Determination of metallicities of red giant stars using machine learning techniques applied to the narrow and broadband photometry of the S-PLUS survey

F. Molina-Jorquera[1,★], G. Damke[2], D. Fernández-Olivares[1], M. Jaque-Arancibia[1,3], F. Almeida-Fernandes[4,5], S. Daflon[6], W. Schoenell[7], T. Ribeiro[8], A. Kanaan[9], S. Torres-Flores[1], and C. Mendes de Oliveira[4]

[1] Departamento de Astronomía, Universidad de La Serena, Av. J. Cisternas 1200 N, 1720236 La Serena, Chile
[2] Cerro Tololo Inter-American Observatory/NSF's NOIRLab, Casilla 603, La Serena, Chile
[3] Instituto Multidisciplinario de Investigación y Postgrado, Universidad de La Serena, Av. R. Bitrán 1305, 1720256 La Serena, Chile

5-9

A&A,691,A144(2024)

Reporter: Baokun Sun

Outline

- 1. Introduction
- 2. Data
- 3. Artificial neural networks
- 4. Giant or dwarf classification
- 5. Determination of stellar metallicities
- 6. Derivation of metallicities of the whole S-PLUS stellar catalogue
- 7. Conclusions

Introduction

- The study of substructures in the Milky Way provides important clues to understanding the formation of the Galaxy in a cosmological context.
- The need to obtain photometric information sensitive to stellar metallicity has triggered the advent of large-area surveys that utilise narrow-band filters centred on relevant spectral features.

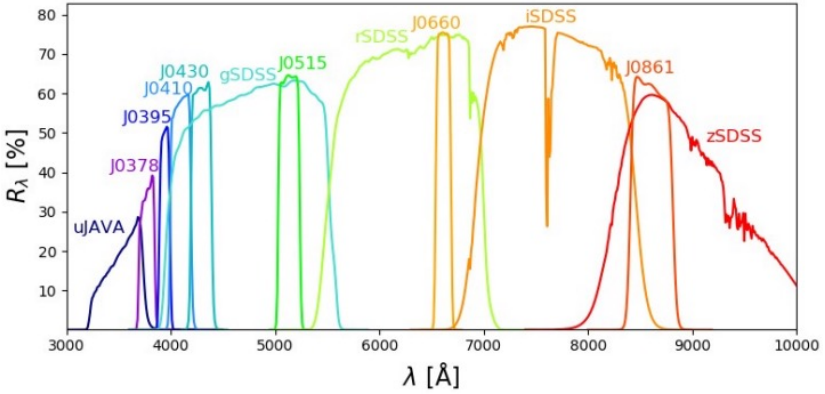

Fig. 1. Total transmission curves of the Javalambre filter system. On the vertical axis is the efficiency; the horizontal axis shows the wavelength in Angstroms. The efficiency includes contributions from filter transmission, atmosphere transmission, CCD efficiency, and primary mirror reflectivity curves. This photometric filter system is composed of seven narrow-band filters (J0378, J0395, J0410, J0430, J0515, J0660, and J0861) and five broadband filters (u, g, r, i, and z) (source: S-PLUS, 2019).

Table 3. J-PLUS filter system.

Filter	Central Wavelength (Å)	$FWHM$ (Å)	Comments
u	3485	508	(a)
$J0378$	3785	168	[O II]; (a)
$J0395$	3950	100	Ca H+K
$J0410$	4100	200	H_δ
$J0430$	4300	200	G band
g	4803	1409	(b)
$J0515$	5150	200	Mg b triplet
r	6254	1388	(b)
$J0660$	6600	138	Hα; (a)
i	7668	1535	(b)
$J0861$	8610	400	Ca triplet
z	9114	1409	(b)

Notes. (a) In common with J-PAS; (b) SDSS. Rest-frame key spectral features matching the location of narrow- and intermediate-band filters are also indicated.

