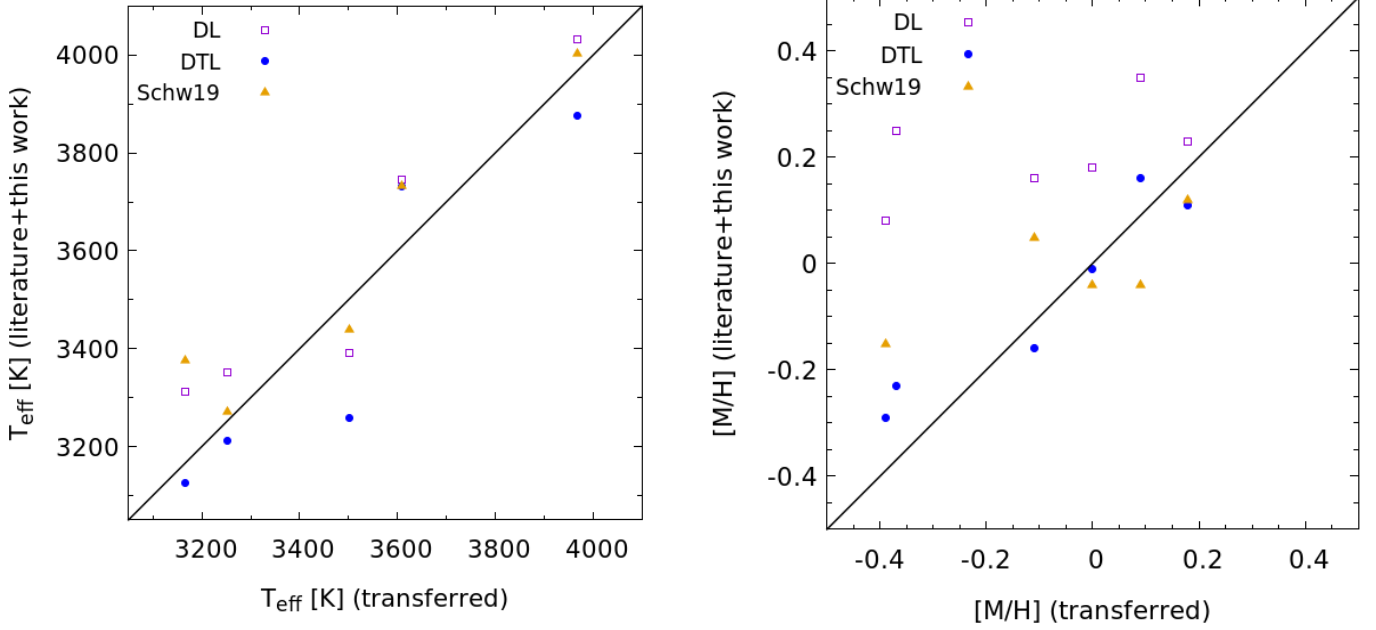
In this application, we used GPU cards with 11 GB of RAM and 4352 computing cores. The training time for a model experiencing proper convergence depended on the training data size, but also on the architecture and number of epochs, and varied between 45 minutes and two hours. For a more detailed description of the general design of a DL neural network, we refer to Passegger et al. (2020).

The same methodology used by Passegger et al. (2020) for uncertainty estimation was considered here, where parameter estimations from each DTL model were collected and the probability density function was determined using KDE (Scott 2015; Terrell & Scott 1992; Wang & Li 2017). Based on such a probability density function, the maximum was retained as a confident estimation of the parameter. This was done for each star and stellar parameter separately. To provide the uncertainty for each star and parameter, the  $\pm 1\sigma$  thresholds of the predictions were calculated.

## 4. Results and discussion

We introduced an algorithm-independent assessment of precision in the prediction of  $T_{\text{eff}}$  and [M/H]. This was carried out thanks to the testing data set, where stars not seen before by the model during its knowledge transference were used as the gold standard for its assessment for estimating stellar parameters. First, we applied the DL method presented by Passegger et al. (2020) by selecting the best DL model that predicts  $T_{\text{eff}}$  from PHOENIX-ACES. In doing so, the hypothesis that the derived models use the most relevant feature set after the convolution step remains acceptable. This first step also provided a set of comparisons of stellar parameters derived from DL.

According to the steps indicated in Sect. 3.2, starting from the best DL model, the network-based transfer procedure was carried out within the samples from Tables 1 and 2. In this procedure, 14 stars were used during the four-fold cross-validation approach, keeping aside the remaining five stars in Table 7. After creating the most suitable model, its quality was assessed by measuring the residual errors with the available information from interferometry. The results for the five stars kept out of the training validation process of DTL were then used for independent quality assessment. In the left panel of Fig. 7, we compare the interferometric values for these five stars to results from DTL, DL, and Schweitzer et al. (2019, Schw19), as all these studies used PHOENIX-ACES models. For more than half of the stars, the accuracy of the DTL approach exceeds that of the



**Fig. 7.** Comparison of  $T_{\text{eff}}$  (left) and  $[M/H]$  (right) from different techniques for stars not used in the DTL training process. Values are presented in Tables 7 and 8.

DL and Schw19 approaches, while for only two stars (J15194–077 and J09144+526) the literature gives closer results to the interferometric values than DTL.

By using the same strategy as for  $T_{\text{eff}}$ ,  $[M/H]$  data from Tables 3 and 4 enabled us to use 18 stars for a four-fold cross-validation approach, keeping another 6 stars aside this time (see Table 8). The right panel of Fig. 7 shows that the accuracy of the best deep transferred model for those six stars is significantly better than other existing estimations. For only one star (J22565+165), DL gives a closer parameter estimation to the transferred value than DTL. For the rest of the stars, DTL produces more accurate results than the other methods.

