

Photometric metallicity prediction of RR Lyrae stars in the Gaia optical and K_s infrared wavebands by deep learning

1 ISTVÁN DÉKÁNY  ¹ AND EVA K. GREBEL  ¹

2 ¹ *Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr. 12-14, 69120 Heidelberg, Germany*

3 ABSTRACT

4 RR Lyrae stars are useful chemical tracers thanks to the empirical relationship between their heavy-
5 element abundance and the shape of their light curves. However, the consistent and accurate calibration
6 of this relation across multiple photometric wavebands has been lacking. We have devised a new method
7 for the metallicity estimation of fundamental-mode RR Lyrae stars in the Gaia optical G and near-
8 infrared VISTA K_s wavebands by deep learning. First, an existing metallicity prediction method is
9 applied to large photometric data sets, which are then used to train long short-term memory recurrent
10 neural networks for the regression of the [Fe/H] to the light curves in other wavebands. This approach
11 allows an unbiased transfer of our accurate, spectroscopically calibrated I -band formula to additional
12 bands at the expense of minimal noise. We achieve a low mean absolute error of 0.1 dex and high
13 R^2 regression performance of 0.84 and 0.93 for the K_s and G bands, respectively, measured by cross-
14 validation. The resulting predictive models are deployed on the Gaia DR2 and VVV inner-bulge
15 RR Lyrae catalogs. We estimate mean metallicities of ~ -1.35 dex for the inner bulge and ~ -1.7 for
16 the halo, which are significantly less than values obtained by earlier photometric prediction methods.
17 Using our results, we establish a public catalog of photometric metallicities of over 60,000 Galactic
18 RR Lyrae stars, and provide an all-sky map of the resulting RR Lyrae metallicity distribution. The
19 software code used for training and deploying our recurrent neural networks is made publicly available
20 in the open-source domain.

21 **Keywords:** RR Lyrae variable stars(1410) – Metallicity(1031) – Light curves(918) – Deep learning
22 (1938)

23 1. INTRODUCTION

24 RR Lyrae variables are low-mass, core helium-burning stars undergoing radial pulsation. Since these bright stars
25 abound in old stellar populations, are easy to identify, and their absolute luminosities and heavy-element abundances
26 can be inferred from their light variations (see Bhardwaj 2022, for a recent review), they have been widely employed as
27 astrophysical tracer objects within the Local Group of galaxies (e.g., Tanakul & Sarajedini 2018; Clementini et al. 2019;
28 Soszyński et al. 2019; Dékány & Grebel 2020). In addition, the photometric estimation of the metallicity distributions
29 of large RR Lyrae samples offer the possibility to constrain the formation epoch and early chemical enrichment of the
30 underlying old stellar populations (see, e.g., Savino et al. 2020, and references therein).

31 Although the existence of an empirical relationship between the shape of an RR Lyrae star’s light curve and its
32 metallicity has been known since the pioneering work by Jurcsik & Kovács (1996), its accurate calibration in multiple
33 photometric bands has been a challenging task, mainly due to the small number of stars with high-dispersion spectro-
34scopic (HDS) metallicity measurements. Over the past ~ 25 years, various empirical formulae have been established to
35 predict the [Fe/H] from the Fourier regression parameters of the light curves in various photometric wavebands, follow-
36 ing different strategies. Since direct regression to the HDS metallicities has long suffered from small sample sizes (e.g.,
37 Nemec et al. 2013), most authors (e.g., Jurcsik & Kovács 1996; Smolec 2005; Ngeow et al. 2016; Iorio & Belokurov
38 2020; Mullen et al. 2021) addressed the problem by relying on low-dispersion spectroscopic or spectro-photometric
39 metallicities estimated from spectral indices (see, e.g., Layden 1994; Crestani et al. 2021). Although the diversity
40 of light-curve shapes in the training sets could be increased by a few times this way, such an approach required the
41 intermediate calibration of the spectral indices to the same limited amount of HDS data. Another approach was to

transfer a predictive formula established for one waveband to another by linearly transforming the regressors, i.e., the Fourier parameters of the light curve, between these bands (e.g., Skowron et al. 2016; Clementini et al. 2019). However, such transformations are not only inherently noisy, but they also depend on the metallicity.

Generally speaking, the limited amount of heterogeneous data, strong heteroskedasticity, and large errors in the regressors lead to various systematic biases in the metallicity prediction formulae. Such systematics were often also difficult to explore and quantify due to the lack of large samples with accurate light curves in multiple wavebands. Fortunately, HDS analyses of RR Lyrae stars have recently proliferated (see Crestani et al. 2021, and references therein), calling for the revision of earlier photometric [Fe/H] estimation methods. Exploiting the new state-of-the-art abundance measurements, we recently established new empirical predictive models of the [Fe/H] of both fundamental-mode (RRab) and first-overtone (RRc) RR Lyrae stars from their light curves in the Cousins I filter, a highly important waveband in contemporary studies of the Milky Way and the Magellanic Clouds. In the case of this band, the training sample was sufficiently large and accurate, so that with careful feature selection and probabilistic modeling of the uncertainties, it allowed a direct regression to the HDS data, removing the additional error-prone but customary intermediate step of calibrating spectral indices to the HDS [Fe/H] values. Our analysis revealed large systematic biases in several earlier formulae, that led to an ~ 0.4 dex positive bias in previous estimates of the metallicity distribution functions of old stellar populations, such as the bulge and the Magellanic Clouds.

In the era of large time-domain sky surveys that discovered and monitor $\sim 10^5$ RR Lyrae stars in the Local Group and keep on detecting new ones, accurate photometric predictive models of their metallicities are needed in order to unlock their full potential as population tracers; and ideally we need them in all the various optical and infrared wavebands in which such surveys operate. One way to accomplish that is to consider this task as a set of independent regression problems of the HDS measurements on the light curves in each waveband. However, this approach is hindered by the fact that in most cases, the number of stars with both HDS abundance measurements and accurate light curves is still too small. Using small, separate training data sets for different wavebands is not only prone to systematics in the predicted [Fe/H], but also makes it hard to keep all separate relations up-to-date as new data become available.

An alternative approach is to fit a single predictive formula to the HDS measurements using a waveband in which a sufficiently large HDS training data set already exists. Then, transfer the resulting *base estimator* to other wavebands; but instead of transforming the regressors, handle each problem as a regression of the light curve shapes to the predicted *photometric* [Fe/H] as the response variable. By taking this approach, large training data sets become available, allowing us to use deep learning for the regression, i.e., highly complex models capable of extracting every last bit of predictive information from the light curves. This, in turn, offers the possibility of a highly precise and unbiased transfer of the original relation to additional wavebands, with the cost of introducing only minimal noise. Therefore, the safe combination of photometric data across different surveys for the inference of metallicity distribution functions of large stellar populations becomes possible. Moreover, upon the availability of additional HDS measurements, an immediate homogeneous update of the predictive models in all wavebands becomes very straightforward.

It is important to emphasize that by adopting this method, we make the important assumption of physical similarity between the HDS training data set of the base estimator and the training sets of the deep learned models. Specifically, we assume that these data sets are not systematically different in terms of other, latent physical variables that affect the light-curve shapes (e.g., in their helium abundances; see, e.g., Dékány et al. 2021). We note that the same assumption is also implicitly made when any other estimator of photometric metallicity estimation is either calibrated or applied to a new data set.

The above approach of photometric metallicity estimation was originally put forward by Hajdu et al. (2018), who trained an ensemble of multi-layer perceptrons (i.e., classical fully connected neural networks) to predict the [Fe/H] of RRab stars from their K_s -band photometric time series, using a parametric representation of the light curves as regressors, and I -band photometric [Fe/H] estimates obtained with the formula of Smolec (2005) as response variables. In this study, we build upon this idea to obtain deep-learned predictive models of the [Fe/H] of RRab stars from their light curves in the Gaia G and VISTA K_s wavebands. Our analysis comprises the following major steps. First, using our recent predictive formula (Dékány et al. 2021, hereafter D21), we create our development data sets by computing photometric [Fe/H] estimates from the I -band light curves of a large number of RRab stars that were also observed in either the Gaia optical G or the VISTA near-infrared (near-IR) K_s filters. These $[Fe/H]_I$ estimates are then used as our response variables for the development of additional predictive models. To estimate the $[Fe/H]_I$ from G and K_s light curves, we train Recurrent Neural Networks (RNNs), which allow us to use the original photometric time series as input data, instead of relying on a classical parametric representation of the light curves. Finally, the trained RNNs

Table 1. I -band photometric parameters and metallicities of RRab stars in the bulge and disk areas of the OCVS.

Field ^a	ID	R.A. ^b [hms]	DEC. ^b [dms]	[Fe/H] _{I}
b	00001	17:05:07.49	-32:37:57.2	-1.712
b	00003	17:05:09.88	-32:39:52.8	-1.296
b	00004	17:05:15.22	-32:50:13.2	-1.868
b	00005	17:05:21.98	-32:39:42.5	-2.596
b	00006	17:05:29.00	-32:33:37.4	-1.426

$\langle I \rangle$	$N_{ep.}$	Period [d]	$A_{tot.}$	A_1	A_2	A_3	ϕ_{21}	ϕ_{31}	C_φ ^c	S/N
15.595	121	0.732586	0.447	0.167	0.073	0.043	9.5708	6.680	0.956	413.2
16.451	121	0.515465	0.696	0.215	0.118	0.077	8.9977	5.700	0.950	428.2
15.965	122	0.651696	0.264	0.113	0.036	0.018	9.3813	6.315	0.965	154.5
15.336	121	0.565511	0.606	0.241	0.098	0.029	8.9175	5.006	0.962	72.3
15.846	121	0.478877	0.804	0.245	0.121	0.097	8.8742	5.396	0.958	547.3

^a‘b’: bulge, ‘d’: disk; original OCVS designations: OGLE-BLG-RRLYR-ID and OGLE-GD-RRLYR-ID, respectively.

^b Coordinates of J2000.0 epoch.

^cPhase coverage: 1 – maximum phase lag.

NOTE— Results are shown for stars that pass the following criteria: $C_\varphi \geq 0.8$, $A_{tot.} \leq 1.2$, $S/N \geq 50$.