Data

- In this work, we have used the photometric catalogues of S-PLUS Data Release 3 (S-PLUS DR3). The S-PLUS DR3 has over 50 million detections (Almeida-Fernandes 2020). The DR3 catalogues cover an area of the sky corresponding to 2214 deg2 distributed in 1107 fields.
- They obtained E(B−V) from the 2D Schlegel et al. (1998) dust maps, recalibrated by Schlafly & Finkbeiner (2011).
- Their spectroscopic data come from APOGEE DR17. We limited the temperature in our sample to $3500 \lesssim$ Teff (K) $\lesssim 7000$ to avoid systematic problems (Abdurro'uf et al. 2022), and a signal-to-noise ratio of >20.

Training, validation, and testing catalogue

Signal-to-noise s2n_[Filter]_PSTOTAL > 20 in all the S-PLUS bands.

The stars with flags PhotoFlag_[Filter] ≠ 0 – which excludes objects with contamination by neighboring sources, saturated pixels, truncated objects, etc. – were also eliminated.

Leaving a catalogue with 6433 stars, with $-0.01 \lesssim \log g \lesssim 5.19$, $-2.45 \lesssim [M/H] \lesssim 0.50$ and $3580 \lesssim \mathrm{Teff}\ (\mathrm{K}) \lesssim 6890$.

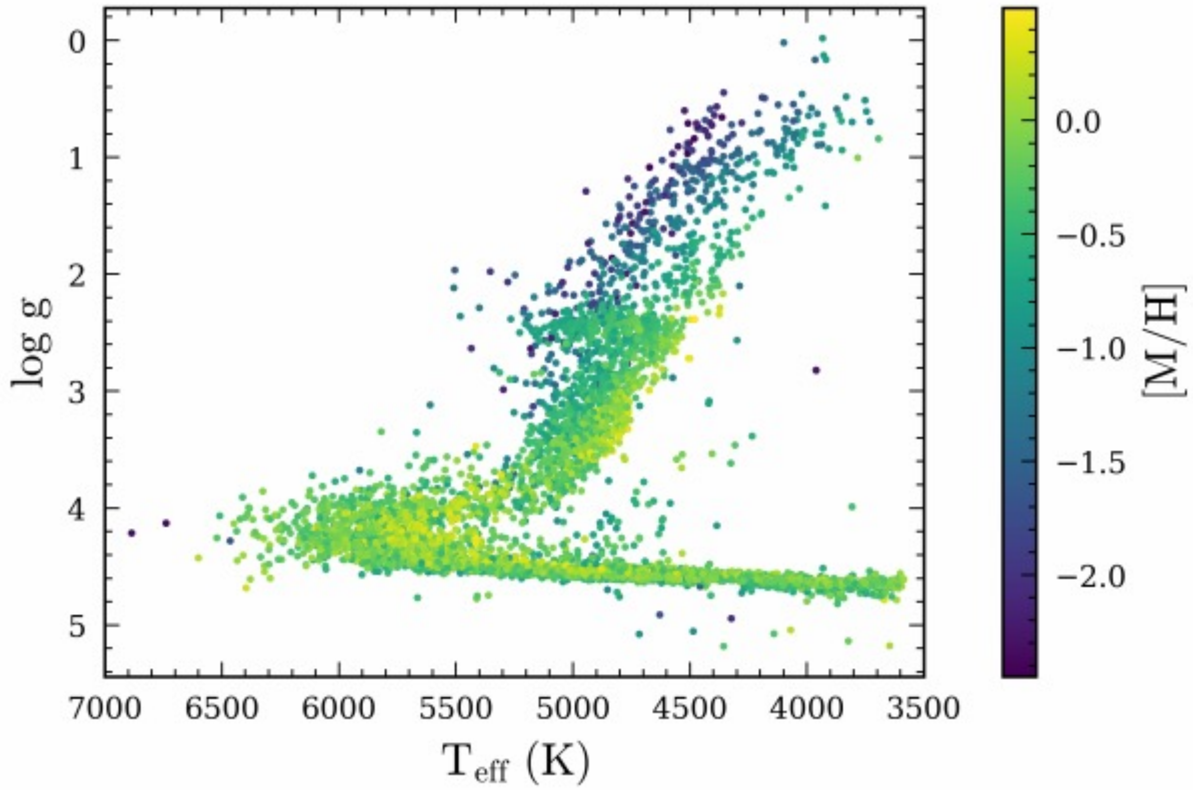

Fig. 2. Distribution of the 6433 stars resulting from the cross-match between S-PLUS DR3 and APOGEE DR17 for the TVT catalogues in the Kiel diagram. The colour map represents spectroscopic metallicities.

