

# J-PLUS: Identification of low-metallicity stars with artificial neural networks using SPHINX

D. D. Whitten<sup>1,2</sup>, V. M. Placco<sup>1,2</sup>, T. C. Beers<sup>1,2</sup>, A. L. Chies-Santos<sup>3</sup>, C. Bonatto<sup>3</sup>, J. Varela<sup>4</sup>,  
D. Cristóbal-Hornillos<sup>4</sup>, A. Ederoclite<sup>4</sup>, T. Masseron<sup>5,6</sup>, Y. S. Lee<sup>7</sup>, S. Akras<sup>8</sup>, M. Borges Fernandes<sup>8</sup>,  
J. A. Caballero<sup>9</sup>, A. J. Cenarro<sup>4</sup>, P. Coelho<sup>10</sup>, M. V. Costa-Duarte<sup>10</sup>, S. Daflon<sup>8</sup>, R. A. Dupke<sup>8,13,14</sup>,  
R. Lopes de Oliveira<sup>15,16,11,8</sup>, C. López-Sanjuan<sup>4</sup>, A. Marín-Franch<sup>4</sup>, C. Mendes de Oliveira<sup>10</sup>, M. Moles<sup>12</sup>,  
A. A. Orsi<sup>12</sup>, S. Rossi<sup>10</sup>, L. Sodr e<sup>10</sup>, and H. Vázquez Rami o<sup>12</sup>

<sup>1</sup> Department of Physics, University of Notre Dame, Notre Dame, IN. 46556, USA  
e-mail: dwhitten@nd.edu

<sup>2</sup> JINA Center for the Evolution of the Elements, East Lansing, USA

<sup>3</sup> Departamento de Astronomia, Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

<sup>4</sup> Centro de Estudios de Física del Cosmos de Aragón (CEFCA) – Unidad Asociada al CSIC, Plaza San Juan, Planta 2, 44001 Teruel, Spain

<sup>5</sup> Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain

<sup>6</sup> Departamento de Astrofísica, Universidad de La Laguna, 38206 La Laguna, Tenerife, Spain

<sup>7</sup> Department of Astronomy and Space Science, Chungnam National University, Daejeon 34134, Korea

<sup>8</sup> Observatório Nacional – MCTIC (ON), Rua Gal. José Cristino 77, São Crist v o, 20921-400, Rio de Janeiro, Brazil

<sup>9</sup> Centro de Astrobiología (CSIC-INTA), ESAC, Camino Bajo del Castillo S/N, 28692 Villanueva de la Cañada Madrid, Spain

<sup>10</sup> Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG), Universidade de São Paulo, Rua do Mat o 1226, C. Universit ria, S o Paulo, 05508-090, Brazil

<sup>11</sup> Departamento de Física, Universidade Federal de Sergipe, Av. Marechal Rondon, s/n, 49000-000 S o Crist v o, SE, Brazil

<sup>12</sup> Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan, 1, 44001, Teruel, Spain

<sup>13</sup> Department of Astronomy, University of Michigan, 1085 S. Ave, University Ann Arbor, MI 48109, USA

<sup>14</sup> Department of Phys. & Astronomy, University of Alabama, Gallalee Hall, Tuscaloosa, AL 35401, USA

<sup>15</sup> X-ray Astrophysics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

<sup>16</sup> Department of Physics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA

Received 3 May 2018 / Accepted 5 November 2018

## ABSTRACT

**Context.** We present a new methodology for the estimation of stellar atmospheric parameters from narrow- and intermediate-band photometry of the Javalambre Photometric Local Universe Survey (J-PLUS), and propose a method for target pre-selection of low-metallicity stars for follow-up spectroscopic studies. Photometric metallicity estimates for stars in the globular cluster M15 are determined using this method.

**Aims.** By development of a neural-network-based photometry pipeline, we aim to produce estimates of effective temperature,  $T_{\text{eff}}$ , and metallicity,  $[\text{Fe}/\text{H}]$ , for a large subset of stars in the J-PLUS footprint.

**Methods.** The Stellar Photometric Index Network Explorer, SPHINX, was developed to produce estimates of  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$ , after training on a combination of J-PLUS photometric inputs and synthetic magnitudes computed for medium-resolution ( $R \sim 2000$ ) spectra of the Sloan Digital Sky Survey. This methodology was applied to J-PLUS photometry of the globular cluster M15.

**Results.** Effective temperature estimates made with J-PLUS Early Data Release photometry exhibit low scatter,  $\sigma(T_{\text{eff}}) = 91$  K, over the temperature range  $4500 < T_{\text{eff}} \text{ (K)} < 8500$ . For stars from the J-PLUS First Data Release with  $4500 < T_{\text{eff}} \text{ (K)} < 6200$ ,  $85 \pm 3\%$  of stars known to have  $[\text{Fe}/\text{H}] < -2.0$  are recovered by SPHINX. A mean metallicity of  $[\text{Fe}/\text{H}] = -2.32 \pm 0.01$ , with a residual spread of 0.3 dex, is determined for M15 using J-PLUS photometry of 664 likely cluster members.

**Conclusions.** We confirm the performance of SPHINX within the ranges specified, and verify its utility as a stand-alone tool for photometric estimation of effective temperature and metallicity, and for pre-selection of metal-poor spectroscopic targets.

**Key words.** stars: chemically peculiar – stars: fundamental parameters – stars: abundances – techniques: photometric – methods: data analysis

## 1. Introduction

The chemical properties of individual stars in the Milky Way are crucial in order to develop an understanding of our Galaxy’s chemical evolution and assembly history. In particular, the metallicity distribution function of Galactic halo stars is among the most important observational constraints for cosmological

models (Beers & Christlieb 2005; Salvadori et al. 2010). The comparatively rare stars with metallicity below 1% of the solar value – described in terms of their metal abundance, very metal-poor (VMP;  $[\text{Fe}/\text{H}]^1 < -2.0$ ), extremely metal-poor (EMP;  $[\text{Fe}/\text{H}] < -3.0$ ), and ultra metal-poor (UMP;  $[\text{Fe}/\text{H}] < -4.0$ ) –

<sup>1</sup>  $[\text{A}/\text{B}] \equiv \log_{10}(N_{\text{A}}/N_{\text{B}})_{*} - \log_{10}(N_{\text{A}}/N_{\text{B}})_{\odot}$ .

are expected to include the earliest generations of stars to have formed since the Big Bang. With the exception of mass-transfer binaries and highly evolved late-type giants, these ancient stars retain the chemical signature of their natal environments. Measurement of the chemical abundances of the earliest stars thereby provides a means to study the nucleosynthetic pathways and astrophysical mechanisms that were in operation during the first generations of stars born in the early Universe. One example is the most iron-poor star presently known, SMSS J031300.36-670839.3 ( $[\text{Fe}/\text{H}] \leq -7.1$ ), from which the carbon ( $[\text{C}/\text{Fe}]$ ) and  $[\text{Mg}/\text{Ca}]$  abundance ratios are believed to have originated from a single explosion of a metal-free  $\sim 60 M_{\odot}$ -mass star (Keller et al. 2014).

A substantial fraction of VMP stars exhibit large enhancements of their  $[\text{C}/\text{Fe}]$  ratios, increasing rapidly with declining  $[\text{Fe}/\text{H}]$  (see Yoon et al. 2018 and references therein); these are known collectively as carbon-enhanced metal-poor (CEMP) stars (Beers & Christlieb 2005). Guided by the work of Spite et al. (2013) and Bonifacio et al. (2015), Yoon et al. (2016) explored the complex morphology of CEMP stars in the  $A(\text{C})$ - $[\text{Fe}/\text{H}]$  space<sup>2</sup>. The so-called Yoon–Beers diagram (Fig. 1 of Yoon et al. 2016) provides evidence for multiple progenitors and environments in which different sub-classes of CEMP stars are found. The identification and study of significantly larger samples of CEMP stars is crucial for future studies, as only 300+ of these stars with available high-resolution spectroscopy are currently known.

Obtaining chemical abundances and overall estimates of metallicity is a costly endeavor, however, requiring pre-selection and follow-up spectroscopic analysis for confirmation. While there are now tens of thousands of VMP stars with well-measured (medium-resolution) spectroscopic metallicities, the numbers of known EMP and UMP stars are considerably smaller; in particular,  $\sim 30$  UMP stars have been discovered to date (the compilation as given in Placco et al. 2015, 2016, and Abohalima & Frebel 2018; Starkenburg et al. 2018; Frebel et al. 2019; Aguado et al. 2018).

Wide-field photometry offers an alternative means to probe the chemical characteristics of stars, and presents a method for pre-selection of targets based on their colors. Using broadband photometry from the Sloan Digital Sky Survey (SDSS; York et al. 2000), Ivezić et al. (2008) developed a methodology for estimating metallicity for F- and G-type main-sequence stars ( $5000 < T_{\text{eff}} (\text{K}) < 7000$ ,  $3 < \log g < 5$ ) using polynomial regressions based on de-reddened  $u-g$  and  $g-r$  colors. This approach was found to be effective down to  $[\text{Fe}/\text{H}]$  approximately  $-2.0$ .

Later, An et al. (2013, 2015) used fiducial isochrone fits to SDSS *ugriz* photometry to extend metallicity determinations for main-sequence stars with broadband photometry down to at least  $[\text{Fe}/\text{H}]$  approximately  $-2.5$ . This new threshold is below the peak in the metallicity distribution function of outer-halo population stars ( $[\text{Fe}/\text{H}] = -2.2$ ; Carollo et al. 2007, 2010), leading them to conclude that  $\sim 35$ – $55\%$  of local halo stars belong to this population. Other notable moderate- and narrow-band photometric metallicity estimation techniques have existed for many years, for instance, the Washington (Canterna 1976) and Strömgren (Strömgren 1963, 1964) systems. However, no wide-field large sky-coverage surveys in these systems have been carried out to date. The Canada-France Imaging Survey,

which will cover  $10\,000 \text{ deg}^2$  of the northern sky with a  $u$ -band, demonstrated metallicity sensitivity of  $\sigma[\text{Fe}/\text{H}] = 0.2 \text{ dex}$  down to  $[\text{Fe}/\text{H}]$  approximately  $-2.5$  for dwarf stars ( $3 < \log g < 5$ ) when combined with SDSS and PS1 photometry (Ibata et al. 2017a,b).

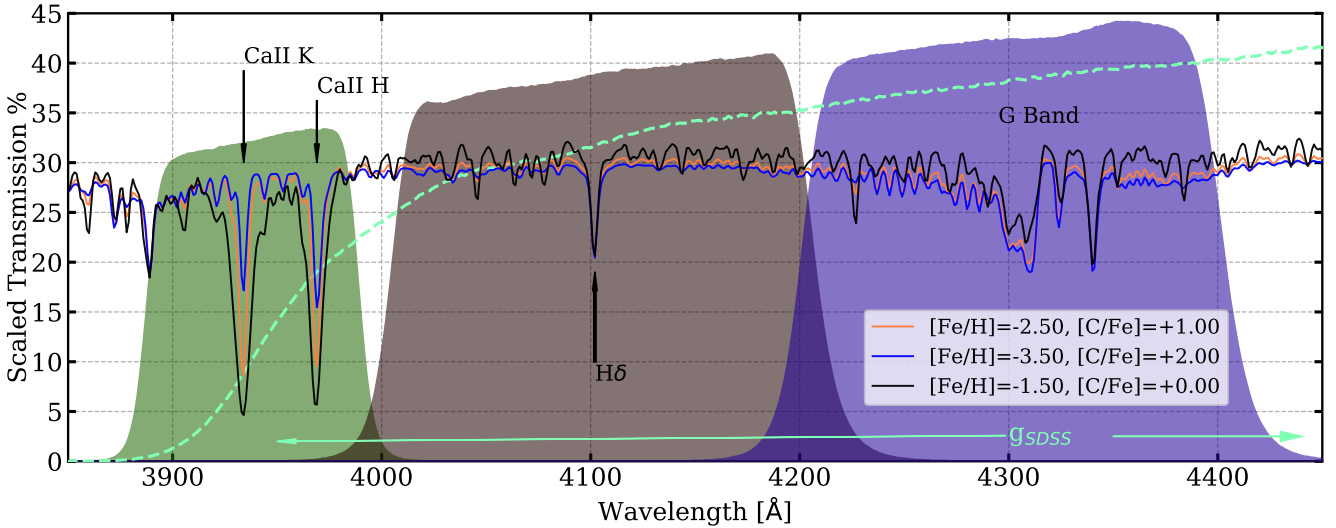
Pre-selection of targets for spectroscopic follow-up can dramatically increase the success rate for the identification of large numbers of metal-poor stars. For example, using all-sky APASS optical, 2MASS near-infrared, and WISE mid-infrared photometry, Schlaufman & Casey (2014) developed an efficient method for selecting bright ( $V < 14$ ) metal-poor candidates based on their lack of molecular absorption near  $4.6 \mu\text{m}$ . This effect was demonstrated to be present in atmospheres of all surface gravities in the effective temperature range  $4500 \leq T_{\text{eff}} (\text{K}) \leq 5500$ . Of these targets,  $32.5\%$  were found to have  $-3.0 < [\text{Fe}/\text{H}] < -2.0$ . The identification of bright metal-poor candidates is of great benefit to high-resolution follow-up observations, for which the acquisition of high-resolution spectra with high signal-to-noise ratio (S/N) can become prohibitively expensive in terms of telescope time.