This table is available in its entirety in machine-readable form.

are deployed on the RRab catalogs of the Gaia and the VISTA Variables in the Vía Láctea surveys, in order to deliver a public database of accurate photometric metallicity estimates.

2. PHOTOMETRIC DATA AND THEIR REPRESENTATION

As the first step of establishing the photometric data sets for the development of our predictive models for the metallicity, we compute I -band photometric [Fe/H] estimates for a very large number of Galactic RRab stars using the D21 formula, that is directly calibrated to HDS spectroscopic measurements. These [Fe/H] _{I} estimates will serve as our response variables (i.e., the ground truth) for training and validating the predictive models for the optical and near-IR bands. We used the public I -band light curves of RRab stars acquired by the Optical Gravitational Lensing Experiment (OGLE, [Udalski et al. 2015](#)) cataloged in the OGLE Collection of Variable Stars¹ (OCVS) of the Galactic bulge and disk, comprising tens of thousands of objects.

The I -band ϕ_{31} and A_1 Fourier parameters that serve as input features of the D21 formula were computed following the regression procedure² discussed by [Dékány et al. \(2021\)](#). In brief, a robust fitting of a truncated Fourier-sum with iterative outlier rejection and order optimization via cross-validation is followed by a Gaussian Process Regression (GPR) of the phase-folded light curves. The final parameters are obtained from the Fourier representation of the mean GPR model. The results of our analysis are displayed Table 1. The uncertainties in the resulting [Fe/H] _{I} values were estimated by drawing numerous random samples from the GPR model at the original observational phases for each object, repeating the same regression procedure as above for each realization, and calculating the standard deviation of the resulting [Fe/H] predictions. These uncertainties will be used for sample weighting in the training and validation of the predictive models.

At the next step, we cross-matched the celestial coordinates of the OCVS RRab stars with the RR Lyrae catalog from the Gaia data release 2 (DR2) ([Clementini et al. 2019](#)), and the K_s -band point source catalogs of the VISTA

¹ <http://ogledb.astroww.edu.pl/~ogle/OCVS/>

² The corresponding software library is available at <https://github.com/idekany/lcfit>

Variables in the Vía Láctea (VVV) ESO Public Survey (Minniti et al. 2010) created by the VISTA Data Flow System (VDFS, Emerson et al. 2004) and provided by the Cambridge Astronomy Survey Unit (CASU). The cross-match yielded $\sim 13,700$ objects with Gaia and $\sim 29,600$ stars with VVV photometry.

The Gaia light curves of the cross-matched objects were directly retrieved through the Gaia@AIP database interface of the Leibniz Institute for Astrophysics Potsdam. The corresponding K_s -band VVV light curves were obtained by following the procedure discussed in Dékány et al. (2018), and subsequently applying the photometric zero-point corrections of Hajdu et al. (2020).

Both the Gaia and VVV light curves were pre-processed by subjecting them to the previously discussed regression procedure, which was also used for the OGLE light curves. Since the sampling, precision, and temporal baseline of the OGLE light curves are far superior than those of the Gaia and VVV data, we always used the periods derived from the former. Likewise, we adopted the OGLE variable star classifications for similar reasons. In case of the Gaia data, we opted not to use the per-epoch photometric quality flags provided in DR2 for the omission of bad data, instead we relied on our own outlier rejection mechanism because the former would have led to the culling of too many useful measurements. In case of the VVV data, we selected the optimal photometric aperture for each object by maximizing the signal-to-noise (S/N) ratio in the cleaned light curve.

For the predictive modeling of the [Fe/H] from the light curves, we use the following 2-dimensional sequences as input variables (i.e., regressors):

$$\mathbf{x}^{} = \begin{pmatrix} m^{} - \langle m \rangle \\ \phi_P^{} \cdot P \end{pmatrix}, \quad t = \{1, \dots, N_{ep}\}, \quad (1)$$

$$\phi_P(T) = \text{mod}[(T + P \cdot \Phi_1/(2\pi))/P], \quad (2)$$

Here, $m^{}$ and $\phi_P^{}$ are the magnitudes and corresponding pulsation phases of the light curve, respectively, $\langle m \rangle$ is the mean magnitude, Φ_1 is the phase of the first Fourier term, T is the observation time, and N_{ep} is the number of observational epochs. In addition, the VVV K_s light curves were binned to a maximum of 60 points per time series in order to speed up the training of the neural networks.

By using the original (phased) measurements instead of a traditional parametric representation of the light curves, we virtually eliminate any potential representation error in the regressors, and allow a natural propagation of uncertainties in the photometric time series in the form of scatter in the input sequences. We note that the above representation of the time series is not strictly non-parametric due to the phase-folding of the light curves: an accurate knowledge of the period is obviously required or else the input sequence will suffer from bias. In addition, the input sequences are phase-aligned by Φ_1 , which thereby acts as a nuisance parameter, and its uncertainty can still lead to some representation error due to sequence misalignment. In practice, however, the robustness of this phase alignment method ensures negligible bias in the input for all but the noisiest and most ill-sampled light curves.

In addition to the broad optical G waveband, Gaia also acquires simultaneous photometric measurements in its bp (blue) and rp (red) filters. In principle, we could use all three time-series as 6-dimensional sequences (3 magnitudes and their corresponding phases) at the input of the RNN. However, the bp and rp light curves in the common OCVS–Gaia data set have much lower S/N and significantly less useful data points than their G -band counterparts (Fig. 1). Therefore, they would mostly add noise to our model, and introducing quality criteria to counter this would shrink the data set too much. Moreover, bad data points in the different bands occur at different phases, thus their independent omission would disrupt the matching phase values across the different bands for the same star, the handling of which would require increased complexity of the RNN model. In view of the above points, we decided to rely solely on the G -band Gaia light curves in our analysis.

The obtained database of phase-folded, phase-aligned, and cleaned photometric time series was narrowed down by imposing various criteria for controlling their quality. Light curves with very few measurements, low S/N , poor phase coverage, and uncertain $[\text{Fe}/\text{H}]_I$ estimates were excluded from the regressions. We also trimmed outliers beyond the main locus of the period–amplitude–metallicity distributions. Formally, such criteria can be considered as tunable hyper-parameters whose optimal values maximize the predictive model’s performance at validation (see Sect. 3.2), and provide a good trade-off between data quality and quantity. Our final criteria for the development data sets are summarized below:

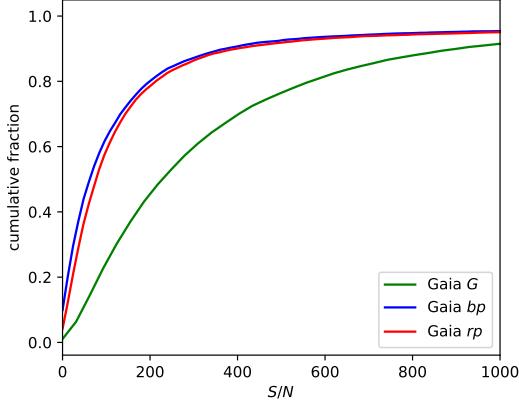


Figure 1. Cumulative distributions of the S/N of the Gaia light curves in the G , bp , and rp wavebands for the Galactic RRab stars in the common sample with the OCVS.

$$\mathbf{Gaia} \; G: \; C_\varphi > 0.85; \; S/N > 30; \; N_{ep.} > 20; \; A_{tot.} < 1.4; \quad -2.7 < [\text{Fe}/\text{H}]_I < 0; \; \sigma_{[\text{Fe}/\text{H}]_I} < 0.3 \quad (3)$$

$$\mathbf{VVV} \; K_S: \; C_\varphi > 0.9; \; S/N > 100; \; N_{ep.} > 100; \; A_{tot.} < 0.45; \; -2.5 < [\text{Fe}/\text{H}]_I < 0; \; \sigma_{[\text{Fe}/\text{H}]_I} < 0.15; \; P_{Bl.} < 0.7(4)$$

Here, C_φ denotes the phase coverage (i.e., 1–maximum phase lag), A_{tot} is the total (peak-to-valley) amplitude of the GPR light-curve model, N_{ep} is the number of epochs in the light curve, and $\sigma_{[\text{Fe}/\text{H}]_I}$ is the uncertainty in the $[\text{Fe}/\text{H}]_I$ values, estimated from the GPR model of the I -band light curve. Moreover, P_{Bl} is the probability that a star shows the Blazhko effect (see, e.g., [Smolec 2016](#), for a review), i.e., a periodic amplitude and/or phase modulation of the light curve according to the machine-learned classifier of [Prudil & Skarka \(2017\)](#). Our selection criterion excludes stars with strongly modulated light curves from the K_s development set. Although the mean light curves of Blazhko RRab stars do not seem to be discrepant from the general relationship between metallicity and light curve shape in the optical ([Dékány et al. 2021](#)), a similar conclusion has never been drawn in the infrared. Also, undersampled modulation cycles can lead to bias in the light curve shape and thus affect metallicity prediction. Even though the strength of the Blazhko modulation is generally larger in the optical than in the near-IR [Jurcsik et al. \(2018\)](#) for the same star, introducing a selection criterion by P_{Bl} for the G -band data, unlike for the K_s band, did not give any advantage. The much more subtle dependence of the light curve shape on the metallicity in the K_s band and the highly non-uniform sampling of the VVV light curves probably both contribute to this behavior.

By applying the selection criteria in Eqs. 3 and 4, we remove most Gaia stars from the crowded inner bulge where they have generally poor sampling, and exclude most stars from the relatively small common disk sample of OGLE and VVV, resulting in almost disjunct development data sets for the G and K_s bands. They comprise 4458 G -band and 7534 K_s -band light curves, which are shown in Fig. 2. The Bailey (period–amplitude) diagrams of the development data sets are shown in Fig. 3 with the corresponding $[\text{Fe}/\text{H}]_I$ values color-coded; highlighting the strong dependency of the period on the metallicity, and the more non-linear structure of this diagram in the near-IR compared to the optical domain.

Figure 4 shows the $[\text{Fe}/\text{H}]_I$ distributions of these data sets, revealing the strong data imbalance, which particularly affects the K_s band. Since both data sets predominantly comprise bulge stars, their metallicity distributions are strongly peaked at around -1.4 dex, with relatively weak tails on both their metal-rich and metal-poor sides.