Defining labels for giants and dwarf stars

- They used the agglomerative hierarchical clustering algorithm from scikit-learn to identify two groups in the (Teff, log g) plane.

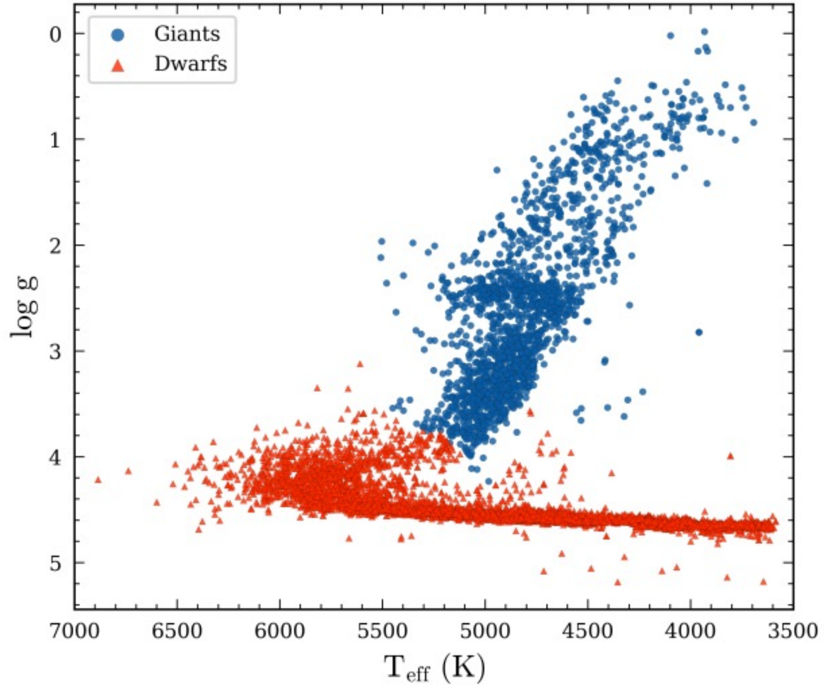

Fig. 3. Distribution of dwarf (red triangles) and giant (blue circles) stars determined by the hierarchical clustering algorithm in the ($T_{\rm eff}$, log g) plane.

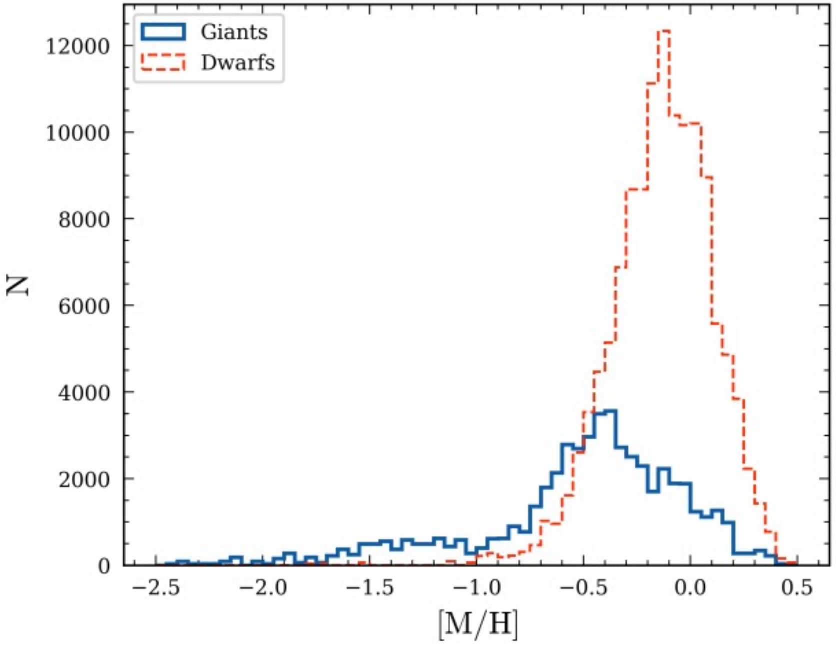

Fig. 4. Distribution of metallicities of giant (thick solid blue line) and dwarf (thin dashed red line) stars in the TVT set. Spectroscopic metallicities of giant stars have standard errors of $\sigma \sim 0.02$ dex and dwarfs of $\sigma \sim 0.09$ dex.