Finally, we quantitatively assessed the accuracy improvements of our method with respect to previous implementations. For metallicity, the median absolute differences (MADs) with respect to the transferred values are 0.07 dex for DTL, 0.27 dex for DL, and 0.13 dex for Schw19. On the other hand, for  $T_{\text{eff}}$ , Schw19 provides the smallest MAD with 60 K. However, DTL outperforms DL with differences of 91 K and 110 K, respectively.

#### 4.1. Uncertainties

Until this point, this paper proposes a technique to transfer the features identified as relevant according to the stellar parameter of interest and the selected flux window. This technique uses precise estimations of stellar properties from interferometry and binary observations to adapt the knowledge domain based on the previous DL feature identification. Our technique provides good estimations for the stars used for quality control as it outperforms the reference estimations for the sample of selected stars in the majority of cases.

When the same set of features is transferred through a different model to adapt the knowledge domain, a different subset of features (components of the feature vector) can be selected. This leads to different parameter estimations after a converged training process, which reflects the effects of the different components of the feature vector. Therefore, we defined uncertainties

**Table 8.**  $[M/H]$  comparison for the stars not used in the training phase.

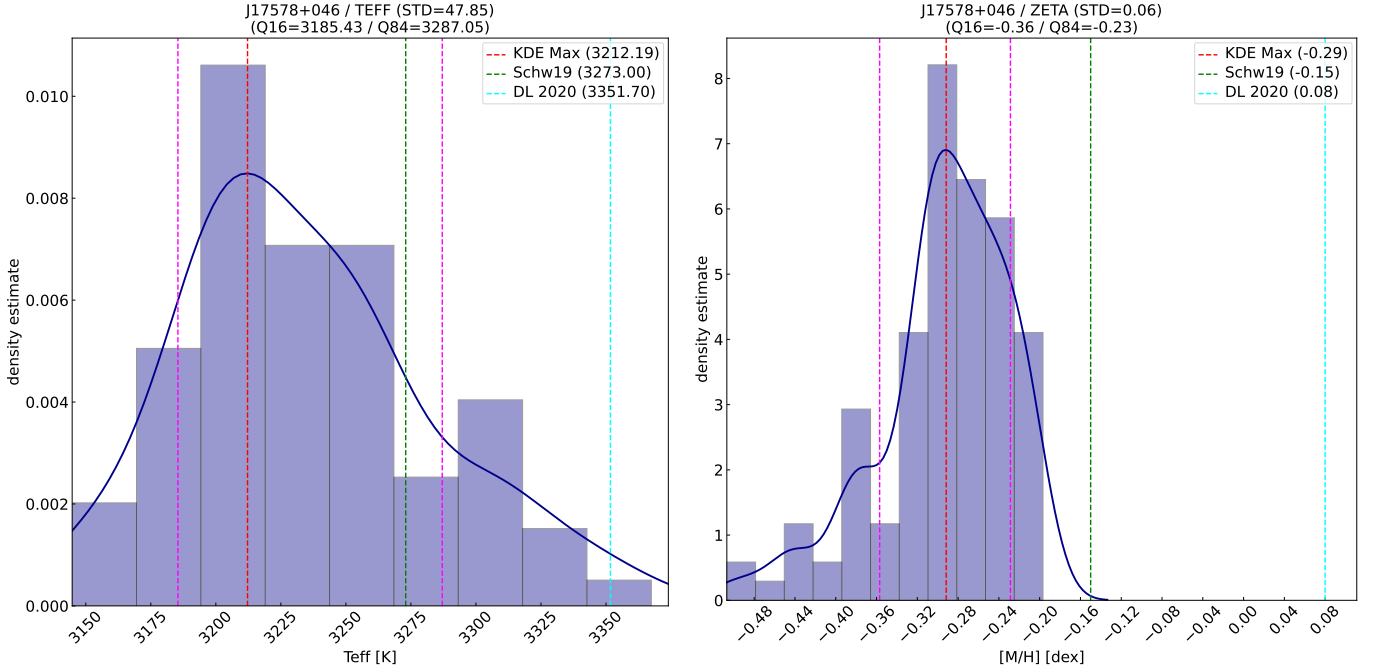
| Karmin     | $[M/H]_{\text{Bin,lit}}$<br>[dex] | $[M/H]_{\text{DL}}$<br>[dex] | $[M/H]_{\text{DTL}}$<br>[dex] | $[M/H]_{\text{Schw19}}$<br>[dex] |
|------------|-----------------------------------|------------------------------|-------------------------------|----------------------------------|
| J04153–076 | $-0.37 \pm 0.02$                  | $+0.25 \pm 0.19$             | $-0.23 \pm 0.05$              | ...                              |
| J07361–031 | $-0.11 \pm 0.03$                  | $+0.16 \pm 0.10$             | $-0.16 \pm 0.04$              | $+0.05 \pm 0.16$                 |
| J11421+267 | $+0.00 \pm 0.11$                  | $+0.18 \pm 0.09$             | $-0.01 \pm 0.03$              | $-0.04 \pm 0.16$                 |
| J13005+056 | $+0.09 \pm 0.10$                  | $+0.35 \pm 0.14$             | $+0.16 \pm 0.10$              | $-0.04 \pm 0.16$                 |
| J17578+046 | $-0.39 \pm 0.12$                  | $+0.08 \pm 0.14$             | $-0.29 \pm 0.06$              | $-0.15 \pm 0.16$                 |
| J22565+165 | $+0.18 \pm 0.09$                  | $+0.23 \pm 0.09$             | $+0.11 \pm 0.03$              | $+0.12 \pm 0.16$                 |

**Notes.** Observed (Bin,lit, i.e.,  $[M/H]$  from primaries of FGK+M systems in the literature), DL-estimated (DL), and DTL-predicted (DTL) values. Uncertainties are their standard deviations.

in the DTL process for estimating stellar parameters as  $\pm 1\sigma$  around the most frequently predicted value for each parameter. In this way, each star has its own uncertainty interval, which does not have to be symmetric.