In order to handle the data imbalance, we introduced density-dependent sample weights for the training of our regression model. First, we computed Gaussian kernel density estimates of the $[\text{Fe}/\text{H}]_I$ distributions, evaluated them for every object in the development sets, and assigned a density weight w_d to each data point by taking the inverse of the estimated normalized density. Additionally, we introduced density threshold values ρ_G and ρ_{K_s} , beyond which uniform w_d values are used in order to prevent an excessive influence of the relatively few data points in the tails of the $[\text{Fe}/\text{H}]_I$ distributions in the regression. We treat these thresholds as tunable hyper-parameters, and find their optimal

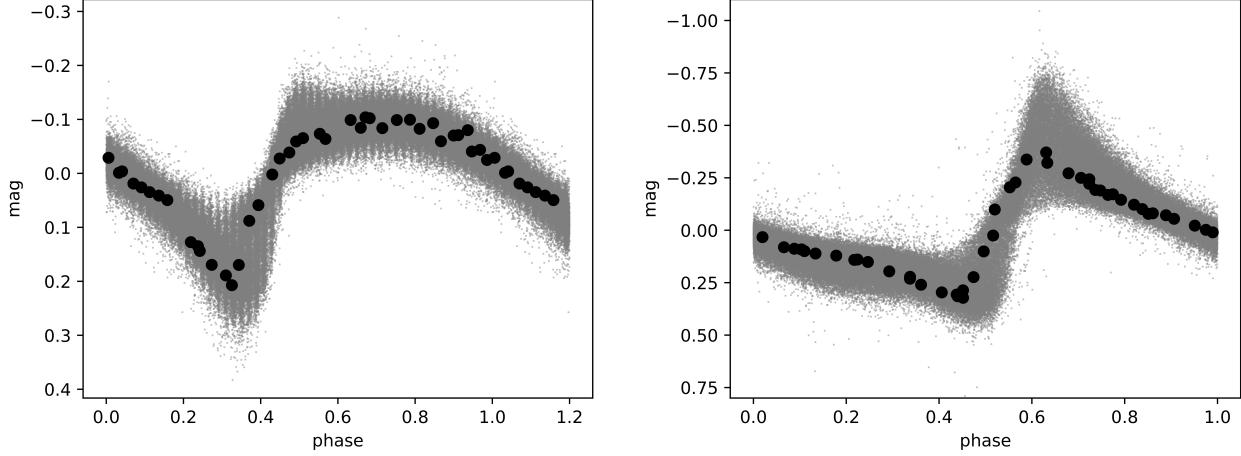


Figure 2. Phase-folded, phase-aligned, mean-subtracted light curves of the K_s -band (left) and G -band (right) development data sets after all pre-processing steps have been applied. Gray dots show all data, black points highlight the data for a single RRab star.

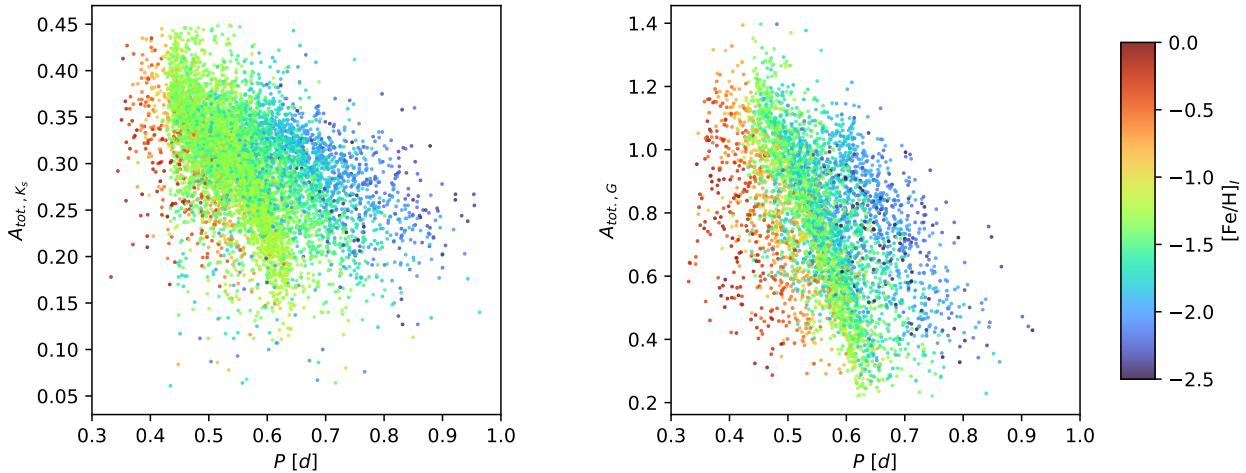


Figure 3. Period–amplitude–metallicity distributions of the VVV K_s -band (left) and Gaia G -band (right) light curves of the RRab stars in our respective development data sets. The metallicity values were estimated from the I -band light curves of the same objects (see text) and are denoted with a color map shown at the right side of the figure.

values to be $\rho_G = 0.5$ and $\rho_{K_s} = 0.9$. In addition, to take the uncertainties in individual $[Fe/H]_I$ values also into account, we introduce w_u weights as their normalized squared inverse. The final sample weights are then computed as the product $w = w_d \cdot w_u$. Figure 4 shows the distributions of w_d and w for both the K_s - and G -band development sets.

3. PREDICTIVE MODELING

3.1. Long Short-Term Memory Recurrent Neural Networks

Recurrent neural networks (RNNs) encompass a variety of mathematical models that are capable to accurately approximate extremely complex, highly non-linear interrelations in numerical sequences. They have been successfully

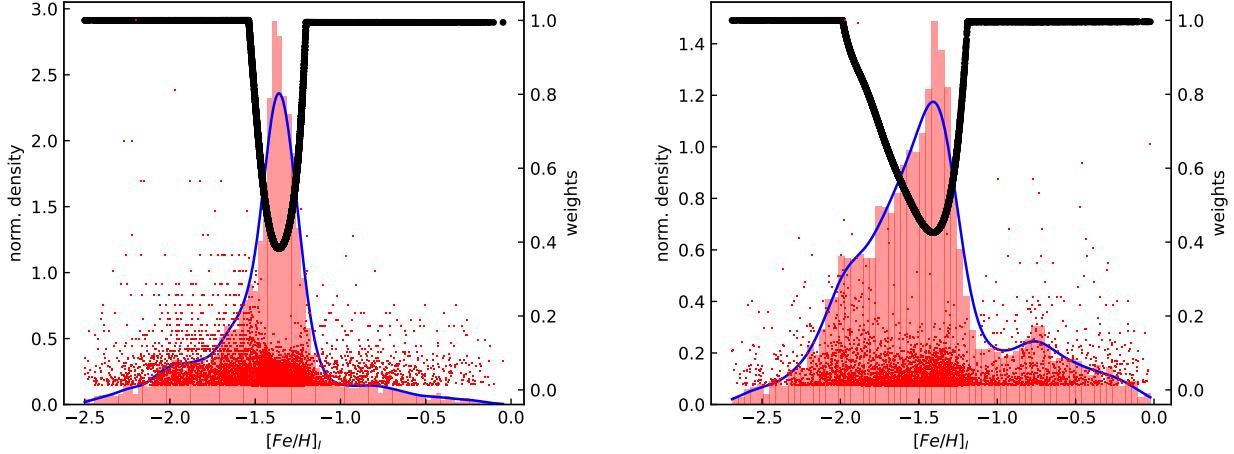


Figure 4. Distributions of the metallicities of the K_s - (left) and G -band (right) development data sets and their sample weights. Red bars and blue curves show the histograms and kernel density estimates of the $[Fe/H]_I$ values. Black symbols denote the (normalized) weights computed from the inverse of the density. Red points show the final sample weights.

employed for supervised learning scenarios in a wide range of research areas involving sequential data, ranging from neural translation to finances (see, e.g., Yu et al. 2019, for a recent review).

The simplest type of RNN takes a sequence $\mathbf{x}^{<t>}$ at its input, and performs the following recurrent transformation at every time step $t \in \{1, T_x\}$:

$$\mathbf{a}^{<t>} = \tanh(\mathbf{W}_{aa}\mathbf{a}^{<t-1>} + \mathbf{W}_{ax}\mathbf{x}^{<t>} + \mathbf{b}_a). \quad (5)$$

Here, the output $a^{<t>}$ is called the activation of the t -th time step, and the \mathbf{W}_{aa} and \mathbf{W}_{ax} weight matrices and \mathbf{b}_a bias vector are free parameters of the RNN, which are shared across all time steps. In case of using the RNN for predicting a single real number from each input sequence, the following operation is performed on the activation vector from the last time step:

$$\hat{y} = (\mathbf{w}_{ya}\mathbf{a}^{<T_x>} + b_y), \quad (6)$$

where \hat{y} is the predicted value of the response variable y , and the w_{ya} weight vector and the b_y bias term are free parameters of the model.

The RNN's complexity can be adjusted to a specific problem by changing the dimensionality of the weight matrices (a.k.a. the number of neurons) and hence the dimension of the resulting activation vectors, and by using multiple recurrent layers. In case of the latter, the input of layer l will be the sequence of activation vectors $\mathbf{a}^{[l-1]<t>}$ computed by the previous layer $l-1$; each layer will have their own $\mathbf{W}_{aa}^{[l]}$, $\mathbf{W}_{ax}^{[l]}$ weight matrices and $\mathbf{b}_a^{[l]}$ bias vectors, and the prediction will be computed from the activation vector of the last time step of the last layer by Eq. 3.1. The optimal model parameters of the RNN are found by minimizing an appropriate cost function, e.g., the mean squared error (MSE) in case of a regression problem.

Modern alternatives to the classical RNN architecture outlined in Eq. 3.1 include modifications aiming to increase the speed of training and to enable a better representation of long-term dependencies in lengthy input sequences. Perhaps the most successful of these has proven to be the Long Short-Term Memory (LSTM) network originally proposed by Hochreiter & Schmidhuber (1997), which has been the primary architecture of choice in a wide spectrum of use cases (see, e.g., Houdt et al. 2020, and references herein). Recently, LSTMs have been successfully employed in time-domain astronomy for the classification of photometric time series, e.g. to identify RR Lyrae stars (Dékány & Grebel 2020) and detect stellar flares (Vida et al. 2021).