New multiband photometric surveys build upon previous broadband photometric metallicity determinations by implementation of one or more narrowband filters across the optical and near-infrared spectrum to target key stellar absorption features. The SkyMapper Southern Survey (Keller et al. 2007) makes use of a  $310 \text{ \AA}$   $v$ -band filter, which covers (but is not centered on) the Ca II K line, along with SDSS-like *ugriz* filters, to survey much of the entire southern hemisphere sky. Other recent ongoing surveys, such as Pristine (Starkenburg et al. 2017), are employing a narrowband ( $\sim 100 \text{ \AA}$ ) filter centered on the Ca II H & K lines, a technique pioneered by Anthony-Twarog et al. (1991). When used in conjunction with pre-existing SDSS or Pan-STARRs (Tonry et al. 2012) broadband filter photometry, they obtain improved photometric metallicity estimates for numerous stars over large swaths of sky. As described by Starkenburg et al. (2017), Pristine’s resulting success rate for recovering EMP stars is  $26\%$ , with  $80\%$  of the remaining candidates being VMP stars. When combined with SDSS photometry, Pristine has demonstrated an accuracy of  $\sim 0.2 \text{ dex}$  down to  $[\text{Fe}/\text{H}] < -3.0$ . The use of Ca II H & K photometry from Pristine has already led to the discovery of a  $[\text{Fe}/\text{H}] < -4.66$  star, Pristine\_221.8781+9.7844 (Starkenburg et al. 2018).

Located at the Observatorio Astrofísico de Javalambre (OAJ, Teruel, Spain), the Javalambre Auxiliary Survey Telescope (JAST/T80) is a  $83 \text{ cm}$  telescope that is currently carrying out the Javalambre Photometric Local Universe Survey (J-PLUS; Cenarro et al. 2019). A twin of JAST/T80, based at the Cerro Tololo InterAmerican Observatory (CTIO, Chile), is executing the Southern Photometric Local Universe Survey (S-PLUS; Mendes de Oliveira et al. 2018), a Southern Hemisphere counterpart of J-PLUS.

J-PLUS is the first large-sky survey conducted at the OAJ, and was initially conceived to aid with photometric calibration of the upcoming Javalambre Physics of the Accelerating Universe Survey (J-PAS; Benítez et al. 2014), to be executed as well at the OAJ. While J-PAS is motivated by cosmological goals and, hence, has formally stronger requirements than J-PLUS in terms of photometric depth and number of narrowband filters (see Benítez et al. 2014 for details), J-PLUS has scientific goals that are largely (but not exclusively) related to the science of the Milky Way Halo and local Universe studies. The seven narrowband ( $\sim 100 \text{ \AA}$ ) filters of J-PLUS (J0378, J0395, J0410, J0430, J0450, J0515, J0660, J0861) are specifically designed to detect a variety of absorption features across the optical

<sup>2</sup>  $A(X) = \log \epsilon(X) = \log(N_X/N_H) + 12$ , where X represents a given element.



**Fig. 1.** Synthetic spectra of varying abundances are shown plotted over the J-PLUS filters  $J0395$ ,  $J0410$ , and  $J0430$ . The  $J0395$  filter is optimally centered on the Ca II H & K lines, which can be used as indicators of the overall stellar metallicity. The temperature-sensitive H $\delta$  line is apparent in the F410 filter. The  $J0430$  filter captures the CH G-band, which serves as a carbon indicator.

spectrum (Cenarro et al. 2019 for details). The filter system also hosts traditional *ugriz* bands, similar to those of SDSS, but with a more stable *u*-band response (Cenarro et al. 2019). Figure 1 depicts the location of some filters of special interest –  $J0395$ ,  $J0410$ , and  $J0430$ , from left to right – along with the corresponding spectral features, illustrated with synthetic spectra.

These filters provide the sensitivity necessary to detect absorption features such as the Ca II H & K lines, and the CH G-band, but mapping their behavior to estimate metallicity, [Fe/H], and the carbon-to-iron abundance ratio, [C/Fe], presents a highly degenerate problem. At temperatures above  $\sim 5750$  K, the CH molecule begins to dissociate, while the H $\gamma$  line broadens to the extent of disrupting the feature. The Ca II H & K lines also exhibit a strong temperature dependence. In addition, many lines respond to varying surface gravity due to Stark pressure broadening. A robust technique capable of unraveling the complex behavior of the parameter space explored by these filters for estimates of stellar parameters is thus required.

Machine learning tools – and in particular, Artificial Neural Networks (ANN) – have a long history (Rosenblatt 1958), but have found success in data-driven astronomical applications. Once adequately trained, these networks are powerful statistical pattern recognition tools capable of modeling highly complex behavior, through implementation of non-linear activation functions according to a user-specified architecture. Fabbro et al. (2018) applied their deep convolutional neural network, StarNet, to SDSS-III APOGEE (Eisenstein et al. 2011) spectra for determination of effective temperature, surface gravity, and metallicity, with precision and accuracy similar to that of the APOGEE pipeline. Using APOGEE DR14 for validation, Ting et al. (2019) applied a neural network function in their method, *The Payne*, to interpolate training sets of physical ab initio spectral models, and produced estimates of stellar parameters and 15 element abundances.

Deep convolutional networks have also been used to determine the evolutionary states of red giants from asteroseismology (Hon et al. 2017). Using frequency power spectra from the *Kepler* mission (Borucki et al. 2010), Hon et al. (2018) classified 426 red giants as red giant branch or helium-core burning stars.

Neural networks have been employed for classification-based problems in astronomy as well, such as star-galaxy classification (Kim & Brunner 2017) and spectral classification (Kheirdastan & Bazarghan 2016).

In this paper, the utility of ANNs for estimates of effective temperature and metallicity is explored. We present the Stellar Photometric Index Network Explorer (SPHINX<sup>3</sup>), a software package based on ANN estimation of stellar parameters ( $T_{\text{eff}}$  and [Fe/H]) from the J-PLUS mixed-bandwidth photometry. All training catalogs and parameter determination routines used in the text are provided. The databases and ANN used in the development and operation of SPHINX are described in Sects. 2 and 3. The basic structure is outlined in Sect. 4. In Sect. 5, the training process and results of effective temperature determinations are described, followed by the results of the metallicity determinations in Sect. 6. An application of SPHINX to a case study of the metal-poor globular cluster M15 is conducted in Sect. 7. A preliminary investigation of carbon sensitivity within the J-PLUS filter system is described in Sect. 8, and a concluding discussion of results and future applications is given in Sect. 9.

## 2. Databases

### 2.1. The J-PLUS early and first data releases

Calibration and testing of the network’s inputs were performed in part with the J-PLUS data set. The J-PLUS First Data Release<sup>4</sup> (hereafter DR1) consists of 511 pointings collected from November 2015 to January 2018 with JAST/T80 in all twelve optical bands described above, and covers approximately  $1022 \text{ deg}^2$ . The reduction and photometric calibration of the J-PLUS DR1, as well as the limiting magnitudes in the twelve bands, are presented in Cenarro et al. (2019). A representative subset of the DR1, the J-PLUS Early Data Release (EDR), comprises 18 pointings ( $36 \text{ deg}^2$ ), and is publicly available<sup>5</sup>. In addition to the present paper, the J-PLUS EDR and science verification data have been used to refine the membership in the nearby galaxy

<sup>3</sup> <https://github.com/DevinWhitten/SPHINX>

<sup>4</sup> [j-plus.es/datareleases/data\\_release\\_dr1](http://j-plus.es/datareleases/data_release_dr1)

<sup>5</sup> [j-plus.es/datareleases/early\\_data\\_release](http://j-plus.es/datareleases/early_data_release)

**Table 1.** Reddening coefficients for the J-PLUS photometric system.

Filter name	$\lambda_{\text{eff}}$ (Å)	$A(\lambda)/E(B - V)$
<i>u</i>	3536	4.295
J0378	3782	4.013
J0395	3939	3.849
J0410	4108	3.685
J0430	4303	3.506
<i>g</i>	4810	3.120
J0515	5141	2.864
<i>r</i>	6271	2.235
J0660	6604	2.070
<i>i</i>	7669	1.668
J0861	8611	1.377
<i>z</i>	8979	1.290

clusters Abell 2589 and Abell 2593 (Molino et al. 2019), analyze the globular cluster M15 (Bonatto et al. 2019), study the H $\alpha$  emission (Logroño-García et al. 2019) and the stellar populations (San Roman et al. 2019) of several local galaxies, and compute the stellar and galaxy number counts up to  $r = 21$  (López-Sanjuan et al. 2019).

Selections of the 6 arcsec aperture photometry were made from the MagABDualObj catalogs for each release. To limit contamination from non-stellar type objects, we removed extended sources using the stellarity index given by SExtractor, CLASS\_STAR  $\geq 0.92$  (Bertin & Arnouts 1996; López-Sanjuan et al. 2019). This morphological classifier was itself based on a neural network, taking into account the pixel scale of the image and full width at half maximum.

From these selections, we obtained 3 132 543 and 78 329 sources for DR1 and EDR, respectively. Minor variations in the 6 arcsec aperture photometry were present, tile by tile, in EDR. Corrections were made, based on a stellar-locus regression technique provided by Cristóbal-Hornillos et al. (priv. comm.). For training and validation of SPHINX, these catalogs were cross-matched, using a 1.0 arcsec search radius, with stars from the SDSS Reservoir (described in Sect. 2.2) with available stellar atmospheric-parameter estimates. This search radius was varied to 3.0 arcsec, and no new stars were obtained.

Interstellar extinction is expected to be small in EDR and DR1. However, application of extinction corrections is still expected to improve the performance of the photometric ANNs. One method to determine reddening coefficients for J-PLUS photometry is to convolve the filter-response functions corresponding to each of the J-PLUS filters with the reddening law from Fitzpatrick (1999). The result is a transmission-weighted sum of the wavelength-dependent extinction contribution within the filter bandpass. These values are provided in Table 1. The effective wavelength,  $\lambda_{\text{eff}}$ , was determined from a transmission-weighted average of the wavelength for each filter.

To estimate the line-of-sight reddening,  $E(B - V)$ , for our sources, we used Schlafly & Finkbeiner (2011). Corrections for reddening estimates for the J-PLUS photometry indeed improved the synthetic magnitude calibrations (Sect. 2.3). However, these corrections are subject to some limitations, particularly in that this extinction map is only two dimensional. In future work, we will utilize three-dimensional estimates of dust extinction (e.g., Green et al. 2018), based on parallaxes from Gaia Data Release 2 (Gaia Collaboration 2016). In this work, we implemented a hard extinction selection, corresponding to  $E(B - V) <$

0.05 dex, so that more reddened sources were not admitted to our samples.

## 2.2. The SDSS Reservoir

Three separate spectroscopic campaigns comprise the training databases for SPHINX, hereafter called the SDSS Reservoir. The first, SDSS's Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013), obtained over 500 000 stellar spectra (including repeated stars) across a wavelength of 3 600–10 000 Å ( $R \approx 2000$ ) over approximately 10 000 deg<sup>2</sup> of the sky. Our database included 80 221 stars from this survey, selected for high S/N and low metallicity ([Fe/H]  $< -0.5$ ). The Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009) took place in two phases, SEGUE-1 and SEGUE-2, and covered a combined area of 2755 deg<sup>2</sup>. We made use of 147 811 of the  $\sim 350$  000 spectra from this sample. Finally, we included 74 572 spectra from the component of SDSS-I known as the Legacy Survey, provided by SDSS DR7 (Abazajian et al. 2009). This survey covered primarily  $\sim 7500$  deg<sup>2</sup> of the North Galactic Cap, with additional stripes in the South Galactic Cap amounting to  $\sim 740$  deg<sup>2</sup>.

Stellar parameters for these databases were derived using the SEGUE Stellar Parameter Pipeline (SSPP; Lee et al. 2008a,b). This pipeline made robust determinations of effective temperature ( $T_{\text{eff}}$ ), surface gravity ( $\log g$ ), metallicity ([Fe/H]), and carbonicity ([C/Fe]) over a temperature range  $4000 < T_{\text{eff}}$  (K)  $< 8000$ . Typical random uncertainties in the effective temperature and metallicity estimates for F- and G-type stars were determined empirically,  $\sigma(T_{\text{eff}}) \sim 130$  K, and  $\sigma([\text{Fe}/\text{H}]) \sim 0.21$  dex (Allende Prieto et al. 2008). The color–magnitude diagram of stars available for training from the SDSS Reservoir is shown in Fig. 2, along with the effective temperature and metallicity distributions of source matches with J-PLUS DR1. In all three plots, stars were limited to those with  $\sigma(T_{\text{eff}}) < 250$  K,  $\sigma([\text{Fe}/\text{H}]) < 0.3$  dex, and  $S/N > 25$ , as dictated by the SSPP.