Data augmentation and Catalogue homogenisation

- We performed the data augmentation by adding a random Δmag sampled randomly from a Gaussian distribution with a standard deviation equal to the reported photometric uncertainty to each S-PLUS band.
- We cut the complete sample in $-1.6 < [M/H] < 0.4$ (since outside this range, the number of stars is much lower). The resultant sample was divided into bins of width equal to 0.05 dex, and 245 stars were randomly taken from each bin.
- This way, we obtained a homogeneous sample in metallicity with 9800 stars.

Input features

- We considered all colours determined from each narrow-band filter and the broadband filter it overlaps. These colours are: (J0378 - u), (J0395 - g), (J0410 - g), (J0430 - g), (J0515 - g), (J0660 - r), and (J0861 -z).
- We considered the colours determined from all possible combinations between broad-band filters. These colours are: (u - g), (u - r), (u - i), (u - z), (g - r), (g - i), (g - z), (r - i), (r -z), and (i-z).
- The colour combinations used to train the networks have between three and seven colours defined using one narrow-band colour index and one or two colours defined only from wide-band colours.

Giant or dwarf classification

- ANN-C3324-2 was trained with the combination of colours (J0378 - u), (J0395 - g), (J0410 - g), (J0515 - g), (J0660 - r), (g - z), and (r - i).

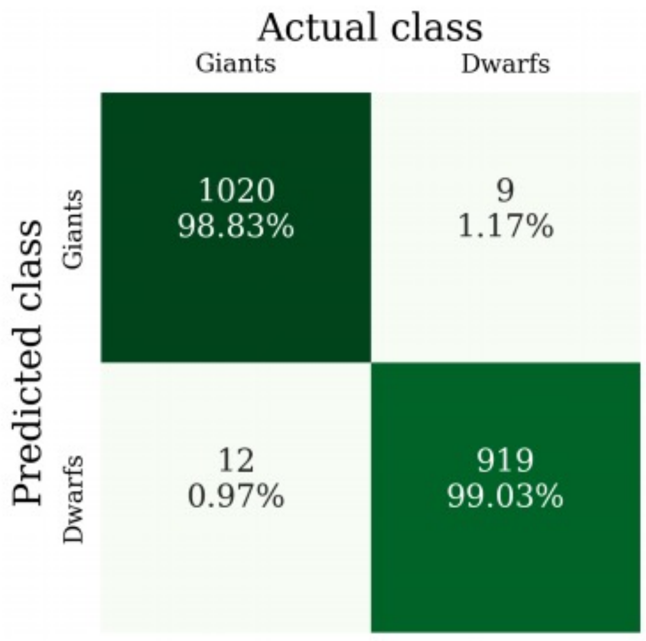

Fig. 6. Confusion matrix of the ANN-C3324-2 predictions for the test set, showing the number of actual and predicted stars in each category.