A standard LSTM unit features a memory cell designed to retain information over an arbitrary number of time steps, and thus enable the network to learn long-term dependencies. It also contains update, forget, and output gates that regulate the information flow to/from the memory cell and enable to reset its state. The cell and the gates each have free parameters that are learned by the model (i.e., fitted to the data). For a more in-depth mathematical discussion of LSTM networks in the context of astronomical light curves, we refer to Dékány & Grebel (2020), see their Eqs. 6–11 and Fig. 1).

3.2. Model selection and optimization

The optimization of a machine-learned predictive model consists of two fundamental steps, training and hyper-parameter tuning. During training, the optimal values of a model’s parameters are found by minimizing a cost function for a training data set. Since the gradients of the cost function can be explicitly expressed in the case of neural networks, gradient descent-based optimization algorithms can be employed. The model’s hyper-parameters that govern its complexity, such as the number of layers and the number of neurons in each of them, along with the data filtering and weighing thresholds (see Sect. 2), as well as the choice of regularization and its parameter(s) have fixed values during training. The optimal values of these hyper-parameters are searched for by maximizing a performance metric for a validation data set that the model has not seen during training.

For training, we used the MSE cost function with sample weights discussed in Sect. 2. In order to prevent overfitting the model to the training set, we experimented with two different methods, namely kernel regularization and dropout. In the former, the J cost function includes a term that is proportional to the norm of weights that are used for the linear transformation of the input sequence and/or the recurrent state (i.e., previous activation vector):

$$J \propto \sum_{i=1}^N (y - \hat{y})^2 + \lambda \|\mathbf{W}\| , \quad (7)$$

where N is the number of training data. The coefficient λ of the weight norm \mathbf{W} is called the regularization parameter, which, together with the type of norm used, are handled as hyper-parameters of the model. In addition, we also employed the dropout (Srivastava et al. 2014) regularization technique, whereby a random P_d fraction of the activation vectors’ elements are randomly dropped (i.e., assigned to 0) during training, with the dropout probability being another hyper-parameter.

The performance metric to be maximized during hyper-parameter tuning should reflect how well the trained model fulfills our expectations on yet unseen data (i.e., the validation set). On the one hand, we would like to minimize our mean prediction errors, which we formulate as maximizing the coefficient of determination, i.e., the R^2 score:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} , \quad (8)$$

where \bar{y} is the mean of the response variable in the validation set. The R^2 score measures predictive performance relative to the total variation of the response variable. Its best possible value is 1 (i.e., no prediction errors), and it can take arbitrarily low values. On the other hand, we expect the predicted distribution of the metallicity to be as close to the real distribution as possible. In other words, we want the predictive performance to be homogeneously high across the entire range of metallicities covered by the development set. We formulate this by minimizing the symmetric difference measure between the true and predicted distributions, i.e., their Jensen–Shannon (JS) divergence:

$$D_{JS} = \frac{1}{2}(D_{KL}(P||M) + D_{KL}(Q||M)) . \quad (9)$$

Here, P and Q are the compared distributions, $M = (P + Q)/2$, and D_{KL} is the Kullback–Leibler divergence. The JS divergence is bound to the $[0, 1]$ interval and takes the value of 0 in case the compared distributions are identical. An optimal value of the R^2 metric does not necessarily mean that D_{JS} takes its optimum as well, in case the predictive performance changes with [Fe/H]. Since the data imbalance in our regression problem may enhance such a skew in the model’s performance, we search for the optimal hyper-parameter values by jointly optimizing R^2 and D_{JS} , which we achieve by using their ratio as our custom metric:

Table 2. Properties of the final K_s - and G -band predictive models of the metallicity

Layer	K_s -band model		G -band model	
	hyper-params.	params.	hyper-params.	params.
biLSTM1	32 neurons, $\lambda = 3 \cdot 10^{-6}$	8960	16 neurons, $\lambda = 5 \cdot 10^{-6}$	2432
dropout1	$P_d = 0.1$	0	$P_d = 0.1$	0
biLSTM2	32 neurons, $\lambda = 3 \cdot 10^{-6}$	24832	16 neurons, $\lambda = 5 \cdot 10^{-6}$	6272
dropout2	$P_d = 0.1$	0	$P_d = 0.1$	0
dense	—	65	—	33

NOTE—The number of neurons refers to a single direction of the network.

275

$$\mathcal{M} = R^2 / D_{JS} \quad (10)$$

276

277 Since our development data sets have modest sizes, a single held-out validation set would significantly decrease
 278 the diversity of the training data. To avoid this, we could apply k -fold cross validation (CV) instead, whereby the
 279 development set is randomly split into k training and k disjunct validation sets, a mean performance metric is computed
 280 from the predictions on the latter; and the model with the best set of hyper-parameters is finally refitted to the entire
 281 development set in order to obtain a single “monolithic” final model. Due to the limited amount of data however,
 282 this approach can still be volatile to the distribution changes introduced by the training-validation splits. Firstly, the
 283 same hyper-parameters can lead to significant over- or underfitting in different folds. Secondly, the true performance
 284 of the refitted model can be significantly different from the performance estimate from CV due to the reduced data
 285 set sizes in the latter. It might even be the case that the optimal hyper-parameters are different for the final fit to the
 286 development set from the ones found by CV.

287 A more robust predictive model can be obtained by a modified approach to k -fold CV. In order to prevent the
 288 model’s volatility to data splitting during each CV fold, we apply the “early stopping” regularization technique, which
 289 is commonly used in deep learning scenarios. During training, we monitor the model’s performance by evaluating
 290 it on both the current training and validation folds after each training epoch. When overfitting occurs, the model’s
 291 performance measured on the training set will keep increasing as the learning algorithm starts to fit the model to
 292 the noise pattern of the training set, while the thus far increasing validation performance will turn over and start
 293 to decrease, due to the model’s poor generalization. We obtain the optimal model for each CV fold by stopping its
 294 training at this turnover point, and estimate its overall performance by the mean \mathcal{M} metric across the folds. Finally,
 295 instead of refitting the model with the best hyper-parameters to the development set as we would in standard k -fold
 296 CV, we instead keep the k models and use their mean prediction. This approach is commonly referred to as training an
 297 “early stopped k -fold ensemble” of models, and is not only more robust than a single monolithic model, but offers the
 298 advantage that instead of providing a point estimate for an input, its prediction’s uncertainty can also be estimated
 299 as the standard deviation of the ensemble’s k predictions.

300 An extensive grid search of hyper-parameters was performed to find the best model ensemble. We tried one- and
 301 two-layer mono- and bidirectional LSTM (biLSTM) architectures with various numbers of neurons in each. The effect
 302 of both L1 and L2 norms were tested in Eq. 3.2 for the cost function’s weight penalty (a.k.a. lasso and Tikhonov
 303 regularization, respectively), along with a grid of λ values and dropout probabilities. We trained each model with
 304 the Adam optimization algorithm (Kingma & Ba 2014) with a learning rate of 0.005 until the optimal early stopping
 305 epoch was found (typically after a few thousand epochs). In order to have a sufficient number of training examples
 306 from our entire [Fe/H] range, we used a large mini-batch size of 256³.

³ Each training iteration is performed on a mini-batch of training data. During one training epoch, all training data are used once.

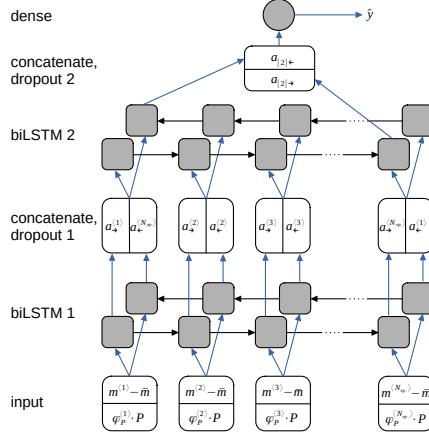


Figure 5. Schematic architectural graph of our predictive models. Grey boxes represent neural network (LSTM and dense) units, empty boxes denote their inputs and outputs. The activation vector at the output of the first biLSTM layer is 64 and 32 dimensional for the K_s - and G -band model, respectively.

Figure 5 shows a schematic picture of our best-performing predictive models, while their main attributes are listed in Table 2. For both the K_s and G bands, they contain two biLSTM layers, and use L1 regularization on the recurrent weights. In addition, dropout is employed after both the first and second layers.

3.3. Regression performance

The regression performance of our final model ensembles has been measured using various common metrics applied on the union of the k -fold validation data sets, except in the case of the R^2 metric, for which the mean was computed. All of these metrics estimate very low generalization error, i.e., high prediction accuracy on unseen data. Their values for both the K_s and the G bands are shown in Table 3, which also includes the same metrics for the training data as reference. The similarity between the values obtained for the training and validation data sets indicates an excellent bias-variance tradeoff of the final model ensembles.

Figure 6 compares our model ensembles' [Fe/H] predictions with their ground truth values, i.e., the metallicities computed from the corresponding I -band light-curves. The residuals do not show any significant structure and contain only few strong outliers. At the very metal-poor end ($[Fe/H] < -2$), some K_s -band metallicity predictions lie out from the main locus with a sizable positive bias for a relatively small number of stars. Upon the inspection of both their K_s - and I -band light curves, we did not identify peculiarity in their photometry. It might be possible that a latent physical parameter with a large influence on the light-curve shape (such as the He content) is systematically different for these objects compared to the rest of the sample. Another possibility is that the dependency of the K_s light-curve shape on the [Fe/H] is too complicated at the metal-poor end for our model to accurately learn it from the available data.

The photometric metallicity distributions predicted by our model ensembles for the training and validation data sets are directly compared to the ground truth distributions (i.e., I -band photometric metallicities) in Fig. 7. For both wavebands, the predicted distributions are virtually identical to the original ones, not showing any bias, and even subtle features are accurately recovered, which tallies with their very low D_{JS} values.