The DTL process can be repeated several times, making it possible to get different transferred models, each of them weighting the different extracted features differently. The proposal was to retain the models that are over a specific quality threshold and then to aggregate the estimates from them using an integrated KDE technique, as mentioned in Sect. 3.2. An example for  $T_{\text{eff}}$  and for  $[M/H]$  is shown in Fig. 8. Indeed, the observed shapes allow us to discuss the number of DTL models considered to produce good-quality estimations. We selected 80 DTL models to be considered as estimate makers, which is a matter of design. However, for some stars, more evidence was needed in order to reduce the uncertainty level, as illustrated in the right half of the left panel in Fig. 8, where the tail increases the uncertainty value for the 84% quantile. The estimated uncertainties for the selected quality stars are presented in Table 7 for  $T_{\text{eff}}$  and Table 8 for  $[M/H]$  parameters.

After testing the quality of the DTL models for estimating  $T_{\text{eff}}$  and  $[M/H]$ , they were applied to the entire CARMENES data set. The outcome can be found in Table A.1. Stars with



**Fig. 8.** Different DTL estimations of  $T_{\text{eff}}$  (left) and  $[M/H]$  (right) for a single star (J17578+046, Barnard’s Star, M3.5 V). Quantiles  $\pm 1\sigma$  are indicated with dashed magenta lines. Results from DL (cyan) and Schw19 (green) are also included for comparison.

high rotational velocities and therefore with significantly higher uncertainties, were included even when the training data set did not contain this type of object. In Table A.1, there are several parameters with significantly small uncertainties, even smaller than those of the training sample (e.g.,  $\Delta T_{\text{eff}} = 10$  K). The uncertainties that we provide for the estimated stellar parameters refer to the internal error of the method, and therefore they do not take into account uncertainties from the interferometric and binary training samples or from the synthetic gap.

#### 4.2. Comparison with the literature

We compare our DTL results for the full CARMENES sample for  $T_{\text{eff}}$  and  $[M/H]$  with other estimations from the literature. For improved readability, we divide the comparisons into three plots. We also present the Pearson correlation coefficient  $r_p$ , which we use to assess the goodness of the correlation. A summary can be found in Table 9.

Figure 9 presents the comparison with literature values for  $T_{\text{eff}}$ . The top panel shows results from DL, Marfil et al. (2021, Mar21), and Schw19. All three studies used CARMENES data and show a similar pattern, with hotter temperature for stars below 3500 K and above 3800 K compared to DTL. For DL, the dispersion is larger than for the other two methods, also showing a slightly smaller  $r_p$  of 0.87 compared to 0.88 for Mar21 and 0.90 for Schw19. The main factor leading to this effect in DTL is the limited number of samples with temperatures above 4000 K in the training and validation sets. Another possible explanation could be a change in opacity at around 3300 K, but a detailed analysis of the synthetic model structures would be necessary to come to any robust conclusions.

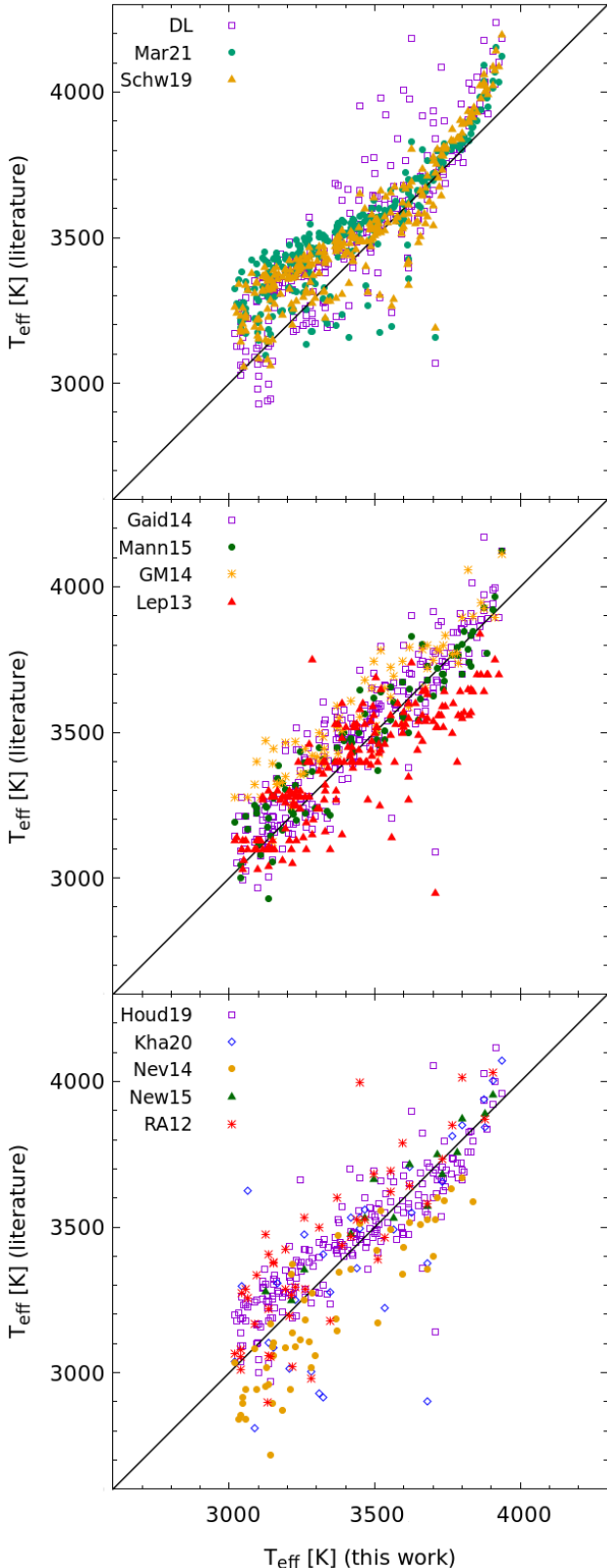
In the middle panel, all literature references determined  $T_{\text{eff}}$  by fitting BT-Settl synthetic models (Allard et al. 2011) to VIS spectra. Gaidos & Mann (2014, GM14) additionally used spectral curvature indices in the K band if there were no VIS spectra available. Perhaps this difference explains a general trend towards hotter temperatures compared to DTL (on average