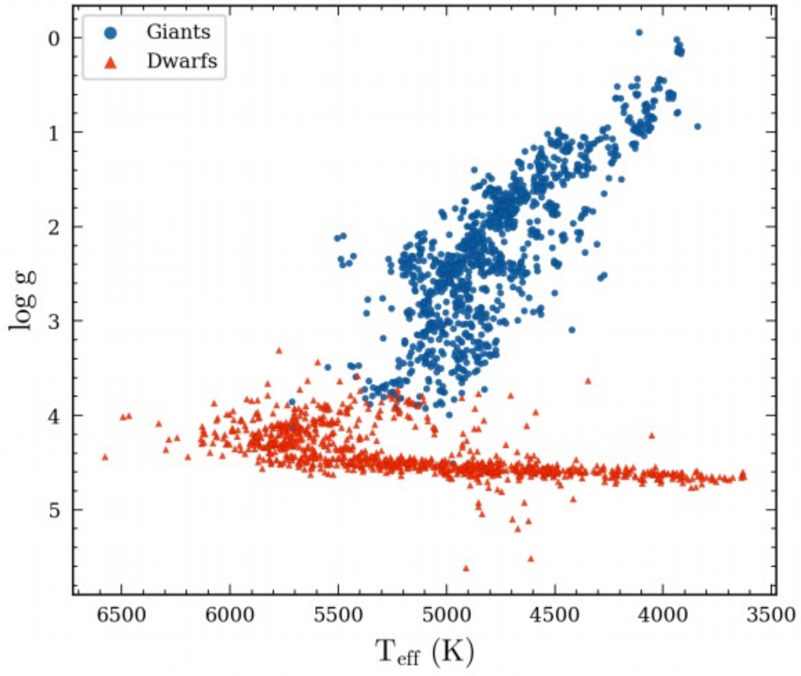

Fig. 7. Kiel diagram showing the predictions of ANN-C3324-2 for the test set. Stars classified by the ANN as dwarfs are shown as red triangles, and those classified as giants are shown as blue circles.

Determination of stellar metallicities

The network with the input colour combination 3292 – (J0378 - u), (J0395 - g), (J0410 - g), (J0515 - g), (J0660 - r), (u - g), and (r - z), the first five corresponding to the spectral characteristics [OII], Ca H+K, Hγ, Mg triplet and Hα, respectively – is the one with the best performance with the complete TVT catalogue.

With giant stars, the best performance was the one with the input colour combination 4290 - (J0378 - u), (J0395 - g), (J0430 - g), (J0515 - g), (J0660 - r), (J0861 - z), (u - r), and (r - z), the first five corresponding to the spectral characteristics [OII], Ca H+K, G band, Mg triplet, Hα and Ca triplet, respectively.

In both cases, Ca H+K and Mg triplet characteristics are present, which are sensitive to metallicity.

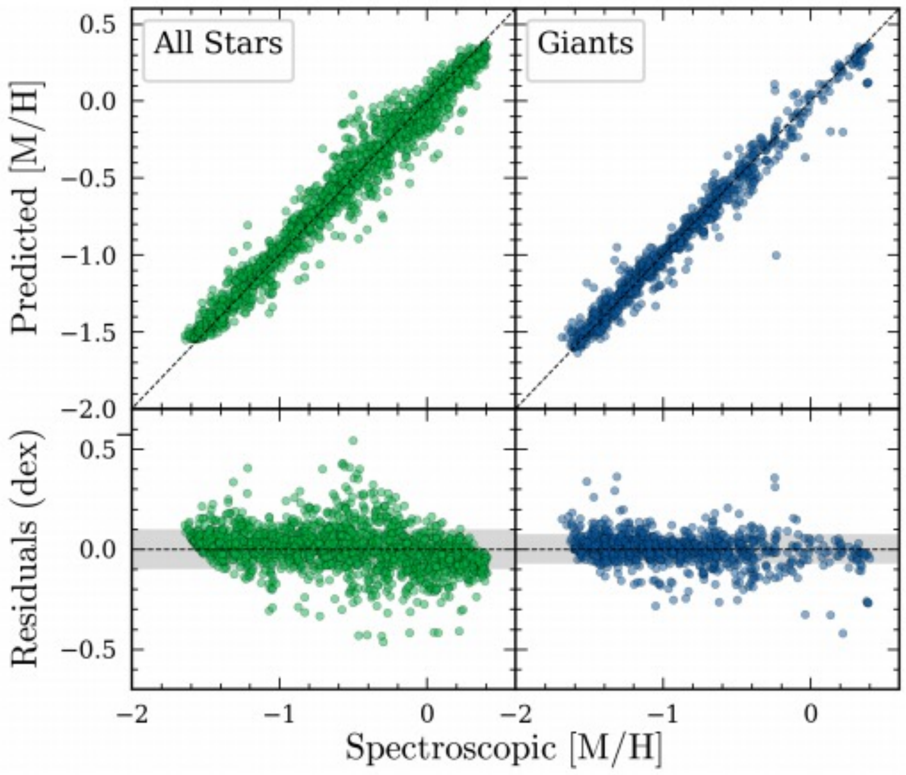

Fig. 10. Results of the inference of the ANN-R3292-2 for all stars (left, green) and the ANN-R4290-2 only for giants (right, blue) for the test set. The upper panels show the predictions vs the spectroscopic metallicities, and the lower panels show the residual for the metallicities. The grey area indicates a range of ±1 in standard deviation, which is 0.1 (dex) for all stars and 0.07 (dex) for giants only. The residuals were calculated as the difference between the predictions and the spectroscopic metallicities.