We further test the consistency of our models by directly comparing their [Fe/H] predictions with HDS measurements of RRab stars in the Gaia DR2 catalog. We cross-matched the latter with the same large spectroscopic sample that was compiled from the literature for the development of our D21 I -band photometric metallicity estimator, using the results of Crestani et al. (2021, C21), For et al. (2011, F11), Chadid et al. (2017, C17), Sneden et al. (2017, S17),

Table 3. Properties of the final K_s - and G -band predictive models of the metallicity

Metric	K_s -band model		G -band model	
	tr.	val.	tr.	val.
R^2	0.87	0.84	0.96	0.93
wRMSE	0.12	0.13	0.10	0.13
wMAE	0.09	0.10	0.07	0.10
RMSE	0.14	0.15	0.15	0.18
MAE	0.10	0.11	0.12	0.13
medAE	0.07	0.08	0.09	0.10
D_{JS}	0.004	0.004	0.001	0.001
R^2/D_{JS}	218	239	956	960

NOTE—(w)RMSE: (weighted) root mean squared error; (w)MAE: (weighted) mean absolute error; medAE: median absolute error; D_{JS} : Jensen-Shannon divergence

³³⁴ Clementini et al. (1995), Fernley & Barnes (1996), Lambert et al. (1996), Liu et al. (2013), Nemec et al. (2013), Govea
³³⁵ et al. (2014), Pancino et al. (2015), and Andrievsky et al. (2018). The corrections for the systematic [Fe/H] offsets
³³⁶ computed by D21 were applied to these measurements to match the homogeneous scale of C21+F11+C17+S17. We
³³⁷ selected the objects with well-sampled and accurate G -band light curves, which resulted 213 individual data points
³³⁸ of 60 unique objects, 28 with single and 32 with multiple HDS [Fe/H] measurements. We note that this data set
³³⁹ largely overlaps the one used for calibrating the D21 predictive formula, but is too small and imbalanced for a reliable
³⁴⁰ direct calibration of a similar formula for the G band. Figure 8 compares the predicted $[Fe/H]_G$ values with the
³⁴¹ corresponding individual HDS measurements. The two sets of values are in good agreement, showing no systematics
³⁴² in their residual, and their scatter is consistent with the prediction uncertainties measured on the validation data set
³⁴³ (c.f. Fig 6). This important cross-check indicates the pertinence of our underlying assumption about the physical
³⁴⁴ similarity between the HDS training set of the D21 formula and the development data set of our LSTM model, thus
³⁴⁵ supporting the consistency of our approach.

4. PHOTOMETRIC METALLICITY CATALOGS

³⁴⁶ As the final goal of this study, we computed photometric metallicities for a large number of RRab stars by deploying
³⁴⁷ the neural networks developed in Sect. 3 on the photometric databases of the Gaia and VVV surveys.

4.1. G -band photometric metallicities of the Gaia DR2 RRab stars

³⁴⁹ We analyzed the G -band light curves of all 98024 RRab stars in the Gaia DR2 RR Lyrae catalog following the
³⁵⁰ same procedure that was applied on the development set, as described in Sect. 2. We adopted the periods and type
³⁵¹ classifications from Clementini et al. (2019). Our predictive model was applied to all light curves that passed the
³⁵² following quality criteria: $C > 0.85$; $S/N > 30$; $N_{ep.} > 20$; $A_{tot.} < 1.4$, resulting in a *target set* of 58652 RRab
³⁵³ stars. Their metallicity estimates, along with various photometric attributes, are presented in Table 4. The sample is
³⁵⁴ dominated by halo stars, with additional large samples contributing to it from the Magellanic Clouds, the bulge, and
³⁵⁵ the Milky Way’s thick disk.

³⁵⁷ In our D21 study, we have shown that several earlier methods for metallicity estimation are affected by systematic
³⁵⁸ positive biases with respect to our new I -band [Fe/H] prediction formula, which is directly calibrated to modern HDS
³⁵⁹ measurements. Since the neural networks in this study were trained on I -band photometric metallicities obtained with
³⁶⁰ the latter, we can expect to see similar differences between our model predictions and other G -band [Fe/H] estimates
³⁶¹ from the literature.

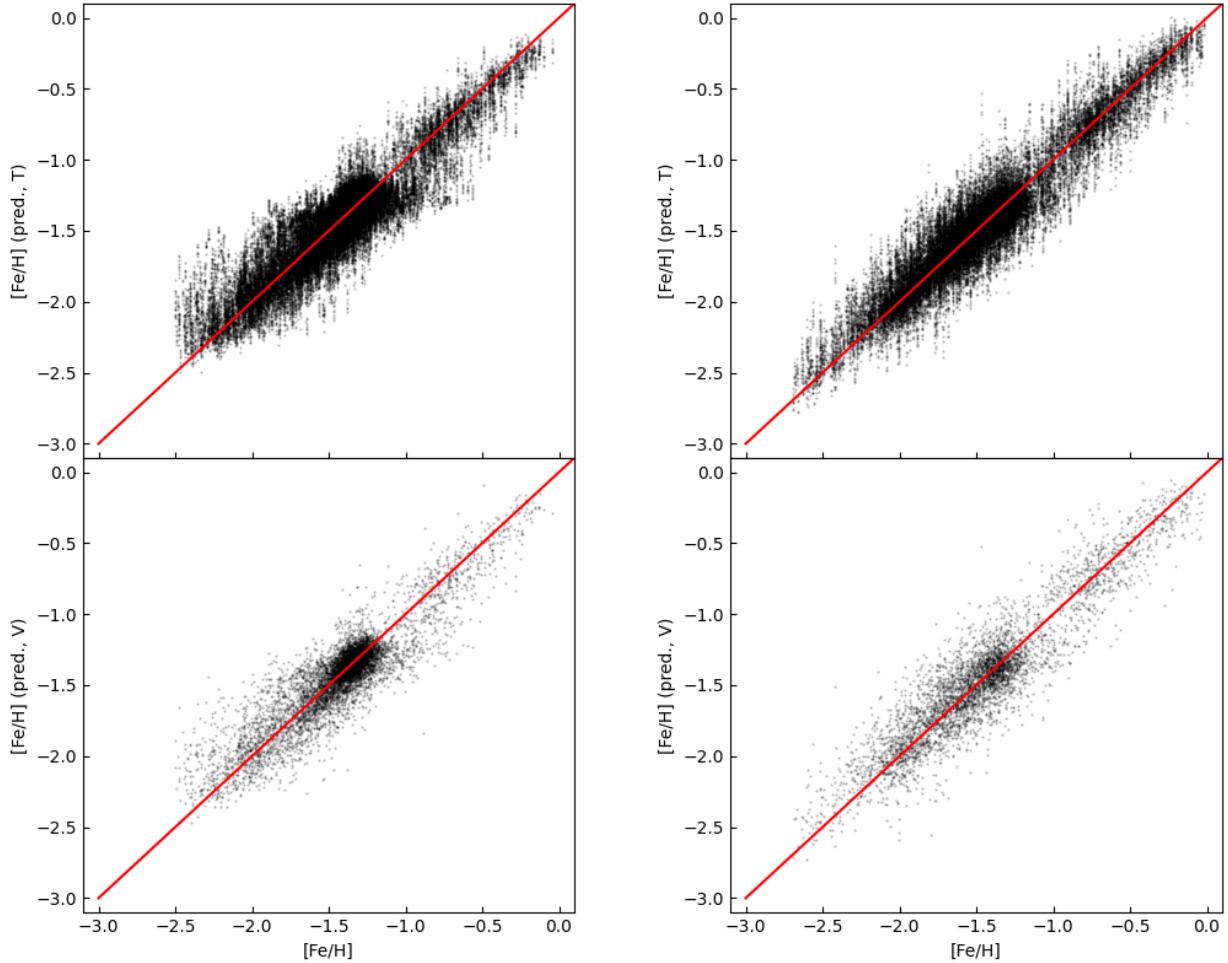


Figure 6. Predicted *vs* true photometric metallicities from the best-performing predictive models for the K_s (left) and G (right) wavebands. The top and bottom panels show the full training and validation data sets, respectively. The red lines denote the identity function.

We directly compared our results to metallicities published in the Gaia DR2 RR Lyrae catalog by Clementini et al. (2019) ($[Fe/H]_{DR2}$), which were available for 43681 objects in our target set. These estimates are based on the quadratic formula by Nemec et al. (2013), which relates the metallicity to the period and the ϕ_{31} Fourier parameter in the *Kepler* photometric waveband, and was calibrated on a tiny sample of 27 HDS measurements. To apply it to G -band light-curve parameters, Clementini et al. (2019) performed two successive linear transformations of the G -band ϕ_{31} parameters, converting them first to the V band, and from there to the *Kepler* band. The left panel of Fig 9 confronts the $[Fe/H]_G$ values from our study with their $[Fe/H]_{DR2}$ equivalents. The latter are affected by a very large positive bias, and the residual strongly correlates with the period, with long-periodic stars having extreme biases of up to ~ 2 dex. In addition, the distribution shows a clump of metal-rich stars standing out from the main locus that

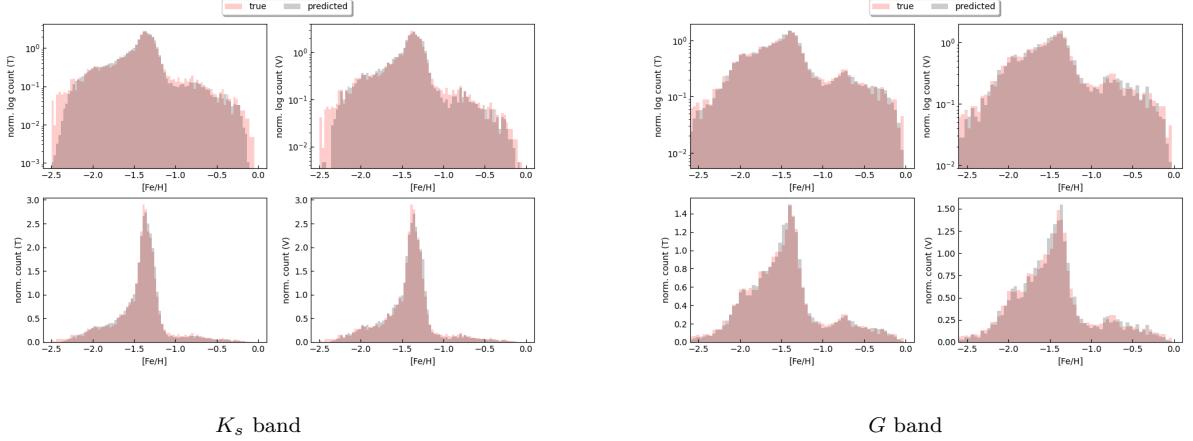


Figure 7. Histograms of the true (red) and predicted (gray) $[\text{Fe}/\text{H}]$ values from our K_s - and G -band model ensembles for the training (T, left) and validation (V, right) data sets. The upper and lower panels show the histograms on logarithmic and linear scales, respectively.