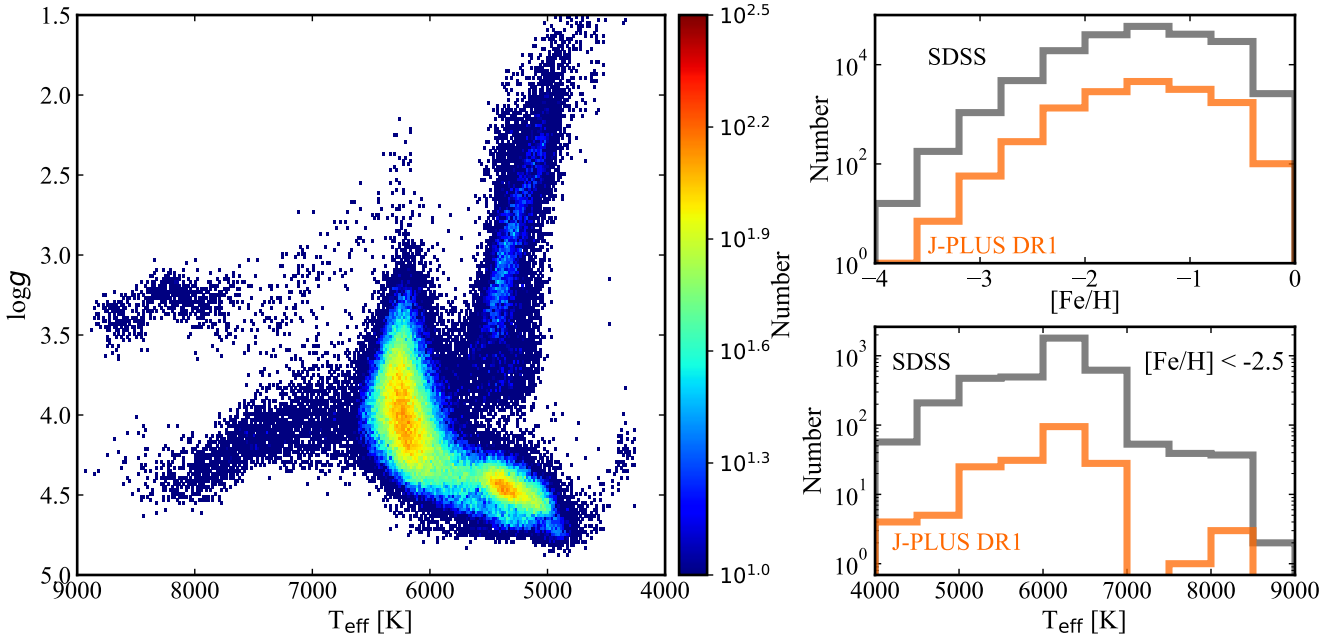
## 2.3. Synthetic magnitudes

Synthetic photometry was computed for all spectra in the databases described above. This was done initially to provide the photometric inputs required for network training before J-PLUS had accumulated an adequate number of source photometry with EDR and DR1. First, filter intensities were computed by taking the inner product of the flux-calibrated spectrum and the J-PLUS filter-response functions. These filter intensities were converted to synthetic magnitudes using the following form of the asinh magnitude, or luptitude (Lupton et al. 1999):

$$m = (-2.5 / \ln 10) \cdot \left[ \sinh^{-1} \frac{f}{2bf_0} + \ln b \right]. \quad (1)$$

Here,  $f$  was the integrated flux through the filter. Appropriate values for the classical zero-point,  $f_0$ , and the softening parameter,  $b$ , were determined by calibrating the synthetic intensities to known J-PLUS magnitudes from a cross-match of EDR and DR1 with synthetic magnitudes derived from the flux-calibrated spectra for stars in common with the SDSS Reservoir. This was done using the orthogonal distance regression technique (Boggs & Rogers 1990), taking into account error estimates in the native photometry as well as the variances in the flux reported in the medium-resolution SDSS spectra.

The resulting synthetic magnitude calibrations for stars in the SDSS Reservoir are shown in Fig. 3. In addition, minor



**Fig. 2.** *Left panel:* Hess diagram of the SDSS Reservoir available for training of the photometric ANN pipeline. *Top right panel:* metallicity distribution of the SDSS Reservoir and the subset with J-PLUS DR1 photometry. *Bottom right panel:* distribution of effective temperatures for the subset of the SDSS Reservoir with  $[\text{Fe}/\text{H}] < -2.5$ .

**Table 2.** Limiting magnitudes for synthetic calibrations.

Filter name	$m_{\text{lim}}$ (mag)	$\sigma_m$ (mag)
J0395	18.46	0.11
J0410	18.52	0.09
J0430	18.69	0.08
<i>g</i>	18.68	0.07
J0515	18.79	0.05
<i>r</i>	18.90	0.04
J0660	18.80	0.04
<i>i</i>	18.75	0.04
J0861	18.48	0.05

offsets ( $\pm 0.2$  mag) were found and removed in a few J-PLUS tiles during the calibration. Residuals in the synthetic magnitude calibrations are moderately dependent on the  $S/N$  of the underlying SDSS spectra, particularly for the J0395 and J0410 filters. Sources with  $S/N < 25$  and magnitude uncertainties larger than 0.1 mag were thus excluded for each calibration. This  $S/N$  cut had the effect of preferentially removing calibration stars at fainter magnitudes, where we found that the scatter for each calibration tended to increase. Considering again a critical limit in the scatter of 0.1 mag, we found limiting magnitude values for each filter, provided in Table 2 along with the standard deviation,  $\sigma_m$ , of the corresponding calibration. Use of synthetic magnitudes beyond these limits is not recommended, as the calibrations are less reliable. Both uncertainties in the underlying SDSS spectra and the native J-PLUS photometry contribute to the scatter seen in the Fig. 3, but if synthetic magnitudes are limited to those with  $S/N > 25$ , the critical residual value of 0.1 mag is reached at characteristically brighter values in the native photometric system.

As expected, the scatter in the photometric calibration is seen to increase for bluer filters, where CCD response and atmospheric absorption become more problematic. This calibration

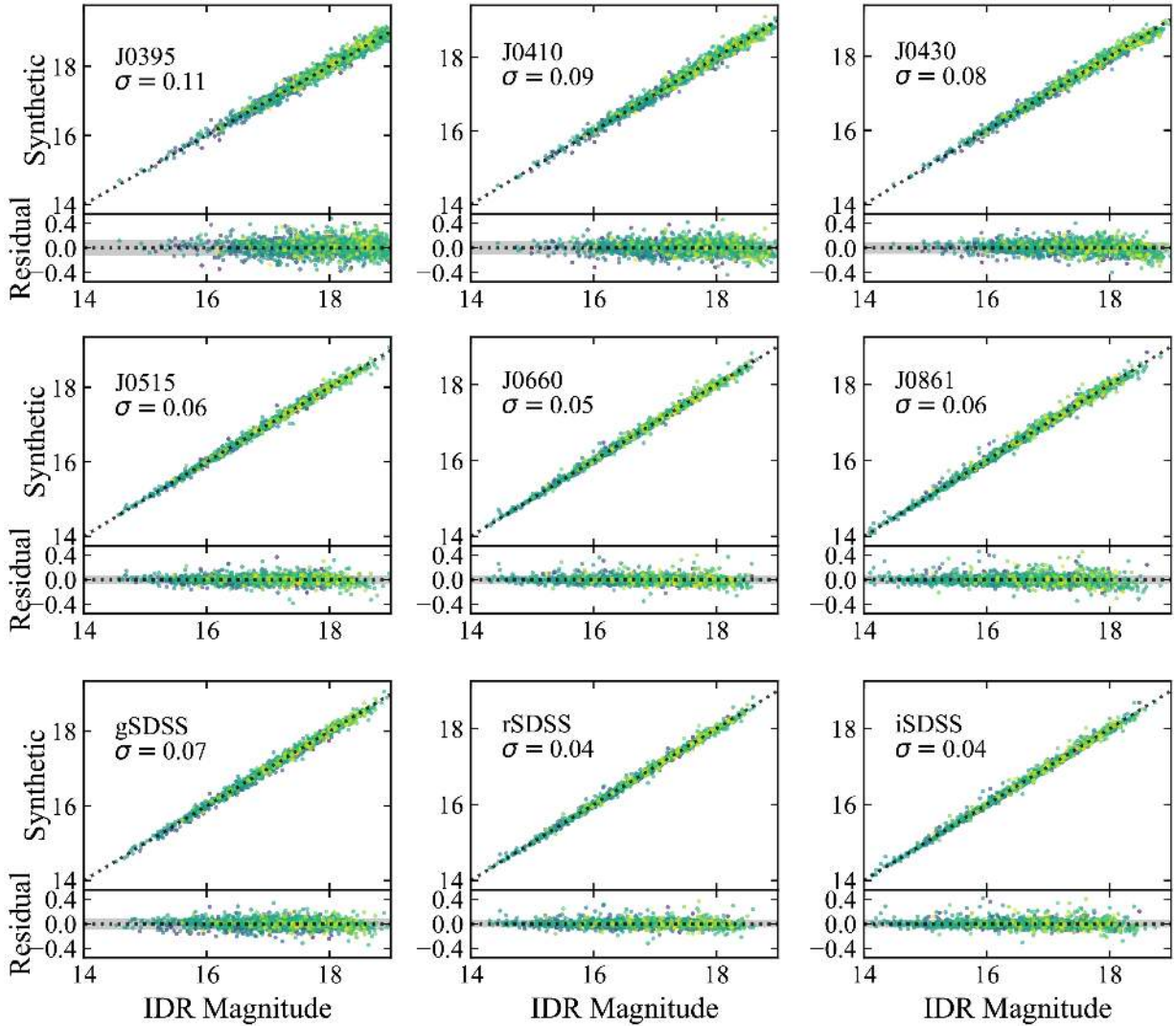
ensured that synthetic magnitudes were on the same scale as the J-PLUS photometric system, thus both synthetic and native photometry could be used interchangeably within the magnitude and  $S/N$  limits specified.

The blue cutoff of the SEGUE spectra ( $\sim 3700$  Å) precluded determination of synthetic magnitudes for the J0378 and *u*-band filters. The same was true for the *z*-band filter, for which the response curve extended beyond the range of the SEGUE spectra. Consequently, these filters were excluded from the methodology. Once photometry is available for a sufficient number of these stars, construction of synthetic magnitudes for training of the ANNs will not be necessary, and these filters should certainly be included, as they each capture important information about the stellar atmospheres.

#### 2.4. The synthetic library

For exploratory and training purposes, we made use of a library of synthetic spectra. To build the synthetic spectra library, we proceeded similarly to the approach described by Lee et al. (2013). In short, we used a specific grid of model atmospheres computed with the MARCS code (Gustafsson et al. 2008), which took into account the impact of the carbon enhancement in the atmosphere. From those model atmospheres we generated the synthetic spectra using the Turbospectrum routine (Álvarez & Plez 1998; Plez 2012). Compared to Lee et al. (2013), we improved the procedure in several aspects. First, we extended the wavelength range, now covering the entire optical region from 3000 to 10,000 Å. In addition, we updated the linelists to include CN, CH, NH,  $\text{C}_2$ , MgH, TiO, ZrO, CaH, VO, and SiH for the molecules, and the ultimate version of VALD3 (Ryabchikova et al. 2015) for the atoms. Finally, we extended the grid to a much larger space of parameters,  $3500 \leq T_{\text{eff}} \text{ (K)} \leq 8000$  in increments of 250 K,  $0.0 \leq \log g \leq 5.0$  in increments of 0.5 dex,  $-4.5 \leq [\text{Fe}/\text{H}] \leq +0.5$  in increments of 0.25 dex, and  $-1.5 \leq [\text{C}/\text{Fe}] \leq +4.5$  in increments of 0.25 dex.

Due to the finite resolution of this library, we interpolated when necessary using a 4D cubic spline interpolation routine



**Fig. 3.** Synthetic magnitudes determined with SDSS Reservoir and J-PLUS filter-response functions are compared with photometry from the J-PLUS First Data Release. Scatter in the calibrations increases for bluer filters, as the uncertainty in the interstellar extinction correction is larger. The colors are indicative of the photometric tile from which the data were obtained, for which no biases are seen.

(Y. S. Lee, priv. comm.). Training sets could then be generated with stellar parameters distributed anywhere within the parameter regions of interest. A crude flux-calibration was performed on these spectra by convolving each normalized spectrum with a blackbody function corresponding to its effective temperature. This library was primarily used for investigation of the photometric carbon sensitivity, discussed in Sect. 8.