**Table 9.** Summary of mean difference ( $\bar{\Delta}$ ), standard deviation (std. dev.), and Pearson correlation coefficient ( $r_p$ ) for the comparison between literature values and DTL results.

| Reference | $\bar{\Delta}$<br>$T_{\text{eff}}/[M/H]$<br>[K]/[dex] | std. dev.<br>$T_{\text{eff}}/[M/H]$<br>[K]/[dex] | $r_p$<br>$T_{\text{eff}}/[M/H]$<br>[K]/[dex] |
|-----------|---|--|--|
| DL        | +83/+0.23   | 133/0.16   | 0.87/0.45                                    |
| Mar21     | +97/−0.10   | 119/0.17   | 0.88/ 0.47                                   |
| Schw19    | +85/−0.01   | 110/0.15   | 0.90/0.54                                    |
| GM14      | +121/+0.01  | 96/0.08  | 0.92/0.89                                    |
| Mann15    | +33/+0.02   | 92/0.11  | 0.94/0.87                                    |
| Gaid14    | +54/+0.04   | 137/0.14   | 0.86/0.72                                    |
| Lep13     | −23/...   | 138/...  | 0.81/...                                     |
| Houd19    | +40/...   | 113/...  | 0.89/...                                     |
| Kha20     | −59/−0.08   | 261/0.12   | 0.73/0.81                                    |
| Nev14     | −132/−0.04  | 119/0.15   | 0.89/0.71                                    |
| New15     | +48/−0.06   | 74/0.17  | 0.95/0.69                                    |
| RA12      | +67/+0.01   | 160/0.10   | 0.83/0.86                                    |
| Ditt16    | .../−0.05   | .../0.20   | .../0.50                                     |
| Ter15     | .../−0.03   | .../0.17   | .../0.69                                     |

+121 K, with  $r_p = 0.92$ ). Gaidos et al. (2014, Gaid14) and Mann et al. (2015, Mann15) achieve similar results to with DTL,  $r_p = 0.86$  and 0.94, respectively. The correlation of Lépine et al. (2013, Lep13) is weaker with  $r_p = 0.81$ , providing some cooler values at the hot end of the  $T_{\text{eff}}$  scale.

For the comparison in the bottom panel, literature values were determined from the  $H_2O$ –K2 index as defined by Rojas-Ayala et al. (2012), empirical relations with EWs, and, in the case of Houdebine et al. (2019, Houd19), photometric relations. Results from Houd19 and Newton et al. (2015, New15) show a good correlation with DTL ( $r_p = 0.89$  and 0.95, respectively), while Rojas-Ayala et al. (2012, RA12) is slightly



**Fig. 9.** Comparisons with estimations of  $T_{\text{eff}}$  from different techniques. The black line corresponds to the 1:1 relation.

on the hotter side of the 1:1 relation with  $r_p = 0.83$ . [Khata et al. \(2020, Kha20\)](#) exhibits a large spread ( $r_p = 0.73$ ), and values from [Neves et al. \(2014, Nev14\)](#) are on average 130 K cooler than DTL, but still well correlated with  $r_p = 0.89$ . Nev14 measured

pEWs of most lines and features in the optical and then used reference photometric scales of  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  from [Casagrande et al. \(2008\)](#) and [Neves et al. \(2012\)](#), respectively. Both scales are based on  $[\text{Fe}/\text{H}]$  determinations from FGK+M binaries. The offset that we see between the findings of these latter authors and our  $T_{\text{eff}}$  values also exists compared to other literature works, which indicates an intrinsic underestimation of temperatures by Nev14. Kha20 determined EWs of Mg ( $1.57 \mu\text{m}$ ,  $1.71 \mu\text{m}$ ), Al ( $1.67 \mu\text{m}$ ), and the  $\text{H}_2\text{O}$  index in the  $H$  band, and performed linear regression on 12 M-dwarf calibrator stars with interferometrically measured  $T_{\text{eff}}$  to derive a temperature relation. The standard deviation of the residuals of the calibrators amounted to 102 K. However, the spread with respect to our results and other literature studies is much larger, with a standard deviation of 261 K. An indication of a higher deviation in the comparison with the literature can already be seen in Fig. 5 of Kha20, where these authors compare their results with  $T_{\text{eff}}$  from Mann15 and RA12, and measure standard deviations of 164 and 158 K, respectively. At this point, we cannot give a clear explanation for this behavior. Overall, the correlation between DTL and the literature is quite good, except for Mar21, Schw19, and Kha20, where we are not sure about the source of the differences seen.