Derivation of metallicities for the whole S-PLUS stellar catalogue

- SNR > 20 on each filter and flags = 0
- CLASS_STAR > 0.75
- Cross-match with Bailer-Jones et al. (2021)
- The distances of 2 729 095 stars from S-PLUS DR3 were obtained.
- 81% of the stars were classified as dwarfs, and the remaining 19% were classified as giants.

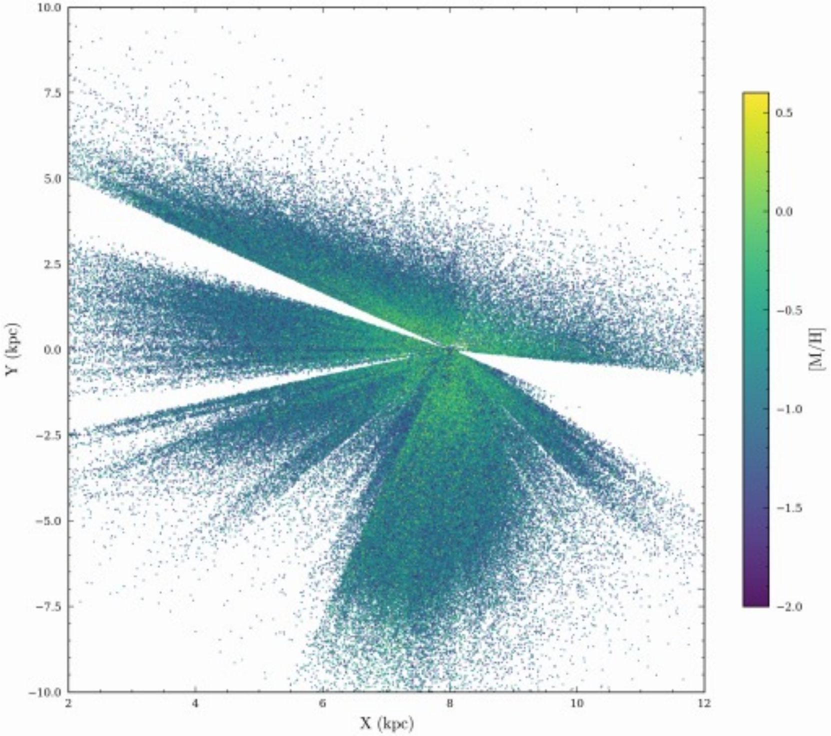

Fig. 14. Metallicity distribution of S-PLUS giant stars in the (X, Y) plane. The colour map indicates the metallicities. The Sun is located at $(X, Y) = (0.8)$.