### 2.5. Narrowband parameter response

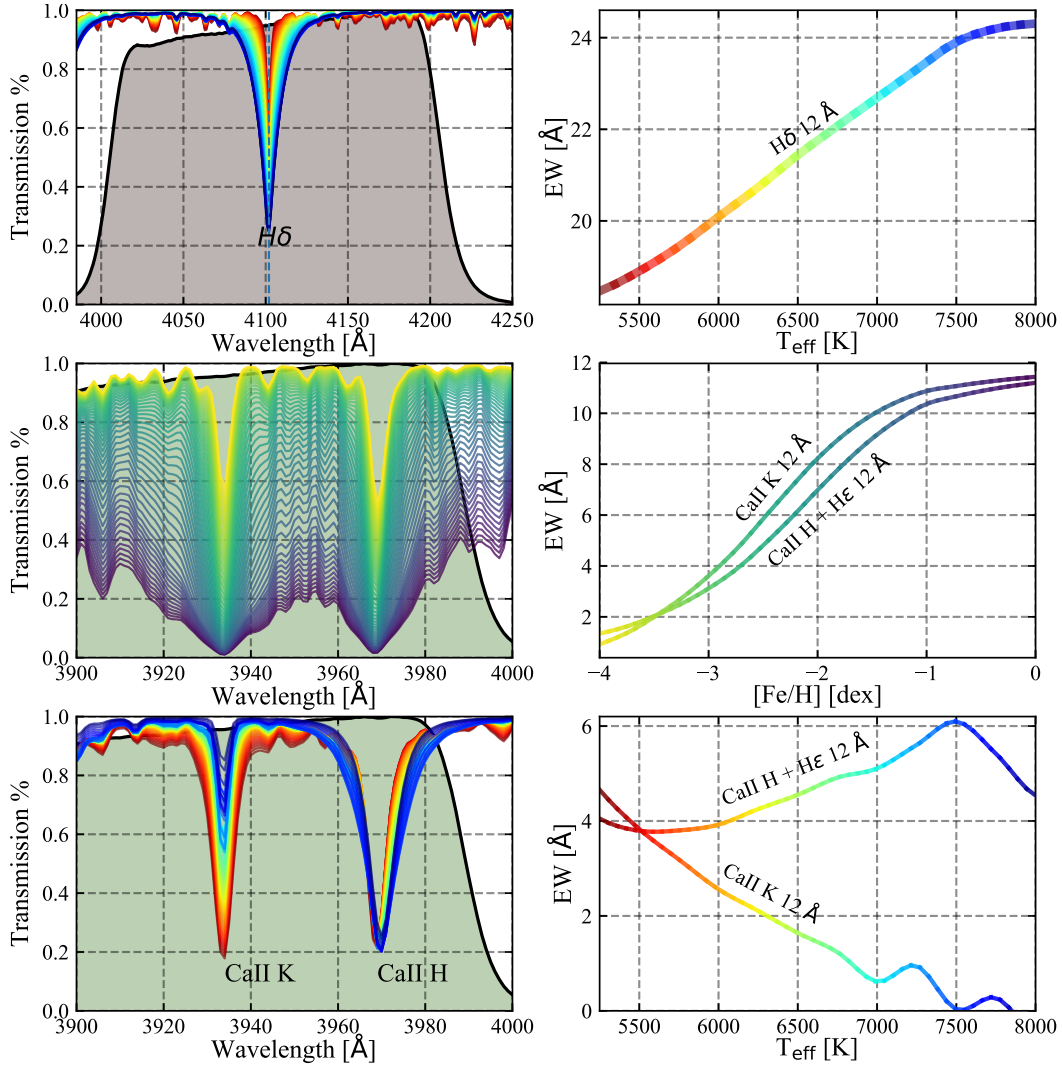
In addition to sampling the overall blackbody function of the spectral-energy distribution (SED), we expect response to a number of temperature-sensitive features present in the narrowband filters. For example, the J0410 filter hosts a particularly temperature-dependent feature, the H $\delta$  line (4102 Å); see Fig. 4. In the upper left panel, we consider the strength of H $\delta$  over the temperature range  $5250 < T_{\text{eff}} \text{ (K)} < 8000$  for a star of  $\log g = 2.5$ ,  $[\text{Fe}/\text{H}] = -2.50$ , and  $[\text{C}/\text{Fe}] = 0.0$ . As seen in the upper right panel, this temperature range corresponds to a 31% increase in the equivalent width of H $\delta$ , from 18.4 to 24.2 Å. Thus, we anticipate success using networks in which training parameters include photometric colors and magnitudes incorporating the J0410 filter.

The Ca II H & K lines (3969 and 3934 Å, respectively) display similar behavior with increasing metallicity. In the middle-left panel, we vary the metallicity of a  $T_{\text{eff}} = 5000$  K star over the range  $-4.00 < [\text{Fe}/\text{H}] < 0.0$ . From inspection of the middle-right panel, for a star of solar metallicity, the equivalent widths of the Ca II H & K lines are 42 and 64% stronger than for a  $[\text{Fe}/\text{H}] = -2.0$  star of the same temperature. However, the Ca II H & K lines also exhibit an asymmetric temperature dependence. In the same range of effective temperature, the Ca II H line strength can increase by as much as 50% for a star with  $[\text{Fe}/\text{H}] = -2.50$ , due to the increasing influence of He at  $\lambda = 3970$  Å. Meanwhile, the Ca II K line nearly vanishes by  $T_{\text{eff}} = 7500$  K. We therefore expect to obtain a keen sensitivity to metallicity by incorporation of the J0395 filter in our methodology, and anticipate the necessity to unravel a degeneracy with effective temperature.

## 3. Artificial neural network

### 3.1. Architecture

In the development of SPHINX, we generalized the use of the core ANN element with what amounts to a three-layered system.



**Fig. 4.** *Top panels:* temperature sensitivity of the H $\delta$  feature. Over an effective temperature range of  $5250 < T_{\text{eff}}$  (K) < 8000, we expect an increase in absorption of 31% for a star of  $\log g = 2.50$ ,  $[\text{Fe}/\text{H}] = -2.50$ , and  $[\text{C}/\text{Fe}] = 0.0$ . *Middle panels:* variation in Ca II H & K line strengths with increasing metallicity for a star of  $T_{\text{eff}} = 5000$  K,  $\log g = 1.0$ , and  $[\text{C}/\text{Fe}] = 0.0$ . *Bottom panels:* Ca II H & K line response to increasing effective temperature. The Ca II H line is nearly superposed with He.

The motivations for this structure are discussed in Sect. 4. Here, we consider the basic function of an ANN.

For the core implementation of the ANN, we made use of the multilayered perceptron class, `MLPRegressor`, from `scikit-learn` (Pedregosa et al. 2011). We chose a feed-forward algorithm with one hidden layer of neurons. As an example, Fig. 5 shows a schematic of this network. Each node computes a non-linear function of the scalar product of the input vector and a weight vector. For  $N$  nodes, the functional form can be expressed as follows:

$$y(\mathbf{x}) = \tilde{g} \left[ \sum_{i=1}^N \omega_{i2} \cdot g(\omega_{i1}^T \mathbf{x}) \right]. \quad (2)$$

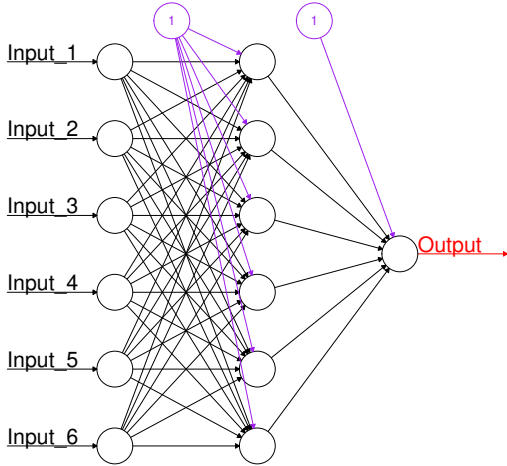
Here,  $\mathbf{x}$  denotes the input vector, in our case, the pre-scaled magnitudes from a subset of the J-PLUS filter set. Each neuron,  $i$  of  $N$ , in the hidden layer receives the sum of the input vector with the weight vector corresponding to the first layer,  $\omega_{i1}$ , represented for simplicity by the inner product, where  $\mathbf{T}$  denotes the transpose. For each hidden layer, we implement a bias neuron, the corresponding weight of which is included in

$\omega_{i1}$ . These bias neurons produce a constant output of 1, and thus are not connected to the previous layer. Bias neurons help to shape the output of the activation function, where they shift the inner product in a way analogous to the y-intercept of a linear equation.

The node then applies an activation function,  $g$ , and the output is provided to the output layer. A similar sum is then performed with the outputs of each neuron in the hidden layer and the weights of the second layer where, in general, the output activation function,  $\tilde{g}$ , is not required to be the same function used in previous layers.

A hyperbolic tangent activation function was implemented in SPHINX, in part because it resulted in faster convergence times, but also because the function maps the input vectors to  $(-1,1)$ , while the logistic sigmoid produces values in an asymmetric range  $(0,1)$ . It has been empirically shown that symmetric activation functions produce faster convergence times (Le Cun et al. 1991). This is the case for the application of effective temperature and metallicity estimates.

Optimal weights for the network are determined using a stochastic gradient descent algorithm. With traditional



**Fig. 5.** Example network element used in the SPHINX ANN pipeline. The ANN unit utilizes a feed-forward algorithm and consists of a single hidden layer. The inputs are scaled combinations of J-PLUS colors and magnitudes.

back-propagation, partial derivatives are computed for each weight with respect to a specified error function,  $E$ , to determine the influence of small perturbations of each weight's value. In this manner, new values of the  $i$ th weight in the  $j$ th layer can be determined as follows (Riedmiller 1994):

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \epsilon \frac{\partial E}{\partial \omega_{ij}}(t). \quad (3)$$

Here,  $\epsilon$  is the learning rate, which scales the update correction. Fixed values of this parameter were found to introduce oscillations, where the updated value over-corrects with each iteration, and so never converges to the optimal weight. To overcome this, a stochastic gradient-based optimizer known as Adam was implemented, in which an adaptive-learning rate is computed from first-order moments of the gradient in the scalar cost function (Kingma & Ba 2014).

### 3.2. Prescaling/normalization

Prior to using the magnitudes of our training set to converge the network, a form of prescaling was applied, whereby all input variables were set to the same scale. The chosen technique was simply a linear rescaling, in which we subtracted the sample mean and divide by the standard deviation, leaving each input vector with a mean of zero and variance of one. A more robust method of prescaling was also included as an option in SPHINX, in which the center of the distribution is taken as the median, and the spread is determined by some specified percentile range. This provides the option to control the fraction of inputs between  $[-1, 1]$ , or the optimal domain of the activation function's response.

Although synthetic magnitudes are calibrated to the J-PLUS photometric system, it is highly desirable to train the ANN with inputs for which the distributions match that of the validation, or science set. Otherwise, prescaling the input colors could push these inputs to a region that requires the ANN to extrapolate from the inputs, or operate near the asymptotic region of the activation function.

This ANN structure serves as the basis for the architecture of SPHINX, where we generalize the use of the ANN in a larger, three-layered structure in an attempt to maximize the

capability of the ANN, while including the variety of input combinations available. We describe the detailed design of SPHINX in Sect. 4.3.

## 4. Parameter determination and optimization with SPHINX

### 4.1. Training set assembly

In contrast to an ANN's capability for modeling patterns in high-dimensionality space, their ability to extrapolate beyond the boundaries of their underlying training set is quite limited. One consequence of this limitation is the need to be aware of the distribution of the validation set prior to training of the network. While we have no a priori knowledge of the chemical abundance and atmospheric-parameter distributions in EDR and DR1, we have all of the necessary information regarding the magnitude distributions, which can be used to construct an ideal training set.