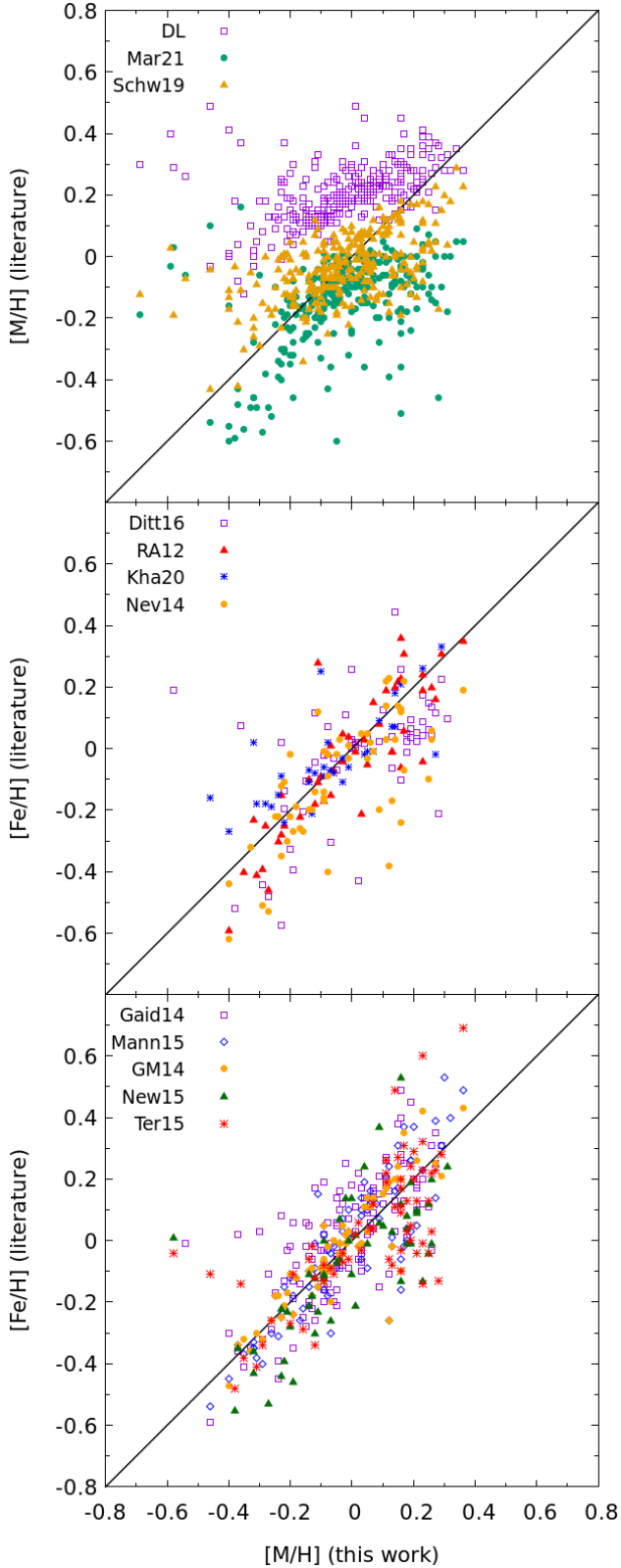
Figure 10 presents the same comparisons for metallicity. As explained by [Passegger et al. \(2020\)](#), our derived  $[\text{M}/\text{H}]$  values directly translate into identical  $[\text{Fe}/\text{H}]$  values. Therefore, we can compare  $[\text{M}/\text{H}]$  values with  $[\text{Fe}/\text{H}]$  from the literature. Furthermore,  $[\text{Fe}/\text{H}]$  is often used as a proxy for  $[\text{M}/\text{H}]$  in the literature.

The top panel again shows the results from DL, Mar21, and Schw19. There is an offset towards more metal-rich values for DL (on average 0.23 dex,  $r_p = 0.45$ ), and towards more metal-poor values for Mar21 (on average  $-0.11$  dex,  $r_p = 0.46$ ). A possible explanation for these offsets might be that Mar21 reported  $[\text{Fe}/\text{H}]$  corrected for  $\alpha$ -enhancement. A trend of more metal-rich values for DL was already shown by [Passegger et al. \(2020, 2022\)](#). The results from Schw19 are consistent overall with DTL, with  $r_p = 0.54$ , although they exhibit a certain spread.

The middle panel of Fig. 10 compares similar determination methods from the literature. RA12 and Kha20 derived  $[\text{Fe}/\text{H}]$  using the  $\text{H}_2\text{O}$ –K2 index and equivalent widths (EWs) of Na I and Ca I in the NIR. Nev14 also incorporated a relation with EWs, but only [Dittmann et al. \(2016, Ditt16\)](#) determined  $[\text{Fe}/\text{H}]$  from a color–magnitude–metallicity relation. This might explain the large spread of the latter values, resulting in  $r_p = 0.50$ . Results from Nev14 are overall more metal-poor than those provided by DTL and show a large spread as well, but a better correlation than Ditt16, with  $r_p = 0.71$ . However, values from RA12 and Kha20 correspond even better with DTL, showing  $r_p = 0.86$  and  $0.81$ , respectively.

The literature values shown in the bottom panel of Fig. 10 were determined using empirical relations between atomic line strength, Na I and Ca I EWs, the  $\text{H}_2\text{O}$ –K2 index, and metallicity calibrated with FGK+M binaries based on [Mann et al. \(2013a,b, 2014\)](#) relationships. The values provided by GM14, Mann15, and Gaid14 are highly correlated, with those of Gaid14 showing the least spread ( $r_p = 0.89$ ,  $0.87$ , and  $0.72$ , respectively). Values from New15 are generally slightly more metal-poor, with a smaller correlation coefficient of  $r_p = 0.69$ . For higher metallicities, [Terrien et al. \(2015, Ter15\)](#) derived a larger spread and some outliers at both ends. The correlation coefficient is the same as for New15. Similar to  $T_{\text{eff}}$ , DTL values for metallicity correspond well with most of the literature, and an improvement with respect to DL can be appreciated, which is very promising.