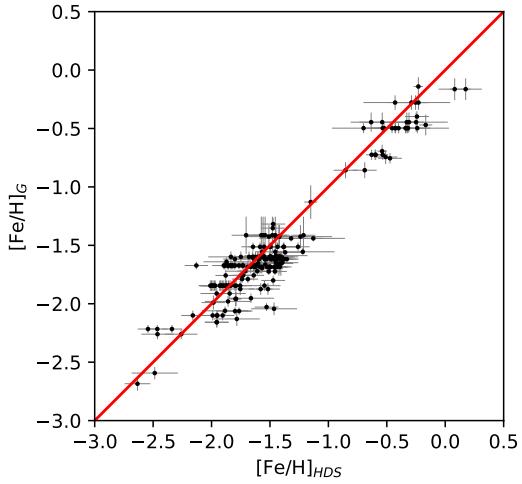


Figure 8. G -band mean $[\text{Fe}/\text{H}]$ predictions of 60 RRab stars by our LSTM model ensemble plotted against the individual HDS metallicity measurements of the same objects. The red line denotes the identity function.

correspond to a metal-rich tail of the $[\text{Fe}/\text{H}]_G$ distribution, but remains blended with the rest of the $[\text{Fe}/\text{H}]_{DR2}$ sample due to its aforementioned bias.

To improve the quality of RR Lyrae metallicity estimation from Gaia photometry, Iorio & Belokurov (2020, IB20) fitted a linear relation to the pulsation period and the G -band ϕ_{31} Fourier parameter using 84 stars with $[\text{Fe}/\text{H}]$ values adopted from Layden (1994). The latter were derived from low-resolution spectral indices using a linear formula that was calibrated to early HDS measurements of a handful of field and cluster RRab stars. We note that the same Layden (1994) sample, with slight adjustments, forms partly the basis of the widely used photometric metallicity prediction formulae of Jurcsik & Kovács (1996) and Smolec (2005). In the middle panel of Fig. 9, we compare the IB20 metallicities with our $[\text{Fe}/\text{H}]_G$ estimates for the same set of stars as for our previous comparison. They show a much better agreement with our results than the $[\text{Fe}/\text{H}]_{DR2}$ values, especially at the metal-rich end, and instead of the huge positive bias observed in the case of DR2 metallicities of long-periodic, metal-poor stars, we see a smaller, but still quite large bias in the opposite, negative direction. Overall, the IB20 formula also overestimates the metallicities

Table 4. G -band photometric metallicity estimates and basic light-curve attributes of Gaia DR2 RRab stars

Gaia DR2 source.id	$[\text{Fe}/\text{H}]_G^a$	$\sigma([\text{Fe}/\text{H}]_G)^b$	$\langle G \rangle$	$N_{ep.}$	Period [d]	$A_{tot.}$	C_φ	S/N
11977018818239872	-1.73	0.04	16.708	38	0.515610	0.903	0.881	125.5
11991514330833408	-2.01	0.03	18.929	36	0.564529	0.824	0.923	152.0
36039689056047872	-1.45	0.02	19.131	29	0.600486	0.570	0.880	108.0
36110989810386560	-2.10	0.17	18.246	29	0.607803	0.775	0.901	43.2
36246122364621440	-1.71	0.03	18.904	30	0.542566	0.893	0.886	177.9

^aMean of predictions from a model ensemble.

^bStandard deviation of predictions from a model ensemble.

NOTE—This table is available in its entirety in machine-readable form.

with respect to our model, with a positive bias of up to 0.2–0.3 dex at the center of the distribution. The particularly poor DR2 and IB20 predictions at the long-periodic, metal-poor limit can probably be attributed mainly to the lack of such stars in the data sets used to calibrate those formulae.

Finally, we compare the photometric metallicity distributions using all three methods in the right panel of Fig. 9, revealing large systematic differences between them. In addition to large offsets between their centers, their shapes are also remarkably different. Our distribution features a long metal-rich tail, which is absent from the DR2 MDF, and only slightly apparent in the IB20 MDF, because it blends with the bulk of the distributions due to the positive bias dominating the latter.

Figure 10 shows the celestial distribution of our target set, with the predicted $[\text{Fe}/\text{H}]_G$ values color-coded. The stars in the metal-poor tail of the MDF lie in close proximity of the Galactic plane, while the high-latitude halo sample and the Magellanic Clouds contribute with the most metal-poor stars. A positive gradient in the mean metallicity from the halo towards the center of the bulge is also apparent. Notably, the metallicities of the stars constituting the metal-rich tail are the ones that show the best correlation between our predictions and those from both DR2 and IB20 (albeit with an offset in the former case; c.f. the left and middle panels of Fig. 9). This is probably due to the fact that the training data sets used for fitting the latter two relations comprise similar objects from the Solar neighborhood.

The distributions in Fig. 9, and thus their modes are dominated by halo objects. To obtain simple estimates of the median halo metallicity, we cross-matched our sample with the RR Lyrae stars in the bulge and Magellanic Clouds listed in the OCVS, and excluded all matches within $1''$, resulting in a reduced sample of ~ 33000 objects. Since the OCVS is a highly complete catalog, this simple approach leaves us with only a small contamination of disk RRab stars, which does not significantly affect the accuracy in terms of estimating the mode of the halo RRab MDF. To measure the latter, we computed kernel density estimates of the MDFs and determined their maxima. Our estimates of the median $[\text{Fe}/\text{H}]$ of the halo obtained in this way is -1.67 dex using our predictive model, -1.43 dex using the IB20 formula, and -1.04 dex according to the DR2 RR Lyrae catalog. Our results are in good agreement with pertinent studies of the inner halo MDF using main-sequence stars by An et al. (2013) and Youakim et al. (2020), while the IB20 and DR2 values both show significant positive bias.

4.2. K_s -band photometric metallicities of the VVV inner bulge RRab stars

The K_s -band predictive model ensemble was deployed on the catalog of RRab stars in the inner bulge area of the VVV survey discovered by Dékány & Grebel (2020, D20). These objects have not been detected by the OGLE survey covering the same area due to extremely high interstellar extinction. Our target data set comprises the public K_s -band light curves of all 4446 RRab stars in the D20 catalog, processed according to Sect. 2. The resulting photometric parameters and metallicity estimates are displayed in Table 5.

The right panel of Figure 11 shows the resulting metallicity distribution of the inner bulge RRab stars, compared to the distribution of the D21 I -band $[\text{Fe}/\text{H}]$ estimates of our development data set. The two distributions are very similar with their modes at -1.35 and -1.36 dex, respectively, showing a smooth continuation of the bulge MDF toward low Galactic latitudes. The figure also indicates the accuracy of both our model ensemble developed in this study, and

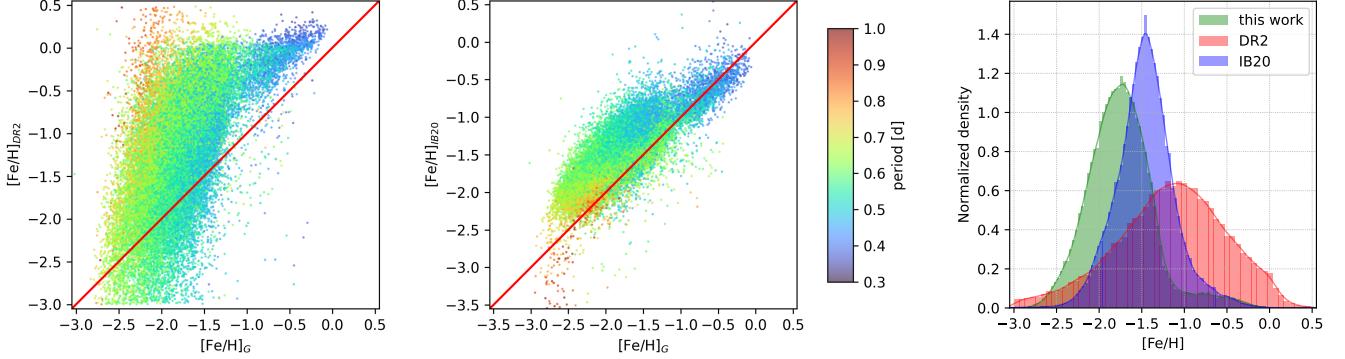


Figure 9. Photometric metallicities predicted by our model ensemble ($[Fe/H]_G$) for the G -band target data set, plotted against the corresponding metallicity estimates from the Gaia DR2 RR Lyrae catalog ($[Fe/H]_{DR2}$, left), and by the IB20 formula ($[Fe/H]_{IB20}$, middle). The pulsation periods are color-coded, the red line denotes the identity function. Right: histograms and corresponding kernel density estimates of the same data set as in the left and middle panels. Green: $[Fe/H]_G$ estimates from this work, blue: IB20 metallicities, red: $[Fe/H]$ values from the DR2 RR Lyrae catalog.