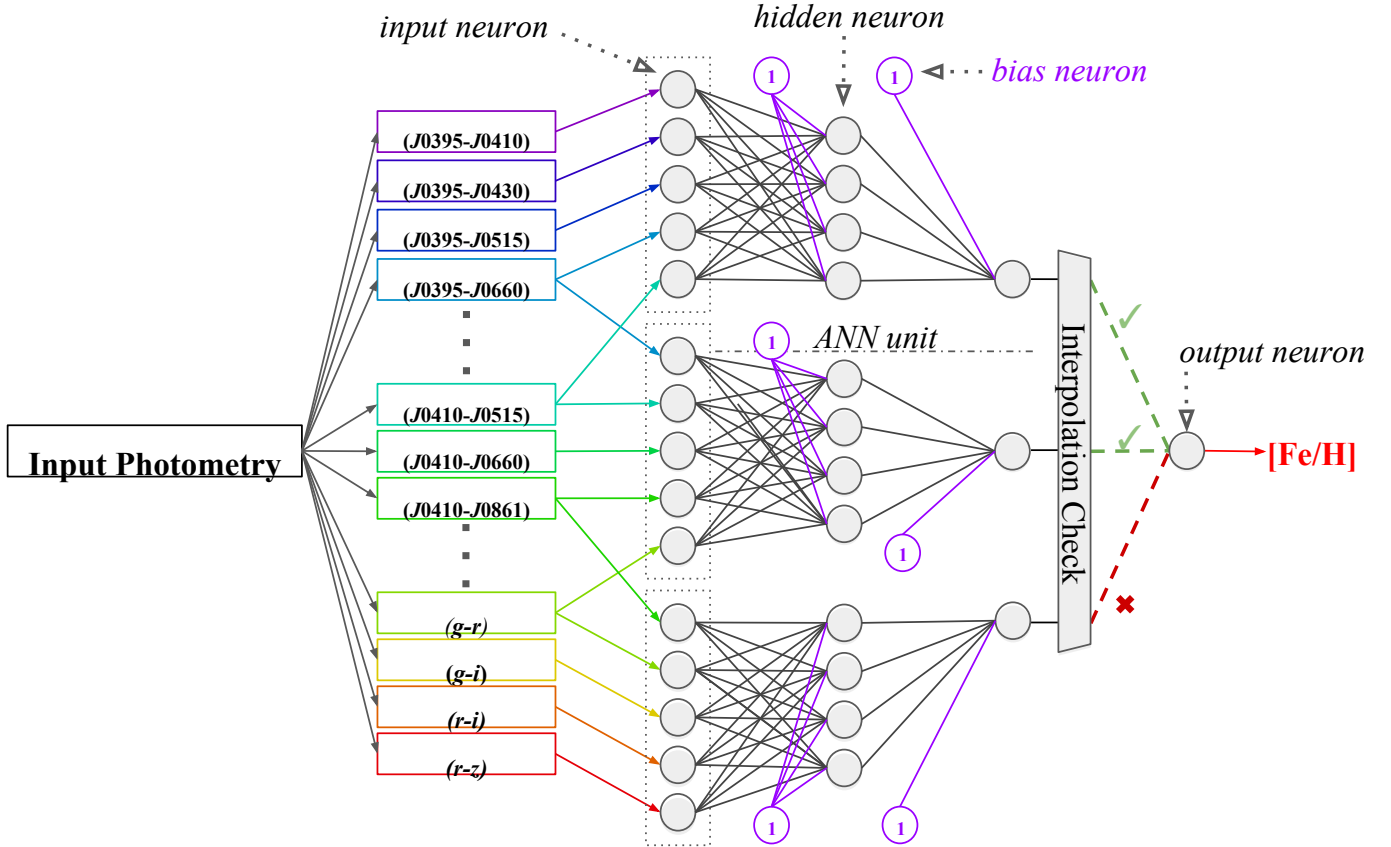
At present, SPHINX allows for selection of training stars from a variety of catalogs, including the subset of EDR and DR1 sources with available SSPP-estimated parameters, as well as SDSS Reservoir sources with synthetic magnitudes calibrated to the J-PLUS photometric system. SPHINX then performs an S/N, photometric error, and  $E(B-V)$  rejection according to the specified thresholds, in addition to the faint and bright limits, all of which are set in the input parameter file. With the appropriate parameter bounds and rejections determined, the training set can be uniformly sampled across the target variable, in this case  $T_{\text{eff}}$  or  $[\text{Fe}/\text{H}]$ . This routine is optional, but is designed to protect against overemphasis by the network on a particular region of the parameter space.

Once all desired processing of a training set is complete, we define a scale frame, which describes each input distribution in the target set. This includes the center location and spread estimates that SPHINX used to linearly scale and unscale all inputs to the networks for training and parameter estimation. If the number of target sources is large, it may be desirable to set the scales according to the target photometry. However, if the target list consists of only a few sources, this is not advised; the scales should be set based on the training set's photometric distributions. The default manner of determining the center and spread of each distribution is to fit a Gaussian to each input. A more robust method utilizing the median and fourth-spread, or  $f$  spread (i.e., the interquartile range), is also available, although it was found that alternative estimates of the scales did not significantly influence the network's performance.

### 4.2. Approximating input distributions

While the SDSS Reservoir and DR1 catalogs are sufficiently large that we can construct training sets across a reasonably wide range of stellar parameters and magnitudes, an effective method to accommodate a variety of target (or testing) distributions is required. For instance, our selection from the J-PLUS M15 photometry (details in Sect. 7) comprises stars of  $14 < g < 18$ , well within the SDSS Reservoir distribution. However, the SDSS Reservoir distribution peaks at  $g = 18.3$ , with a standard deviation of 1.2 mag, while the M15 distribution mean and standard deviation are  $g = 16.3$  and 1.1 mag, respectively. Training stars naively sampled from the SDSS Reservoir catalog would introduce a bias towards the fainter stars in the M15 cluster, which is not ideal.





**Fig. 6.** Example schematic of the Stellar Photometric Index Network Explorer (SPHINX). The architecture consists of a scalable three-layered neural network. The hidden layer performs the hyperbolic tangent activation function on the neuron-specific weighted inputs. The following layer represents each sub-networks best estimate of the stellar parameter. The final layer performs a validation weighted sum of the estimates verified to have resulted from the interpolating networks.

We include the option to force the training set input distributions to best approximate those of the target or science set. For a certain magnitude or color input, it is assumed that the center and spread of the target distribution have previously been set. The following error function is minimized for  $\gamma$ , the scale of the Gaussian distribution in the training set corresponding to the center and scale of the target distribution:

$$\Phi(\gamma|\bar{x}, \sigma) = \sum_i^N \left| y_i - \gamma \cdot \exp \left[ \frac{-(x - \bar{x})^2}{2\sigma^2} \right] \right|^{-\alpha}, \quad (4)$$

where the scale  $\gamma$  sets the maximum number of stars with input value  $x \pm \delta x$  that can be drawn from the training distribution, such that the resulting input distribution conforms to a Gaussian described by  $\bar{x}$  and  $\sigma$ . We leave  $\alpha$  as a free parameter (default  $\alpha = 1/2$ ) to tune the level with which the algorithm penalizes non-conformities to the desired distribution.

#### 4.3. Network optimization with SPHINX

In general, the proper combination of inputs for the network – or, for that matter, their quantity – is not known. We anticipate the use of temperature-sensitive filters such as *J0410* and *J0660*, and metallicity-sensitive filters such as *J0395*, *J0515*, and *J0861*. Beyond these, however, it is not directly apparent what additional combinations facilitate the capabilities of the network. Furthermore, for networks operating with different combinations of photometric inputs, one network may be forced to extrapolate, but not another.

To address these concerns and ensure resilience against potentially faulty inputs in both the training and science case data, we enable a system of networks, or ANN units, each with a distinct combination of the available inputs. SPHINX determines all combinations of the specified filters and their corresponding colors, assigning and training a network for a specified number of these combinations. Throughout the operation and analysis of SPHINX, the number of inputs to each ANN unit is small ( $N = 6$ ), to maximize the ANN’s dependence on the each input in the subset, and also to mitigate the convergence time of each ANN in the array.

The basic schematic of SPHINX is shown in Fig. 6. For simplicity of this illustration, only a sub-sample of available colors is shown, and the number of ANN units is limited to three. As can be seen, SPHINX amounts to a three-layered neural network, consisting of a layer of subordinate ANN units that we call the network array. This array is of scalable size; we evaluate the performance of SPHINX as a function of array size in Sect. 6. For every network estimate made, SPHINX notes when that network is extrapolating outside of its training domain. This is accomplished with a structure that we refer to as an interpolation frame, which stores the domains of all inputs in the training set. All estimates made in the network array layer are subject to an interpolation check, flagging any and all estimates where extrapolation was necessary.

During training, the performance of each ANN unit in the network array is evaluated and recorded using a validation set, and this analysis is used to assign each ANN unit a score for future reference and for computing final estimates. We define

a score for the network, which we base on a modified median absolute deviation (MAD) of the verification set:

$$\xi_N = \text{median}(|\mathbf{E}_N - \text{median}(\mathbf{E}_N)|)^{-\beta}, \quad (5)$$

where  $\mathbf{E}_N$  represents the error in network  $N$ 's estimate of the target variable in the verification set. Here,  $\beta$  is left as a free parameter (default  $\beta = 2$ ) to control the degree to which we penalize ANN units with poorer performance in the final parameter estimate. We then assert that the proper science estimate is the weighted sum of each network estimate and its corresponding score,  $\xi_N$ .

A system of twelve filters corresponds to  $\binom{12}{2} = 66$  possible colors. With a network array size of six ANN units, the resulting number of unique color combinations is  $\binom{66}{6} = 90\,858\,768$ , an infeasible number of networks to train in a finite time. To address this, we allow for a specified number of input combinations to be utilized, whereby a subsample of networks from the underlying ensemble is selected. This can be done at random, or if any specific filters or colors are desired, SPHINX can be restricted to employ combinations of colors in which at least one input is comprised of these specifications. We make use of this feature for effective temperature and metallicity estimates, where SPHINX is forced to use of  $J0410$  and  $J0395$ , respectively, in combination with the other available filters.

Before consideration of the final parameter estimates, each ANN unit estimate is verified by the interpolation check described above. The number of contributing ANN units is recorded for future consideration via the `NET_ARRAY_FLAG`, and the final estimate is simply a weighted sum of the contributing ANN unit estimates and corresponding validation scores,  $\xi_N$ . With the ensemble of contributing ANN units, we compute the MAD of the estimates, as well as the standard deviation, which is also weighted by the validation scores. These serve as the uncertainty in the parameter estimate reported by SPHINX. This procedure can also be used to determine the most optimal networks achieved in the network array. If specified, SPHINX considers only networks with the highest scores,  $\xi_N$ , and excludes all other estimates from the final parameter estimation.

## 5. Effective temperature determination with SPHINX

For effective-temperature training, we restricted the range of interest to  $4500 < T_{\text{eff}} \text{ (K)} < 8500$ . We selected stars with DR1 photometry from the SDSS Reservoir, and reserved all EDR sources for testing purposes. For simplicity, stars for which metallicities exceed solar were excluded, anticipating the influence of high metal abundance on the underlying continua. A limit of 0.1 mag in the error reported for the observed photometry, with a faint limit of 20 on all magnitudes was applied to the training set, in addition to the extinction limit of  $E(B - V) < 0.05$  mag stated previously.

It was possible that, even with excellent synthetic photometric errors and an optimal S/N in the underlying spectrum, the temperature assigned to the training star by the SSPP was incorrect, or at least imprecise. We therefore implemented a  $\pm 120$  K cut on the adopted  $T_{\text{eff}}$  error estimate, and insisted that the individual estimates from the SSPP – in this case the adopted estimate and estimates derived from the  $H\alpha$  &  $H\delta$  Balmer-line strengths<sup>6</sup> – not differ from by more than 75 K.

<sup>6</sup> T1 and T2 as described in Lee et al. (2008a).

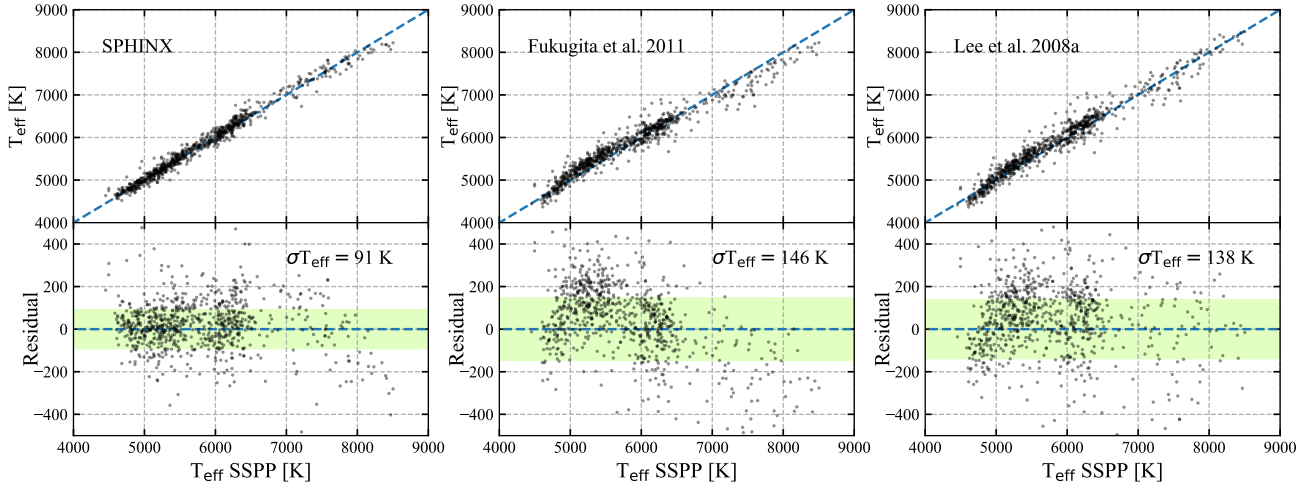
With the optimally cleaned stars selected, training sets were constructed by uniformly sampling temperatures between the maximum and minimum temperature thresholds. The act of sorting and partitioning the catalog into 15 bins prior to randomly sampling 200 stars from each bin ensured that a roughly even distribution of temperatures was obtained. The final DR1 training batch consisted of 1152 stars. Both the training and testing sets included stars with surface gravities in the range  $1.0 < \log g < 5.0$ . The median  $\log g$  in both sets was 4.1, with a standard deviation of 0.5 dex. Both sets were dominated by main-sequence stars, but  $\sim 5\%$  of stars had surface gravities consistent with giants ( $\log g < 3.0$ ).