**Fig. 10.** Comparisons with estimations of  $[M/H]$  from different techniques. The black line corresponds to the 1:1 relation.

## 5. Summary and conclusions

We present a DTL neural network technique that improves the estimation of the stellar parameters  $T_{\text{eff}}$  and  $[M/H]$  for M dwarfs from high-S/N, high-resolution optical spectroscopy obtained with CARMENES. The initial DL model was trained with

PHOENIX-ACES synthetic spectra, which confer the advantage that they allow a sufficient number of spectra to be generated with known stellar parameters. Based on the DL convolutional features, different DTL models were trained and tested. To use a more robust procedure, a cross-validation scheme was adopted. Using the proposed technique could help to bridge the synthetic gap affecting stellar parameter estimation based on synthetic libraries. However, a larger stellar sample covering a wider spectral range is needed to verify this.

Before applying the created models to a large data set, we defined an independent quality assessment procedure based on specific stars for which high-quality stellar parameter estimations are available. This assessment shows that the DTL technique has good prediction capabilities. In addition, we incorporated an uncertainty estimation procedure based on considering the diversity of estimates from different transferred models, as well as an aggregation procedure. Such an estimation is flux-window dependent, but also star dependent, because the trained DTL models below the convergence threshold depend on them.

Another relevant aspect to be considered for parameter estimation concerns the selected flux windows, as they have their own influence on the vector of features, and, in the end, on the estimated parameters. A possible continuation of the line of research in this field is applying reinforced learning techniques based on the behavior of the selected flux windows.

Finally, and importantly, a limitation of the proposed method is the parameter range of the observed transferred knowledge. This means that the parameters of stars with large  $v \sin i$  values cannot be estimated rigorously with this technique, as no interferometric or FGK-star companion values are available with such large values of rotational velocity. In the same way, the transferred knowledge works only for  $T_{\text{eff}}$  higher than 3100 K, as no cooler stars were part of the training set yet. Therefore, our current analysis is limited to spectral types between M0 V and M6 V.

In summary, we propose an innovative technique that can increase the value of its predictions as new high-quality stellar parameters – namely  $T_{\text{eff}}$  from interferometry and  $[M/H]$  from FGK+M systems – become available in the near future. The current data sample used is close to the operational limits of the technique, while in some cases the data set was complemented with stellar parameters estimated from the literature. Therefore, improvements are expected to be possible when more stars with highly reliable stellar parameters and high-resolution spectra become available. In the meantime, the lack of a sufficiently large number of samples is a limitation for the technique.

*Acknowledgements.* We thank an anonymous referee for helpful comments that improved the quality of this paper. 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 German Max-Planck-Gesellschaft (MPG), the Spanish Consejo Superior de Investigaciones Científicas (CSIC), 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 Spanish Ministry of Economy, the German Science Foundation through the Major Research Instrumentation Programme and DFG 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 also acknowledge the financial support from MCIN/AEI/10.13039/501100011033/ and the ERDF through projects PID2021-125627OB-C31, PID2020-112949GB-I00, and PID2019-109522GB-C51/2/3/4,

and fellowship FPU15/01476; the Fundação para a Ciência e a Tecnologia through and ERDF through grants UID/FIS/04434/2019, UIDB/04434/2020, and UIDP/04434/2020, PTDC/FIS-AST/28953/2017; grant RTI2018-094614-B-I00 (SMASHING) into the “Programa Estatal de I+D+i Orientada a los Retos de la Sociedad” funded by MCIN/AEI/10.13039/501100011033, and COMPETE2020 - Programa Operacional Competitividade e Internacionalização POCI-01-0145-FEDER-028953; the University of La Laguna and the EU Next Generation funds through the Margarita Salas Fellowship from the Spanish Ministerio de Universidades UNI/551/2021-May 26; and NASA through grant NNX17AG24G.