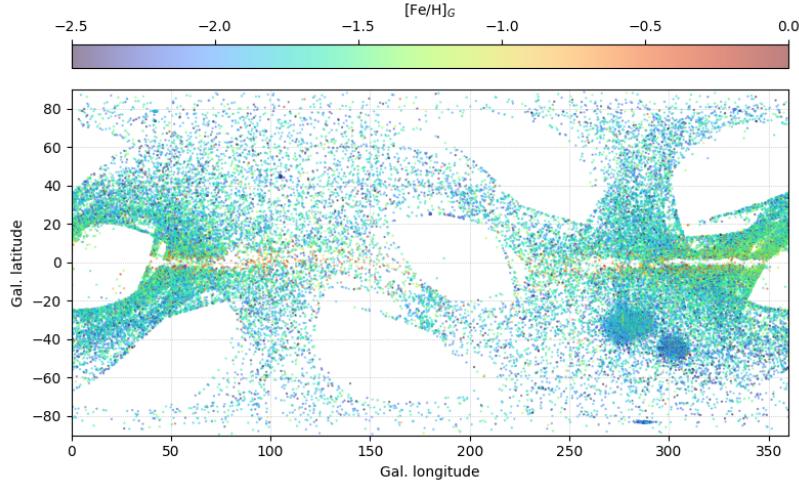


Figure 10. Celestial distribution of the target data set of our Gaia G -band photometric metallicity estimator in the Galactic coordinate system. The predicted $[Fe/H]_G$ values are color-coded.

the D20 deep-learned RRab light-curve classifier. We can observe minor differences between the two MDFs in their metal-rich and metal-poor tails. The MDF of the development set shows a very slightly enhanced metal-poor tail. Based on the OCVS only, a similar difference between the inner and outer bulge's MDF was noted by D21. The MDF of the D20 inner bulge sample also features a slightly more enhanced metal-rich tail, which can be expected, since it covers lower Galactic latitudes, where the relative fraction of metal-rich stars of the thick disk is higher. Figure 12 shows the celestial distribution of the target data set with the $[Fe/H]_{K_s}$ predictions color-coded. Around the Galactic equator, the distribution is dominated by incompleteness due to thick interstellar dust that remains impenetrable for the VISTA telescope even at two microns. The few stars detected around the Galactic Center sight-line are in fact mostly metal-rich stars in the foreground disk. We note that the linear patterns on the northern side are gaps between OGLE fields, where RRab stars that remained previously undiscovered by OGLE were detected in VVV data. The slight overdensity at $(l, b) \sim (9, -1)$ is due to an extinction window where faint metal-rich disk RRab stars behind

Table 5. K_s -band photometric metallicity estimates and basic light-curve attributes of the VVV bulge RRab stars discovered by Dékány & Grebel (2020)

ID ^a	R.A. ^b [hms]	DEC. ^b [dms]	[Fe/H] _{K_s} ^c	$\sigma(\text{[Fe/H]}_{K_s})^d$	$\langle K_s \rangle$	$N_{ep.}$	Period [d]	$A_{tot.}$	C_φ	S/N
7	17:05:27.27	-34:33:50.9	-1.27	0.07	14.657	117	0.643188	0.171	0.939	78.5
27	17:06:58.67	-34:59:07.7	-0.40	0.05	15.538	64	0.440602	0.245	0.938	56.8
36	17:07:41.14	-34:18:33.7	-1.48	0.05	14.421	64	0.769114	0.157	0.933	82.2
64	17:08:58.02	-33:40:19.8	-1.81	0.12	14.808	65	0.737811	0.143	0.915	46.3
108	17:10:26.66	-35:14:50.0	-1.30	0.02	14.573	65	0.632416	0.181	0.924	91.8

^aIdentifier from Dékány & Grebel (2020).

^bCoordinates of J2000.0 epoch.

^cMean of predictions from a model ensemble.

^dStandard deviation of predictions from a model ensemble.

NOTE—This table is available in its entirety in machine-readable form.

the bulge are detected. Together with the OCVS RRab stars (for which the I -band [Fe/H] estimates in Table 1 are preferred), this sample provides an unbiased, homogeneous sample of photometric metallicities.

Finally, we compare our results with another method for metallicity estimation from K_s -band light curves, that was published by Hajdu et al. (2018, H18). Following an approach similar to our study, they developed a fully connected neural network to predict the [Fe/H] of RRab stars from a 4-parameter representation of their light curves. In the latter, the phase-folded light curves were fitted with a linear combination of 4 eigenvectors obtained from a principal component analysis, and the neural network made its predictions from these 4 regression coefficients. The model was trained on I -band photometric metallicities obtained by the formula of Smolec (2005). Since the latter has a positive bias with respect to our D21 formula (see Dékány et al. 2021), we can expect that the H18 model will also output systematically higher predictions compared to our LSTM neural network. We verify this by directly comparing our [Fe/H] _{K_s} predictions to those computed by the H18 model for the target data set, shown in the left panel of Fig. 11. The [Fe/H]_{H18} estimates have a positive bias of 0.4 dex on average, with a significant structure in the residual. The inner bulge MDF obtained by the H18 model is shown with blue in the right panel of Fig. 11. Similarly to the G -band DR2 and IB20 estimates, the metal-rich tail of the distribution is not recovered, due to its blending with the distribution’s overestimated mode. The MDF is also wider due to the significantly higher scatter of the H18 predictions.

5. SUMMARY AND CONCLUSIONS

In this study, we have provided a new method for the estimation of the metallicity of RRab stars from their Gaia G -band and near-infrared K_s -band light curves. In our approach, we have leveraged deep learning to accurately transfer a carefully calibrated photometric [Fe/H] prediction formula to additional wavebands at the cost of minimal additional noise. To achieve this, we used the I -band D21 formula, which had been calibrated to HDS metallicities, to compute photometric metallicities for a large number of objects that also have light curves in the G and/or K_s bands. Subsequently, these metallicities served as response variables in separate regression problems of the G and K_s light curves, solved by state-of-the-art recurrent neural networks. The resulting model ensembles have very high predictive performance measured by cross-validation, with a (weighted) mean absolute error of only 0.1 dex, and are also able to recover the metallicity distributions of the validation data sets without bias, demonstrating that our models generalize well to new data.

Our results were compared with predictions using various other methods from the literature, and found that earlier methods generally overestimate the metallicity, resulting in positive biases of ≥ 0.4 dex ≥ 0.25 dex in the mode of the bulge’s and the halo’s MDFs, respectively. These conclusions are in accordance with our similar comparisons made with our I -band [Fe/H] prediction formula. We found that the [Fe/H] values provided with the Gaia DR2 RR Lyrae catalog are particularly discrepant. We suspect that these disagreements stem from multiple sources, including biases

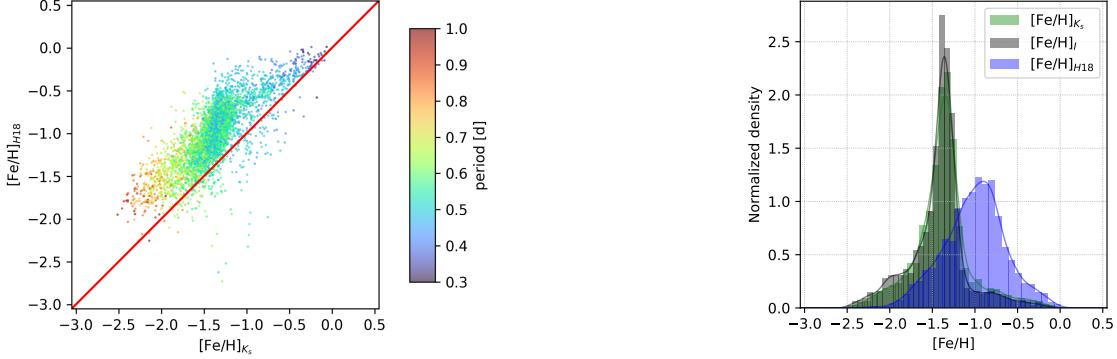


Figure 11. Left: photometric metallicities predicted by our model ensemble ($[Fe/H]_{K_s}$) for the K_s -band target data set, plotted against metallicity estimates by the H18 method ($[Fe/H]_{H18}$). The pulsation periods are color-coded, the red line denotes the identity function. Right: histograms and corresponding kernel density estimates of $[Fe/H]_{K_s}$ (green) and $[Fe/H]_{H18}$ (blue) values for the target data set, and $[Fe/H]_I$ values for the development data set (gray).

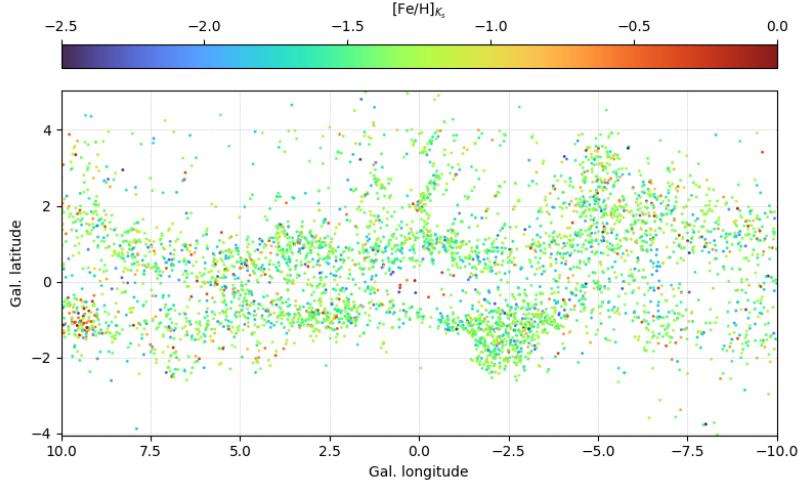


Figure 12. Celestial distribution of the target data set of our Gaia K_s -band photometric metallicity estimator in the Galactic coordinate system. The predicted $[Fe/H]_{K_s}$ values are color-coded.

in regressions due to small and imbalanced calibration data sets, in the transformations of Fourier parameters between wavebands, and in the low-resolution spectroscopic determinations of the metallicity used for the calibration of most of the earlier methods.

We deployed our predictive models to compute the $[Fe/H]$ for numerous Galactic RRab stars observed by the OGLE, Gaia, and VVV surveys. The resulting catalogs of photometric metallicities are provided in the electronic versions of Tables 1, 4, and 5, and can be directly incorporated in various future studies of Galactic archeology. We also make our source code publicly available online⁴, in order to facilitate the further deployment of the neural networks, as well as to allow their eventual re-training and further optimization on other data sets.

The approach of metallicity estimation presented in this paper should be easily applicable in the future to other surveys as well, provided that large enough data sets with well-sampled light curves in multiple photometric wavebands

⁴ The source code and data can be downloaded from https://github.com/idekany/rnl_feh_nn

are available for training the neural networks. We also anticipate that the upcoming Gaia data releases will have improved light-curve sampling and accuracy, and they will include many more RR Lyrae stars, allowing the future improvement of our models' predictive performance. Upon the availability of these data releases, the timely update of our method will be straightforward by retraining our published neural networks on the new data sets.