SPHINX was applied to 1015 stars with EDR photometry. For analysis of the network performance, we excluded testing sources with spectral  $S/N < 20$ , or a larger uncertainty in the SSPP  $T_{\text{eff}}$  estimate ( $\sigma(T_{\text{eff}}) > 150$  K). For comparison, effective temperature estimates were made with two broadband calibrations, Eq. (1) of Fukugita et al. (2011), and  $T_l$  of Lee et al. (2008a). Fukugita et al. (2011) determined an empirical temperature range for their calibration of  $3850 < T_{\text{eff}} \text{ (K)} < 8000$ , and noted the smallest scatter of  $\sigma(T_{\text{eff}}) = 93$  K when using  $g-r$  compared to other broadband colors. The  $T_l$  estimate was derived to be effective for a wide range of temperatures, beyond  $4500 < T_{\text{eff}} \text{ (K)} < 7500$ , making it ideal as a comparison for our calibration with SPHINX.

We made use of a number of robust estimators of central location and scale, and adopted the suggested nomenclature from Beers et al. (1990). For measures of central location, in addition to the median  $C_M$ , we used the trimean  $C_{\text{TRI}}$  – based on the sample median and upper and lower fourths of the distribution(s) – the biweight location  $C_{\text{BI}}$ , and the mean  $C_\mu$ . For estimates of scale, we made use of a normalized median absolute deviation,  $S_{\text{MAD}}$ , which is particularly suited for heavy-tailed distributions (Beers et al. 1990). In addition, a scale estimate based on the fourth spread, the  $f$  pseudosigma,  $S_f$ , was used, along with the biweight scale estimate  $S_{\text{BI}}$  and standard deviation  $S_\sigma$ .

The results of the temperature estimates made with SPHINX and the comparison calibrations on the EDR photometry are shown in Fig. 7 and summarized in Table 3, where we compare as well the statistics of high temperature stars ( $T_{\text{eff}} > 7000$  K). For estimates made with SPHINX, shown in the left-most panel of Fig. 7, central location estimates of the residuals  $C_\mu = +21$  K and  $C_{\text{TRI}} = +20$  K indicated a slight bias towards overestimation for the EDR sample. This behavior was in contrast to the underestimation seen for temperatures above  $T_{\text{eff}} > 7000$  K, where the  $C_M$  and  $C_{\text{TRI}}$  of the residuals were  $-14$  and  $-28$  K, respectively. The scatter in the region, where  $S_f$  increased from 93 to 171 K. Similarly, the  $S_{\text{MAD}}$  increased from 93 K to 172 K. We attributed this behavior both to edge effects emerging near the limit of the network interpolation range. Interestingly, it was found that inclusion of the  $J0410$  filter did not significantly improve the performance of effective temperature estimates made with SPHINX. We conclude that the quality of these estimates is more likely a matter of J-PLUS filters tracing the overall SED, as well as the capability of SPHINX to interpret the behavior of the SED over a wide temperature range.

Using the  $g-r$  calibration given in Eq. (1) of Fukugita et al. (2011), we computed an  $S_f$  of 156 K. This calibration, shown in the middle panels, was biased as well towards overestimation, indicated by the mean of the residuals,  $+66$  K, and  $C_{\text{TRI}}$  of  $+55$  K. Similar to the estimates made with SPHINX, we found underestimation of effective temperature beyond  $T_{\text{eff}} > 7000$  K, where the median of the residuals was  $-230$  K, with a  $C_{\text{TRI}}$  of



**Fig. 7.** Effective-temperature estimates,  $T_{\text{eff}}$ , from SPHINX, as a function of the adopted value from the SSPP for a testing catalog consisting of J-PLUS EDR photometry (*left panels*). For comparison,  $g-r$  calibrations from Fukugita et al. (2011; *middle panels*) and Lee et al. (2008a; *right panels*) were applied to the sample. The residuals for each temperature calibration are shown against the accepted temperature in the *bottom panels*, where the green region depicts the standard deviation from a Gaussian maximum-likelihood fit.

**Table 3.** Central location and scale estimates of photometric temperature calibration residuals.

	SPHINX (K)	Fukugita et al. 2011 (K)	Lee et al. 2008a (K)
$C_M$	+20	+60	+57
$C_{\text{TRI}}$	+20	+56	+56
$C_{\text{BI}}$	+20	+57	+56
$C_\mu$	+21	+66	+64
$S_{\text{MAD}}$	93	157	143
$S_f$	93	156	144
$S_{\text{BI}}$	106	171	156
$S_\sigma$	91	146	138
$T_{\text{eff}} > 7000 \text{ K}$			
$C_M$	-14	-230	-10
$C_{\text{TRI}}$	-28	-233	-15
$C_{\text{BI}}$	-26	-230	-19
$C_\mu$	-5	-201	+13
$S_{\text{MAD}}$	172	218	199
$S_f$	171	213	195
$S_{\text{BI}}$	161	195	208
$S_\sigma$	188	196	196

-233 K. Estimates in this regime became increasingly uncertain, with  $S_f$  of 213 K, and a  $S_{\text{MAD}}$  of 218 K.

In the right panel we compare the network temperature estimates to a polynomial calibration of  $g-rr$  from the SSPP,  $T_I$  (see Lee et al. 2008a, for details). For this calibration, the median in the residuals of +57 K and  $C_{\text{TRI}}$  of +56 K again indicated a systematic overestimation. The standard deviation of the Gaussian maximum-likelihood fit of 138 K increased to 196 K in the high temperature range,  $T_{\text{eff}} > 7000 \text{ K}$ . A number of outliers were apparent around  $T_{\text{eff}} \sim 6200 \text{ K}$  for each calibration, particularly in the Lee et al. (2008a) calibration, which was expected if variable stars such as RR Lyrae stars are present in the EDR sample. Offsets such as these would occur for stars of which spectroscopic and photometric temperature estimates were made at disparate phases of the pulsation cycle.

We concluded that the use of SPHINX with J-PLUS photometry provides effective temperatures with a bias reduced by

a factor of three and a dispersion reduced by  $\sim 40\%$  with respect to the previous broadband calibrations. In addition, SPHINX is less prone to under-estimation in the high-temperature regime beyond 6500 K. No significant influence by surface gravity on the effective temperature was seen in the range  $2.0 < \log g < 5.0$ , however for lower values ( $\log g < 2.0$ ) determinations tended to be overestimated by  $\sim 70 \text{ K}$ .

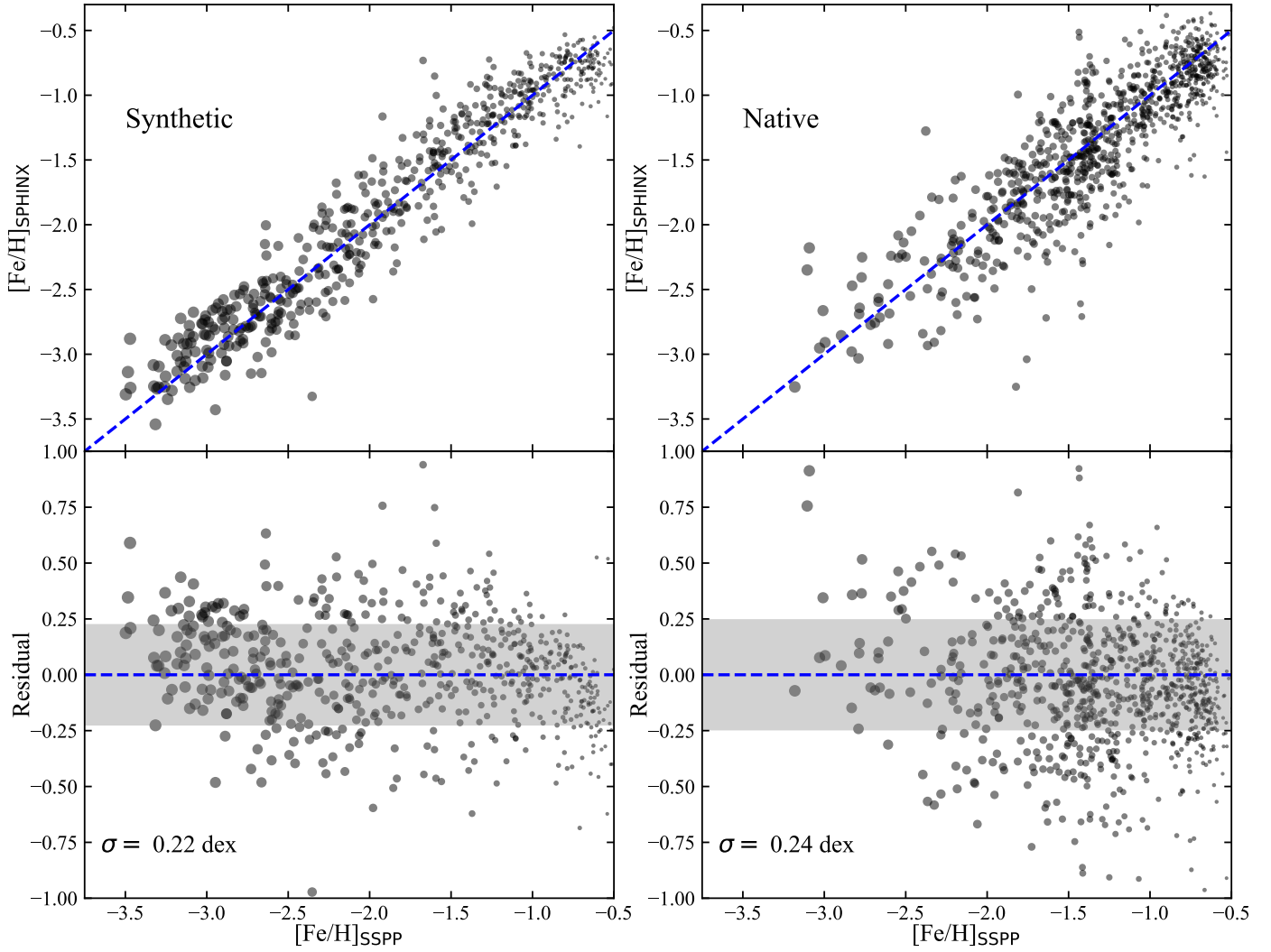
## 6. Metallicity determination with SPHINX

Two separate trials were conducted for validation of the metallicity routine in SPHINX. The maximum error on the biweight estimate of  $[\text{Fe}/\text{H}]$  from the SSPP was set to  $\pm 0.20 \text{ dex}$ . Similar to the temperature training, we insisted that the adopted and biweight estimates of metallicity from the SSPP did not differ by more than  $\pm 0.20 \text{ dex}$ . We refer the interested reader to Lee et al. (2008a) for an in-depth description of the biweight and adopted estimators from the SSPP. In short, the biweight is simply the robust average of all accepted estimates from the pipeline, while the adopted estimate takes into account an average of the biweight and a refined estimate that considers the reduced  $\chi^2$  of a synthetic spectrum match. In the event that the refined and biweight estimates do not differ significantly ( $< 0.15 \text{ dex}$ ), the adopted value is set to the biweight estimate. The reliability of the metallicity estimates from the SSPP was somewhat dependent on the S/N. We therefore set the minimum S/N to 40. The faint limit for all magnitudes was set to 18.5 to insure reliability in the synthetic magnitude calibrations.

For both trials, an array of 100 ANN units was constructed, where SPHINX was set to consider the five highest performing networks in each array. Anticipating the utility of the J0395 filter sensitivity to the Ca II H & K feature, we insisted on the use of J0395 photometry in each of the ANN units. The results of the metallicity estimates for both samples are shown in Fig. 8. Estimates of spread and central location in the residuals are provided in Table 4.