## References

- Abadi, M., Barham, P., Chen, J., et al. 2016, in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265
- Allard, F., Homeier, D., & Freytag, B. 2011, *ASP Conf. Ser.*, **448**, 91
- Andrews, J. J., Chanamé, J., & Agüeros, M. A. 2018, *MNRAS*, **473**, 5393
- Antoniadis-Karnavas, A., Sousa, S. G., Delgado-Mena, E., et al. 2020, *A&A*, **636**, A9
- Awang Iskandar, D. N., Zijlstra, A. A., McDonald, I., et al. 2020, *Galaxies*, **8**, 88
- Bialek, S., Fabbro, S., Venn, K. A., et al. 2020, *MNRAS*, **498**, 3817
- Bonfils, X., Delfosse, X., Udry, S., et al. 2005, *A&A*, **442**, 635
- Boyajian, T. S., von Braun, K., van Belle, G., et al. 2012, *ApJ*, **757**, 112
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
- Caballero, J. A., Guàrdia, J., López del Fresno, M., et al. 2016, *Proc. SPIE*, **9910**, 99100E
- Casagrande, L., Flynn, C., & Bessell, M. 2008, *MNRAS*, **389**, 585
- Chen, Y., Girardi, L., Bressan, A., et al. 2014, *MNRAS*, **444**, 2525
- Chen, Y., Bressan, A., Girardi, L., et al. 2015, *MNRAS*, **452**, 1068
- Cifuentes, C., Caballero, J., Cortés-Contreras, M., et al. 2020, *A&A*, **642**, A115
- Czesla, S., Schröter, S., Schneider, C. P., et al. 2019, *Astrophysics Source Code Library*, [[record ascl:1906.010](https://arxiv.org/abs/1906.010)]
- Demory, B. O., Ségransan, D., Forveille, T., et al. 2009, *A&A*, **505**, 205
- Desidera, S., Gratton, R. G., Lucatello, S., & Claudi, R. U. 2006, *A&A*, **454**, 581
- Dhital, S., West, A. A., Stassun, K. G., et al. 2012, *AJ*, **143**, 67
- Dittmann, J. A., Irwin, J. M., Charbonneau, D., & Newton, E. R. 2016, *ApJ*, **818**, 153
- Fabbro, S., Venn, K., O’Brian, T., et al. 2018, *MNRAS*, **475**, 2978
- Gaia Collaboration (Brown, A. G. A., et al.) 2018, *A&A*, **616**, A1
- Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, **649**, A1
- Gaidos, E., & Mann, A. W. 2014, *ApJ*, **791**, 54
- Gaidos, E., Mann, A. W., Lépine, S., et al. 2014, *MNRAS*, **443**, 2561
- Gao, Y., & Mosalam, K. M. 2018, *Computer-Aided Civil Infrastruc. Eng.*, **33**, 748
- Goodfellow, I., Bengio, Y., & Courville, A. 2016, *Deep Learning* (Cambridge: MIT Press),
- Han, X., Zhang, Z., Ding, N., et al. 2021, *AI Open*, **2**, 225
- Hartman, J. D., Bayliss, D., Brahm, R., et al. 2015, *AJ*, **149**, 166
- Houdebine, É. R., Mullan, D. J., Doyle, J. G., et al. 2019, *AJ*, **158**, 56
- Husser, T.-O., Wende-von Berg, S., Dreizler, S., et al. 2013, *A&A*, **553**, A6
- Johnson, J. A., & Apps, K. 2009, *ApJ*, **699**, 933
- Karpathy, A., & Fei-Fei, L. 2015, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3128
- Khata, D., Mondal, S., Das, R., Ghosh, S., & Ghosh, S. 2020, *MNRAS*, **493**, 4533
- Kielty, C. L., Bialek, S., Fabbro, S., et al. 2018, *Int. Soc. Opt. Photon.*, **10707**, 107072W
- Kraus, M., & Feuerriegel, S. 2017, *Decision Support Syst.*, **104**, 38
- Lépine, S., Hilton, E. J., Mann, A. W., et al. 2013, *AJ*, **145**, 102
- Maldonado, J., Affer, L., Micela, G., et al. 2015, *A&A*, **577**, A132
- 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., Deacon, N. R., Gaidos, E., et al. 2014, *AJ*, **147**, 160
- Mann, A. W., Feiden, G. A., Gaidos, E., Boyajian, T., & von Braun, K. 2015, *ApJ*, **804**, 64
- Marfil, E., Tabernero, H. M., Montes, D., et al. 2021, *A&A*, **656**, A162
- McInnes, L., Healy, J., Saul, N., & Grossberger, L. 2018, *J. Open Source Softw.*, **3**, 861
- Minglei, W., Jingchang, P., Zhenping, Y., Xiaoming, K., & Yude, B. 2020, *Optik*, **218**, 165004
- Mittal, S., & Vaishay, S. 2019, *J. Syst. Architect.*, **99**, 101635
- Montes, D., González-Peinado, R., Tabernero, H. M., et al. 2018, *MNRAS*, **479**, 1332
- Nagel, E., Czesla, S., Kaminski, A., et al. 2020, *A&A*, submitted
- 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
- Newton, E. R., Charbonneau, D., Irwin, J., et al. 2014, *AJ*, **147**, 20
- Newton, E. R., Charbonneau, D., Irwin, J., & Mann, A. W. 2015, *ApJ*, **800**, 85
- O’Brian, T., Ting, Y.-S., Fabbro, S., et al. 2020, *ArXiv e-prints* [[arXiv:2007.03112](https://arxiv.org/abs/2007.03112)]
- Pan, S. J., & Yang, Q. 2010, *IEEE Trans. Knowl. Data Eng.*, **22**, 1345
- Passegger, V. M., Reiners, A., Jeffers, S. V., et al. 2018, *A&A*, **615**, A6
- Passegger, V. M., Schweitzer, A., Shulyak, D., et al. 2019, *A&A*, **627**, A161
- Passegger, V. M., Bello-García, A., Ordieres-Meré, J., et al. 2020, *A&A*, **642**, A22
- Passegger, V. M., Bello-García, A., Ordieres-Meré, J., et al. 2022, *A&A*, **658**, A194
- Quirrenbach, A., Amado, P. J., Ribas, I., et al. 2018, *SPIE Conf. Ser.*, **10702**, 107020W
- Quirrenbach, A., CARMENES Consortium, Amado, P. J., et al. 2020, *SPIE Conf. Ser.*, **114473**, 114473C
- Rabus, M., Lachaume, R., Jordán, A., et al. 2019, *MNRAS*, **484**, 2674
- Raffel, C., Shazeer, N., Roberts, A., et al. 2020, *J. Mach. Learn. Res.*, **21**, 1
- Refaeilzadeh, P., Tang, L., & Liu, H. 2009, *Encyclopedia of Database Systems* (Berlin: Springer), 5, 532
- Reiners, A., Zechmeister, M., Caballero, J. A., et al. 2018, *A&A*, **612**, A49
- Rodríguez Martínez, R., Ballard, S., Mayo, A., et al. 2019, *AJ*, **158**, 135
- 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
- Sarkar, D., Bali, R., & Ghosh, T. 2018, *Hands-On Transfer Learning with Python: Implement Advanced Deep Learning and Neural Network Models Using TensorFlow and Keras* (Birmingham, UK: Packt Publishing Ltd)
- Schlaufman, K. C., & Laughlin, G. 2010, *A&A*, **519**, A105
- Schweitzer, A., Passegger, V. M., Cifuentes, C., et al. 2019, *A&A*, **625**, A68
- Scott, D. W. 2015, *Multivariate Density Estimation: Theory, Practice, and Visualization* (Hoboken: John Wiley & Sons)
- Ségransan, D., Kervella, P., Forveille, T., & Queloz, D. 2003, *A&A*, **397**, L5
- Tabernero, H. M., Marfil, E., Montes, D., & González Hernández, J. I. 2022, *A&A*, **657**, A66
- Tan, C., Sun, F., Kong, T., et al. 2018a, in *International Conference on Artificial Neural Networks* (Berlin: Springer), 270
- Tan, C., Sun, F., Kong, T., et al. 2018b, in *Artificial Neural Networks and Machine Learning – ICANN 2018*, eds. V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, & I. Maglogiannis (Cham: Springer International Publishing), 270
- Tang, J., Bressan, A., Rosenfield, P., et al. 2014, *MNRAS*, **445**, 4287
- Terrell, G. R., & Scott, D. W. 1992, *Ann. Stat.*, **20**, 1236
- Terrien, R. C., Mahadevan, S., Bender, C. F., et al. 2012, *ApJ*, **747**, L38
- Terrien, R. C., Mahadevan, S., Bender, C. F., Deshpande, R., & Robertson, P. 2015, *ApJ*, **802**, L10
- Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. 2019, *PloS one*, **14**, e0224365
- van Leeuwen, F. 2007, *A&A*, **474**, 653
- Vilalta, R. 2018, *J. Phys. Conf. Ser.*, **1085**, 052014
- von Braun, K., Boyajian, T. S., Kane, S. R., et al. 2011, *ApJ*, **729**, L26
- von Braun, K., Boyajian, T. S., Kane, S. R., et al. 2012, *ApJ*, **753**, 171
- von Braun, K., Boyajian, T. S., van Belle, G. T., et al. 2014, *MNRAS*, **438**, 2413
- Wang, D., & Li, M. 2017, *Information Sci.*, **412**, 210
- Wei, W., Huerta, E., Whitmore, B. C., et al. 2020, *MNRAS*, **493**, 3178
- Zechmeister, M., Anglada-Escudé, G., & Reiners, A. 2014, *A&A*, **561**, A59
- Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, *A&A*, **609**, A12
- Zechmeister, M., Dreizler, S., Ribas, I., et al. 2019, *A&A*, **627**, A49
- Zhao, Z., Zhang, Q., Yu, X., et al. 2021, *IEEE Transactions on Instrumentation and Measurement*, **70**, 1