Our photometric prediction method makes it straightforward to obtain consistent metallicities for very large combined samples of RRab stars, spanning across the footprints of various sky surveys conducted in different wavebands. The resulting homogeneous metallicity estimates, combined with astrometric or photometric distances, make RRab stars outstanding chemo-dynamical tracer objects in the era of data-driven astronomy.

The authors were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 138713538 – SFB 881 (“The Milky Way System”, subproject A03).

Software: numpy (Harris et al. 2020), scipy (Virtanen et al. 2020), scikit-learn (Pedregosa et al. 2011), TensorFlow (Abadi et al. 2015)

REFERENCES

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. <https://www.tensorflow.org/>
- An, D., Beers, T. C., Johnson, J. A., et al. 2013, ApJ, 763, 65, doi: [10.1088/0004-637X/763/1/65](https://doi.org/10.1088/0004-637X/763/1/65)
- Andrievsky, S., Wallerstein, G., Korotin, S., et al. 2018, Publications of the Astronomical Society of the Pacific, 130, 024201, doi: [10.1088/1538-3873/aa9783](https://doi.org/10.1088/1538-3873/aa9783)
- Bhardwaj, A. 2022, arXiv e-prints, arXiv:2202.06982. <https://arxiv.org/abs/2202.06982>
- Chadid, M., Sneden, C., & Preston, G. W. 2017, The Astrophysical Journal, 835, 187, doi: [10.3847/1538-4357/835/2/187](https://doi.org/10.3847/1538-4357/835/2/187)
- Clementini, G., Carretta, E., Gratton, R., et al. 1995, The Astronomical Journal, 110, 2319, doi: [10.1086/117692](https://doi.org/10.1086/117692)
- Clementini, G., Ripepi, V., Molinaro, R., et al. 2019, Astronomy & Astrophysics, 622, A60, doi: [10.1051/0004-6361/201833374](https://doi.org/10.1051/0004-6361/201833374)
- Crestani, J., Fabrizio, M., Braga, V. F., et al. 2021, ApJ, 908, 20, doi: [10.3847/1538-4357/abd183](https://doi.org/10.3847/1538-4357/abd183)
- Dékány, I., Grebel, E. K., & Pojmański, G. 2021, ApJ, 920, 33, doi: [10.3847/1538-4357/ac106f](https://doi.org/10.3847/1538-4357/ac106f)
- Dékány, I., & Grebel, E. K. 2020, The Astrophysical Journal, 898, 46, doi: [10.3847/1538-4357/ab9d87](https://doi.org/10.3847/1538-4357/ab9d87)
- Dékány, I., Hajdu, G., Grebel, E. K., et al. 2018, The Astrophysical Journal, 857, 54, doi: [10.3847/1538-4357/aab4fa](https://doi.org/10.3847/1538-4357/aab4fa)
- Emerson, J. P., Irwin, M. J., Lewis, J., et al. 2004, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5493, Optimizing Scientific Return for Astronomy through Information Technologies, ed. P. J. Quinn & A. Bridger, 401–410, doi: [10.1117/12.551582](https://doi.org/10.1117/12.551582)
- Fernley, J., & Barnes, T. G. 1996, Astronomy & Astrophysics, 312, 957. <http://adsabs.harvard.edu/abs/1996A%26A...312..957F>
- For, B.-Q., Sneden, C., & Preston, G. W. 2011, The Astrophysical Journal Supplement Series, 197, 29, doi: [10.1088/0067-0049/197/2/29](https://doi.org/10.1088/0067-0049/197/2/29)
- Govea, J., Gomez, T., Preston, G. W., & Sneden, C. 2014, The Astrophysical Journal, 782, 59, doi: [10.1088/0004-637X/782/2/59](https://doi.org/10.1088/0004-637X/782/2/59)
- Hajdu, G., Dékány, I., Catelan, M., & Grebel, E. K. 2020, Experimental Astronomy, 49, 217, doi: [10.1007/s10686-020-09661-0](https://doi.org/10.1007/s10686-020-09661-0)
- Hajdu, G., Dékány, I., Catelan, M., Grebel, E. K., & Jurcsik, J. 2018, The Astrophysical Journal, 857, 55, doi: [10.3847/1538-4357/aab4fd](https://doi.org/10.3847/1538-4357/aab4fd)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)
- Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
- Houdt, G. V., Mosquera, C., & Nápoles, G. 2020, Artificial Intelligence Review, 1
- Iorio, G., & Belokurov, V. 2020, arXiv e-prints, 2008, arXiv:2008.02280. <http://adsabs.harvard.edu/abs/2020arXiv200802280I>
- Jurcsik, J., Hajdu, G., Dékány, I., et al. 2018, MNRAS, 475, 4208, doi: [10.1093/mnras/sty112](https://doi.org/10.1093/mnras/sty112)

- 541 Jurcsik, J., & Kovács, G. 1996, *Astronomy & Astrophysics*,
 542 312, 111
- 543 Kingma, D. P., & Ba, J. 2014, arXiv e-prints,
 544 arXiv:1412.6980. <https://arxiv.org/abs/1412.6980>
- 545 Lambert, D. L., Heath, J. E., Lemke, M., & Drake, J. 1996,
 546 *The Astrophysical Journal Supplement Series*, 103, 183,
 547 doi: [10.1086/192274](https://doi.org/10.1086/192274)
- 548 Layden, A. C. 1994, *The Astronomical Journal*, 108, 1016,
 549 doi: [10.1086/117132](https://doi.org/10.1086/117132)
- 550 Liu, S., Zhao, G., Chen, Y.-Q., Takeda, Y., & Honda, S.
 551 2013, *Research in Astronomy and Astrophysics*, 13, 1307,
 552 doi: [10.1088/1674-4527/13/11/003](https://doi.org/10.1088/1674-4527/13/11/003)
- 553 Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, \na,
 554 15, 433, doi: [10.1016/j.newast.2009.12.002](https://doi.org/10.1016/j.newast.2009.12.002)
- 555 Mullen, J. P., Marengo, M., Martínez-Vázquez, C. E., et al.
 556 2021, *ApJ*, 912, 144, doi: [10.3847/1538-4357/abefd4](https://doi.org/10.3847/1538-4357/abefd4)
- 557 Nemec, J. M., Cohen, J. G., Ripepi, V., et al. 2013, *The
 558 Astrophysical Journal*, 773, 181,
 559 doi: [10.1088/0004-637X/773/2/181](https://doi.org/10.1088/0004-637X/773/2/181)
- 560 Ngeow, C.-C., Yu, P.-C., Bellm, E., et al. 2016, *The
 561 Astrophysical Journal Supplement Series*, 227, 30,
 562 doi: [10.3847/1538-4365/227/2/30](https://doi.org/10.3847/1538-4365/227/2/30)
- 563 Pancino, E., Britavskiy, N., Romano, D., et al. 2015,
 564 *Monthly Notices of the Royal Astronomical Society*, 447,
 565 2404, doi: [10.1093/mnras/stu2616](https://doi.org/10.1093/mnras/stu2616)
- 566 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011,
 567 *Journal of Machine Learning Research*, 12, 2825
- 568 Prudil, Z., & Skarka, M. 2017, *MNRAS*, 466, 2602,
 569 doi: [10.1093/mnras/stw3231](https://doi.org/10.1093/mnras/stw3231)
- 570 Savino, A., Koch, A., Prudil, Z., Kunder, A., & Smolec, R.
 571 2020, *Astronomy & Astrophysics*, 641, A96,
 572 doi: [10.1051/0004-6361/202038305](https://doi.org/10.1051/0004-6361/202038305)
- 573 Skowron, D. M., Soszyński, I., Udalski, A., et al. 2016, *Acta
 574 Astronomica*, 66, 269.
<http://adsabs.harvard.edu/abs/2016AcA....66..269S>
- 575 Smolec, R. 2005, *Acta Astronomica*, 55, 59
- 576 Smolec, R. 2016, in 37th Meeting of the Polish
 577 Astronomical Society, ed. A. Różańska & M. Bejger,
 578 Vol. 3, 22–25. <https://arxiv.org/abs/1603.01252>
- 579 Sneden, C., Preston, G. W., Chadid, M., & Adamów, M.
 580 2017, *\apj*, 848, 68, doi: [10.3847/1538-4357/aa8b10](https://doi.org/10.3847/1538-4357/aa8b10)
- 581 Soszyński, I., Udalski, A., Wrona, M., et al. 2019, *Acta
 582 Astronomica*, 69, 321, doi: [10.32023/0001-5237/69.4.2](https://doi.org/10.32023/0001-5237/69.4.2)
- 583 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., &
 584 Salakhutdinov, R. 2014, *Journal of Machine Learning
 585 Research*, 15, 1929.
<http://jmlr.org/papers/v15/srivastava14a.html>
- 586 Tanakul, N., & Sarajedini, A. 2018, *MNRAS*, 478, 4590,
 587 doi: [10.1093/mnras/sty1311](https://doi.org/10.1093/mnras/sty1311)
- 588 Udalski, A., Szymański, M. K., & Szymański, G. 2015,
 589 *Acta Astronomica*, 65, 1.
<http://adsabs.harvard.edu/abs/2015AcA....65....1U>
- 590 Vida, K., Bódi, A., Szklenár, T., & Seli, B. 2021, *A&A*,
 591 652, A107, doi: [10.1051/0004-6361/202141068](https://doi.org/10.1051/0004-6361/202141068)
- 592 Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020,
 593 *Nature Methods*, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- 594 Youakim, K., Starkenburg, E., Martin, N. F., et al. 2020,
 595 *MNRAS*, 492, 4986, doi: [10.1093/mnras/stz3619](https://doi.org/10.1093/mnras/stz3619)
- 596 Yu, Y., Si, X., Hu, C., & Zhang, J. 2019, *Neural
 597 Computation*, 31, 1235, doi: [10.1162/neco_a_01199](https://doi.org/10.1162/neco_a_01199)
- 598
- 599
- 600