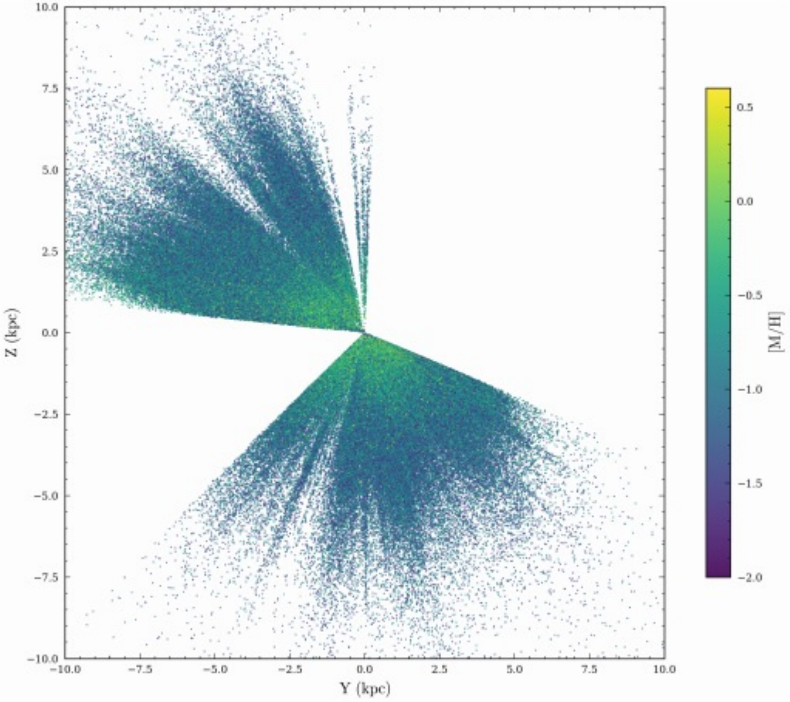

Fig. 13. Metallicity distribution of S-PLUS giant stars in the (Y,Z) plane. The colour map indicates the metallicities. The Sun is located at the point $(Y,Z) = (0,0)$.

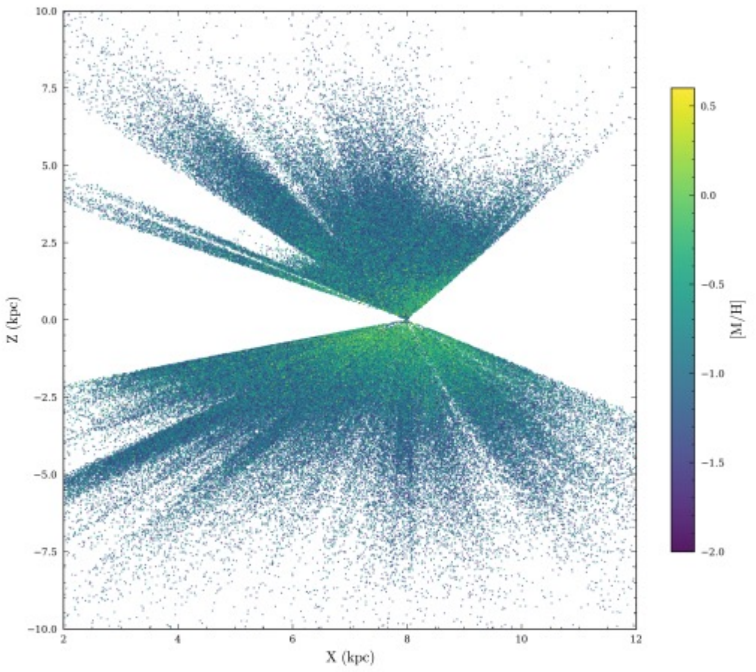

Fig. 12. Metallicity distribution of S-PLUS giant stars in the (X, Z) plane. The colour map indicates the metallicities. The Sun is located at $(X, Z) = (8,0)$.

It is observed that stars become poorer in metals the farther from the Sun they are. Considering that the thin and thick discs of the Milky Way have a scale height of hZ ~ 300 pc and ~900 pc (Juric et al. 2008), respectively, in Figures 12 and 13 metal-poor stars are found in the Galactic halo.

It can also be seen that only the panels of the solar neighbourhood show that the internal disc is more metal-rich than the outer one.

We obtained a vertical gradient of $\Delta[M/H]/\Delta|Z| = -0.087 \pm 0.002$ dex kpc−1 ($7 < R$ (kpc) < 9, $|Z|$ (kpc) < 2) and a radial gradient of $\Delta[M/H]/\Delta R = -0.193 \pm 0.002$ dex kpc−1 ($6 < R$ (kpc) < 10, $|Z|$ (kpc) < 1).