### 6.1. J-PLUS DR1 trial

The first trial implemented a testing set consisting entirely of J-PLUS DR1 photometry for stars in the range of  $4500 < T_{\text{eff}} (\text{K}) < 6200$ . The subset of the SDSS Reservoir with DR1



**Fig. 8.** Metallicity estimates,  $[\text{Fe}/\text{H}]$ , from SPHINX, as a function of the biweight estimate from the SSPP, for stars of effective temperature  $4500 < T_{\text{eff}} \text{ (K)} < 6200$ . The gray region corresponds to the standard deviation of the residuals, determined from a maximum-likelihood Gaussian fit. Results from SPHINX using synthetic magnitudes from the SDSS Reservoir (*left panels*) are shown with the results from native J-PLUS DR1 photometry (*right panels*) for comparison.

**Table 4.** Central location and scale estimates of photometric metallicity determinations.

Trial	$C_{\mu}$ (dex)	$C_M$ (dex)	$C_{\text{TRI}}$ (dex)	$C_{\text{BI}}$ (dex)	$S_{\sigma}$ (dex)	$S_{\text{MAD}}$ (dex)	$S_f$ (dex)	$S_{\text{BI}}$ (dex)
SDSS Reservoir	+0.02	+0.03	+0.03	+0.02	0.22	0.22	0.22	0.23
J-PLUS DR1	-0.06	-0.06	-0.05	-0.06	0.25	0.25	0.25	0.27

photometry in this range, after the error criteria described previously, was limited to 935 stars, which included only 28 stars with  $[\text{Fe}/\text{H}] < -2.5$ , of which five had  $[\text{Fe}/\text{H}] < -3.0$ . Therefore, all DR1 stars with  $[\text{Fe}/\text{H}] < -2.5$  were reserved for the testing set, and we supplemented the DR1 training photometry with 500 stars with  $[\text{Fe}/\text{H}] < -2.5$  and synthetic magnitudes from the SDSS Reservoir, of which 250 had  $[\text{Fe}/\text{H}] < -3.0$ . By doing so, the testing set consisted entirely of native J-PLUS photometry, while the training set maximized the number of low-metallicity stars in the sample. The surface gravity range for the testing set was  $1.3 < \log g < 4.8$ , with a median of 4.2 and standard deviation of 0.7 dex.

The median,  $C_M$ , and biweight central location,  $C_{\text{BI}}$ , of the residuals were both found to be  $-0.06$  dex for the DR1 trial, so the predictions were prone to slight underestimation when compared to the SSPP values. The Gaussian fit to the residuals revealed a standard deviation of  $S_{\sigma} = 0.25$  dex. The largest estimate of scale for the DR1 trial was the biweight estimate,  $S_{\text{BI}} = 0.27$  dex. The scatter in the metallicity was independent of surface gravity in the range of  $1.75 < \log g < 4.5$ . SPHINX tended to overestimate the metallicity for low surface gravity stars ( $\log g < 1.75$ ), for which the residual median increased to  $C_M = +0.14$  dex. However, these stars are correspondingly cool ( $T_{\text{eff}} \lesssim 4750$  K), and it was found that stars in the range

$4500 < T_{\text{eff}} \text{ (K)} < 4850$  tended to produce over-estimations regardless of the surface gravity.

For both the classification and recovery fractions of VMP stars with SPHINX, errors were expressed using the Wilson score approximation of the binomial confidence interval (Wilson 1927; Brown et al. 2001). Of the 81 VMP stars in the DR1 sample,  $85 \pm 3\%$  were recovered by the  $[\text{Fe}/\text{H}]$  estimate from SPHINX. SPHINX classified 109 stars in the DR1 sample as VMP, in which  $63 \pm 4\%$  were confirmed by the biweight estimate from the SSPP.

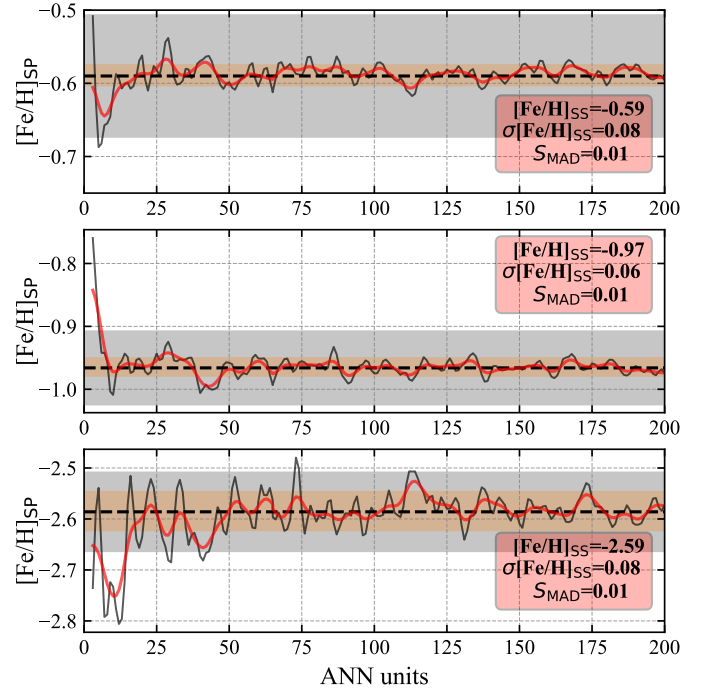
## 6.2. Synthetic SDSS trial

To explore the extent of current metallicity sensitivity with SPHINX, the second trial consisted of training and testing sets constructed from the SDSS Reservoir, using synthetic magnitudes. Effective temperature for both training and testing sets was again restricted to  $4500 < T_{\text{eff}} \text{ (K)} < 6200$ , with a surface gravity range of  $1.0 < \log g < 4.8$ . We included a wider distribution of surface gravities in the synthetic SDSS trial, with a median  $\log g$  of 3.8 and standard deviation of 0.78 dex, to further explore the potential influence on the final determinations. The training set consisted of 1986 stars, 590 of which had  $[\text{Fe}/\text{H}] < -2.5$ , with 269 having  $[\text{Fe}/\text{H}] < -3.0$ .

A slight overestimation was seen in the residuals of the synthetic SDSS trial, where the median and trimean estimates were both  $+0.03$  dex. Sensitivity was found to diminish below  $[\text{Fe}/\text{H}] < -3.0$ , where the residual median,  $C_M$ , increased to 0.17 dex. The standard deviation of the residuals was  $S_\sigma = 0.22$  dex, somewhat improved from the DR1 trial. As seen in Table 4, all estimates of scale were found to be smaller for the synthetic SDSS trial. Similar to the DR1 trial, scatter in the residuals was independent of surface gravity in the range  $1.5 < \log g < 5.0$ . Overestimation was also seen for cooler stars in SDSS trial, where the median residual increased to  $C_M = +0.32$  dex for stars of  $T_{\text{eff}} < 4750$  K. The synthetic SDSS trial was repeated while excluding the use of the  $J0395$  filter, resulting in a significant increase in scatter, where  $S_\sigma = 0.41$  dex and  $S_{\text{MAD}} = 0.43$  dex. We concluded that the  $J0395$  photometry is indeed a crucial component for metallicity sensitivity with SPHINX.

Of the 214 VMP stars in the synthetic SDSS trial,  $91 \pm 2\%$  were recovered, while  $93 \pm 2\%$  of the 209 stars classified as VMP by SPHINX were confirmed the SSPP estimate. While there is an insufficient number of EMP stars in the DR1 sample at present, we can investigate the classification and recovery fractions of EMP using the synthetic SDSS sample. Of the 43 stars determined to be  $[\text{Fe}/\text{H}] < -3.0$  by the SSPP, SPHINX recovered  $53 \pm 6\%$ . Of the 33 stars classified as EMP by SPHINX,  $70 \pm 6\%$  were confirmed by the SSPP estimate, while all remaining stars were VMP. These results are comparable to those obtained by the Pristine survey (Starkenburg et al. 2017).

To evaluate the performance of SPHINX as a function of the number of ANN units employed in the network array, another series of trials were run on the DR1 photometric set. We made use of the DR1 training catalog supplemented with synthetic VMP magnitudes, as described above. The number of ANN units was varied from three to 500, and the classification and recovery fractions for all cases was evaluated. We noted a small but consistent increase in the classification fraction, from 49 to 51% over the range of 3 to 100 ANN units. We found large oscillations in the recovery fractions over the full range of employed ANN units, but a general increase of a few percent. In the case of individual estimates, we found generally better convergence



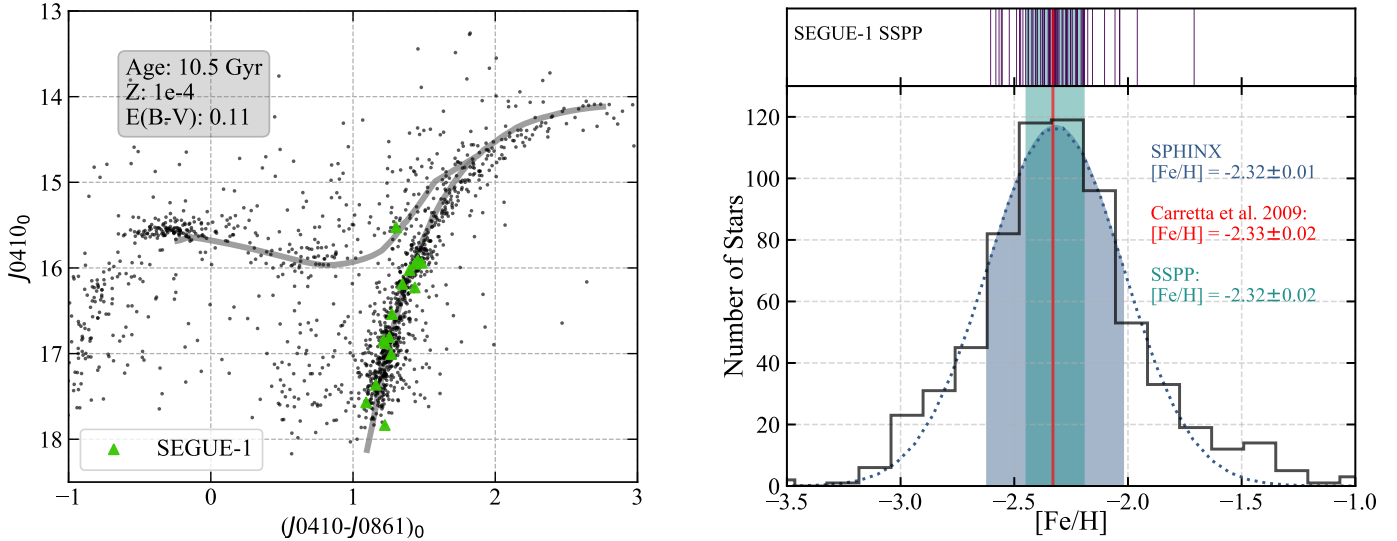
**Fig. 9.** Metallicity estimates from SPHINX for three example stars with DR1 photometry, as a function of the number of ANN units employed. The scale estimate,  $S_{\text{MAD}}$ , is represented by the orange band. In all three cases the scale is within the uncertainty in the spectroscopic value determined by the SSPP, shown in gray.

to a limiting value as the array size was increased. In many cases, estimates varied over 0.20 dex with smaller array size. This behavior is shown for three example stars with different metallicities in Fig. 9. In all cases, the variation in the estimates was seen to be stabilized after array sizes  $N > 50$ . For all three cases considered, the variation in the estimates of the photometric metallicities, determined from the MAD, was well within the uncertainty reported from the SSPP estimate.

## 7. A case study with M15

Photometric sources from the globular cluster M15 (Messier 15, NGC 7078) provide an opportunity to test both the accuracy and precision of metallicity estimates from SPHINX. Located at  $\sim 10.4$  kpc from the Sun, M15 is a particularly bright globular cluster with a horizontal-branch magnitude of  $V = 15.8$  and tidal radius of 21.5 arcmin (Harris 1996). M15 has a well-defined spectroscopically determined age of  $10.56 \pm 0.47$  Gy (Koleva et al. 2008) and metallicity of  $[\text{Fe}/\text{H}] = -2.33 \pm 0.02$  (Harris 1996; Carretta et al. 2009). Furthermore, with an intrinsic scatter in the metallicity of  $\sigma([\text{Fe}/\text{H}]) < 0.05$  dex, we regard this cluster as essentially mono-metallic (Carretta et al. 2009). The accuracy of SPHINX determinations is thereby investigated by considering the central location of the photometric metallicity estimates for the cluster, while the spread is reflective of the precision.