## Appendix A: DTL estimations for CARMENES survey stars

Table A.1:  $T_{\text{eff}}$  and [M/H] values for CARMENES stars estimated by the DTL method<sup>(a)</sup>.

| Karmn                     | $T_{\text{eff}}$ [K]                | [M/H] [dex]                             |
|---------------------------|-------------------------------------|---|
| JJ00051+457               | 3721 <sup>+8</sup> <sub>-8</sub>    | -0.02 <sup>+0.04</sup> <sub>-0.02</sub> |
| J00067-075 <sup>1,2</sup> | 3140 <sup>+249</sup> <sub>-38</sub> | -0.19 <sup>+0.28</sup> <sub>-0.03</sub> |
| J00162+198E               | 3149 <sup>+46</sup> <sub>-31</sub>  | -0.02 <sup>+0.06</sup> <sub>-0.03</sub> |
| J00183+440 <sup>1,2</sup> | 3563 <sup>+1</sup> <sub>-9</sub>    | -0.26 <sup>+...</sup> <sub>-...</sub>   |
| J00184+440                | 3219 <sup>+71</sup> <sub>-41</sub>  | -0.07 <sup>+0.06</sup> <sub>-0.10</sub> |
| J00286-066                | 3246 <sup>+40</sup> <sub>-22</sub>  | -0.06 <sup>+0.02</sup> <sub>-0.04</sub> |
| J00389+306                | 3503 <sup>+10</sup> <sub>-13</sub>  | -0.09 <sup>+0.02</sup> <sub>-0.03</sub> |
| J00570+450                | 3341 <sup>+30</sup> <sub>-19</sub>  | -0.09 <sup>+0.01</sup> <sub>-0.04</sub> |
| J01013+613                | 3489 <sup>+20</sup> <sub>-16</sub>  | -0.18 <sup>+0.03</sup> <sub>-0.02</sub> |
| J01025+716                | 3449 <sup>+33</sup> <sub>-18</sub>  | -0.03 <sup>+0.03</sup> <sub>-0.05</sub> |
| J01026+623                | 3743 <sup>+9</sup> <sub>-8</sub>    | +0.10 <sup>+0.06</sup> <sub>-0.01</sub> |
| ...                       |                                     |   |

**Notes.** <sup>(a)</sup>Stars marked with “1” and “2” were used for training the DTL  $T_{\text{eff}}$  and [M/H] models, respectively. The number “3” refers to stars with rotational velocity  $v \sin i > 10 \text{ km s}^{-1}$ . Some  $1\sigma$  uncertainties and quantiles for stars marked 1 and 2 are missing because the different predictions were too close. In these few cases, the quantiles could be substituted by the standard deviation. The full table is available at the CDS.