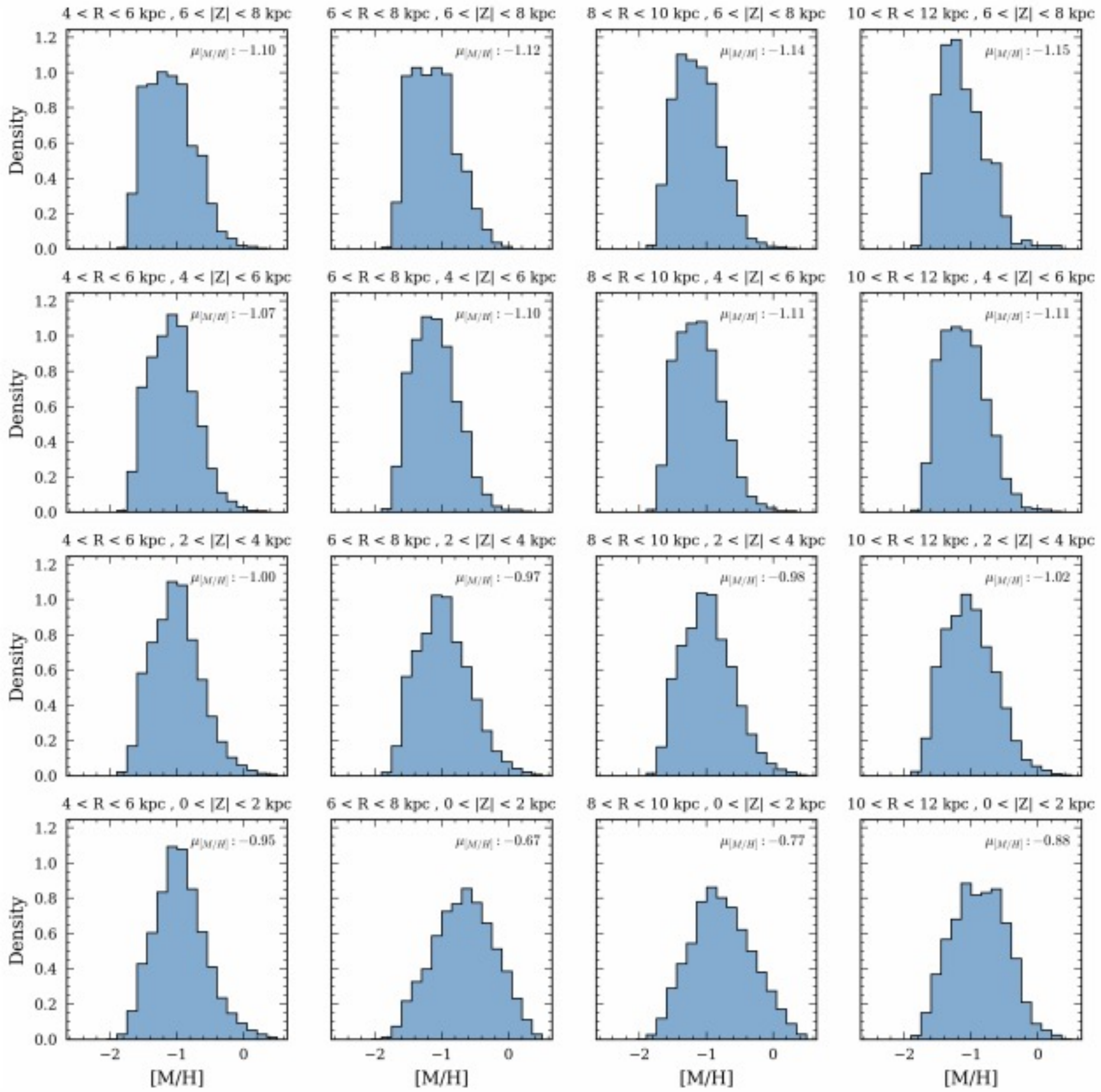

Fig. 15. Probability density of metallicities of giant stars of S-PLUS in different ranges of R and Z normalised to the number of giant stars in every panel.

Conclusion

- We have shown that the application of machine learning methods to derivation the stellar metallicities from a combination of broad and narrow-band photometry can achieve an accuracy comparable to medium-resolution spectroscopy but with the advantage of area coverage.
- The network that presented the lowest MAE over the test set, applied to the determination of photometric metallicities of giant stars, uses the colours (J0378 - u), (J0395 - g), (J0410 - g), (J0515 - g), (J0660 - r), (u - g), and (r - z) as input features, the first six corresponding to the spectral features from [OII], Ca H+K, Hδ, Mg triplet, and Hα, respectively
- This network derived metallicities of stars in the test set with a standard deviation of σgiants ~ 0.07 dex with respect to the spectroscopic metallicities of APOGEE DR17.