For this analysis, we made use of a stellar sample from Bonatto et al. (2019). Data were obtained as a science verification set prior to the J-PLUS Early Data Release. The  $1.4 \times 1.4 \text{ deg}^2$  field of the J-PLUS T80Cam was sufficiently large to contain the projected area of M15 in a single pointing. Photometry extended to a limiting magnitude of  $g \sim 21.5$  for over 40 000 stars. The photometric uncertainty at  $g \sim 21.5$  was estimated to be  $\sigma_m \sim 0.2$  however typical uncertainties for magnitudes within



**Fig. 10.** *Left panel:* color–magnitude diagram of the J-PLUS M15 stars. The population is over-plotted with a PARSEC isochrone corresponding to an age of 10.5 Gyr, and metallicity of  $Z = 0.0001$  (Bressan et al. 2012), with an assumed extinction of  $E(B - V) = 0.11$  mag. Of the 1437 stars in the sample, the 15 with SEGUE-1 medium-resolution spectra are denoted by green triangles. *Right panel:* distribution of photometric metallicity estimates from M15. A Gaussian distribution was fit to the photometric distribution, with the  $\sigma = 0.33$  dex region shaded. A stripe density plot is shown for the 66 confirmed M15 members with medium-resolution parameters from the SSPP. The shaded region in the density plot depicts the  $S_f$  of 0.13 dex.

the calibration limit for SPHINX ( $g \sim 18.79$ ) were less than 0.05 mag. Contrary to the standard procedure of SExtractor used by OAJ, DAOPHOT PSF (Stetson 1987) was used to extract photometry of stars for M15, as the standard pipeline is not currently able to cope with the level of stellar crowding typical of GCs, and DAOPHOT PSF can typically reach deeper photometric limits than SExtractor (Bonatto et al. 2019). Bonatto et al. (2019) applied a decontamination algorithm using color–magnitude diagrams, based on the work of Bonatto & Bica (2007), resulting in 1437 candidate member stars.

Our analysis included 98 stars with medium-resolution spectra from SEGUE-1. Stellar parameters and radial velocities for these stars were determined from the SSPP. For stars with high-quality estimates from the adopted and biweight procedures, effective temperatures and metallicities were taken to be the average of the two. Otherwise, a star was assigned the best of the two estimates, evaluated on the basis of a visual comparison with model synthetic spectra. Typical uncertainties in the radial velocity estimates were  $\pm 7.0$  km s $^{-1}$ . These stars were originally selected as candidates of M15 on the basis of their proximity to the center of the cluster and thus required a check for foreground and background contamination.

We rejected non-member SEGUE-1 stars on the basis of radial velocity. First, we computed the biweight location and scale for the distribution of radial velocities in the sample. With the tuning constant of the biweight estimator set at 6.0, the central location was found to be  $C_{BI} = -108.0$  km s $^{-1}$ , commensurate with a  $C_{TRI} = -107.9$  km s $^{-1}$ . For estimates of scale, we computed  $S_{BI} = 11.5$  km s $^{-1}$ ,  $S_f = 11.5$  km s $^{-1}$ . We rejected as outliers stars with radial velocities outside of the range  $C_{BI} \pm 2 S_{BI}$ . This process was repeated until no new stars were rejected; 66 stars remained. We recomputed the  $C_{BI}$  and  $S_{BI}$  again, and found  $-108.0$  and  $6.9$  km s $^{-1}$ , respectively. Finally, we compared these results with those determined by Pryor & Meylan (1993), on the basis of high-dispersion spectroscopy. They measured a mean velocity of  $-107.09 \pm 0.82$  km s $^{-1}$ , and a velocity dispersion of  $8.95 \pm 0.59$  km s $^{-1}$ , very similar to our values.

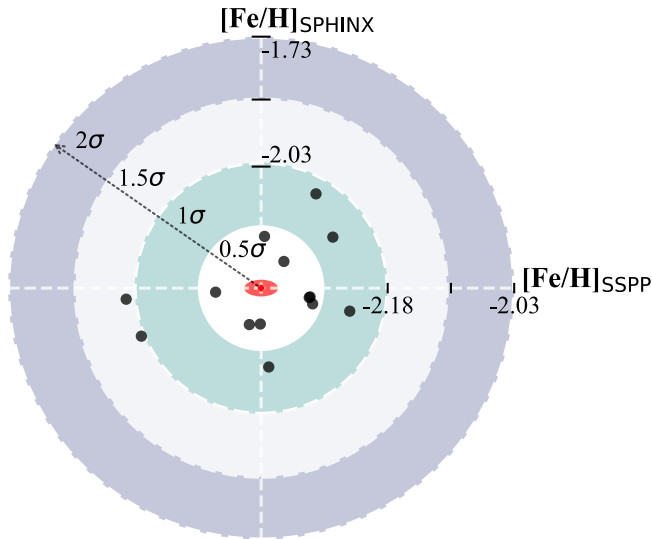
Of the 66 SEGUE-1 stars selected on the basis of radial velocity, a cross-match with the J-PLUS M15 catalog using a  $3''$  search radius identified 15 stars in common. These stars are shown in the left panel of Fig. 10, along with 1083 J-PLUS stars and a  $Z = 10^{-4}$ , 10.5 Gyr isochrone generated with the Padova and Trieste Stellar Evolution Code (PARSEC; Bressan et al. 2012). Here, the distance modulus for the cluster was taken to be 15.4 mag (Harris 1996). A correction of 0.3 mag was applied to the  $(J0410 - J0861)_0$  color prediction from PARSEC to align the isochrone with the giant branch seen in the M15 photometry. These spectroscopic targets largely occupied the giant branch, with one possible horizontal-branch star.

SPHINX was applied to J-PLUS photometry of the 1437 M15 stars with a limiting magnitude of  $g \sim 18.0$  using an array of 50 ANN units, and resulted in parameter estimates for 1041 sources (96% of the initial sample). The remaining 396 were rejected by the pipeline due to restrictions on the photometric errors, or extrapolation of all contributing networks. We excluded estimates produced with less than 25 of the available ANN units, after which 664 (61% of the initial sample) stars remained. The right panel in Fig. 10 compares the distribution of photometric  $[Fe/H]$  estimates to both the 66 SSPP medium-resolution parameters, as well as to the external spectroscopic value of  $[Fe/H] = -2.33 \pm 0.02$  for the cluster (Carretta et al. 2009). A central metallicity of  $[Fe/H] = -2.32 \pm 0.01$  was found from a Gaussian maximum-likelihood fit to the distribution of metallicities from SPHINX. The error estimate of the central metallicity,  $\pm 0.01$  dex, was determined from a bootstrap procedure, for which the 664 stars were randomly sampled with replacement for 1000 trials. The standard deviation of the residuals for the photometric estimates was determined to be 0.29 dex. We compared this to the median value of  $[Fe/H] = -2.31 \pm 0.02$  for the 66 SEGUE-1 stars with SSPP parameters, for which the error in the median was determined from the bootstrap procedure. The  $f$  spread of the SSPP metallicity distribution was 0.17 dex, corresponding to  $S_f = 0.13$  dex.

We consider the accuracy of the photometric metallicity estimates for the cluster in Fig. 11. With the  $\sigma$  estimate from the

**Table 5.** Spectroscopic stellar parameters of 13 SEGUE-1 + J-PLUS stars in M15.

RA (hh:mm:ss)	Dec (dd:mm:ss)	$T_{\text{SP}}$ (K)	$\sigma(T_{\text{SP}})$ (K)	$T_{\text{SS}}$ (K)	$\sigma(T_{\text{SS}})$ (K)	$[\text{Fe}/\text{H}]_{\text{SP}}$	$\sigma([\text{Fe}/\text{H}]_{\text{SP}})$ (dex)	$[\text{Fe}/\text{H}]_{\text{SS}}$	$\sigma([\text{Fe}/\text{H}]_{\text{SS}})$ (dex)
21:29:37.1	12:13:41	5570	93	5447	75	-2.44	0.24	-2.47	0.24
21:29:39.4	12:18:21	5411	84	5299	44	-2.34	0.26	-2.38	0.10
21:29:43.1	12:00:18	5690	55	5521	60	-2.21	0.17	-2.25	0.13
21:29:48.5	12:20:20	5510	81	5315	37	-2.37	0.28	-2.27	0.16
21:29:49.9	12:18:12	5288	93	5206	48	-2.42	0.25	-2.34	0.17
21:29:52.3	12:19:40	5128	108	5215	48	-2.24	0.63	-2.27	0.20
21:30:00.7	12:22:32	5568	136	5320	33	-2.35	0.27	-2.27	0.16
21:30:13.5	12:00:38	5523	91	5390	36	-2.42	0.22	-2.33	0.18
21:30:15.2	12:19:46	5352	114	5269	50	-2.21	0.26	-2.33	0.18
21:30:24.5	12:02:38	5592	130	5420	33	-2.95	0.79	-2.35	0.18
21:30:35.2	12:18:10	5626	117	5314	32	-2.36	0.26	-2.49	0.17
21:30:49.3	12:07:31	5532	114	5344	31	-2.52	0.26	-2.32	0.14
21:30:54.4	12:07:11	5569	121	5419	45	-2.39	0.22	-2.23	0.21



**Fig. 11.** Metallicity estimates for 13 stars in the M15 cluster with both SEGUE-1 spectra and J-PLUS photometry. Estimates are shown above the  $1\sigma$  and  $2\sigma$  regions corresponding to the standard deviation of SPHINX estimates for the entire cluster, and the median error estimate from the SSPP. The small red ellipse in the center represents the intrinsic scatter of the M15 cluster,  $[\text{Fe}/\text{H}] = -2.33 \pm 0.02$  (Carretta et al. 2009).

Gaussian distribution of 0.30 dex and the uncertainty from the SSPP of 0.13 dex, we depict the individual estimates for 13 stars of the original 15 with both photometric and spectroscopic estimates. The elliptical regions of Fig. 11 corresponding to 0.5, 1.0, 1.5, and  $2\sigma$  from the center value of  $[\text{Fe}/\text{H}] = -2.33$  for the cluster were vertically scaled to a circular region for simplicity. We excluded one star from this sample due to a large error estimate from SPHINX ( $[\text{Fe}/\text{H}]_{\text{SP}} = -2.94 \pm 0.79$ ). Another star, thought originally to occupy the horizontal branch, was found to have a large deviation between the photometric and spectroscopic estimates. We therefore excluded this estimate on the basis of potential error in the photometry. Of the 13 stars considered with SSPP and SPHINX metallicity estimates, 11 (85%) fall within the  $1\sigma$  region. The remaining two stars each lie within the  $1.5\sigma$  region. Both the median and  $C_{\text{BI}}$  computed for the residuals were found to be  $-0.02$  dex. These 13 sources with spectroscopic

parameters are provided in Table 5. We use the SP subscript to denote values determined by SPHINX, while SS refers to spectroscopically determined estimates from the SSPP.

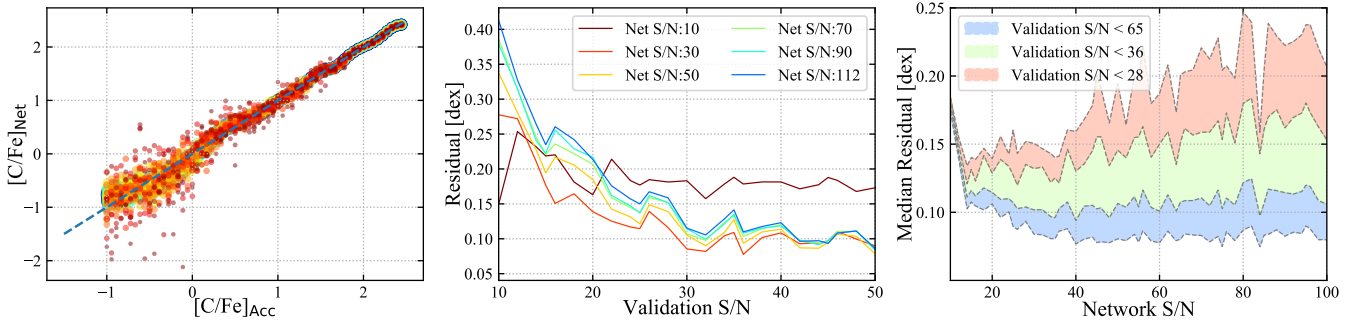
J-PLUS photometry of M15 provides an opportunity to validate the effective temperature and metallicity routines on a single stellar population. While determinations of metallicity made with SPHINX exhibit a scatter of  $\sim 0.26$  dex, we are able to make estimations of central metallicity for M15 to remarkably high accuracy, in addition to estimates of photometric effective temperature that are commensurate with estimates from the SSPP. In fact, the uncertainty in the mean of the metallicity distribution generated from estimates using J-PLUS photometry is within the uncertainty in the value determined from high-resolution spectra. While the capability of SPHINX is currently limited with respect to individual precision measurements, SPHINX enables studies of stellar populations as well.

## 8. Preliminary results on carbonicity

While SPHINX is presently limited to obtaining effective temperature and metallicity determinations, we anticipate the addition of carbonicity ( $[\text{C}/\text{Fe}]$ ) and surface gravity ( $\log g$ ) estimates in the near future. This will require more stars with J-PLUS photometry for proper training on their carbon features. However, with our library of synthetic spectra, we are able to explore the validity of a neural network approach to carbon detection. The application of the synthetic library also permits an in-depth study of the influence of underlying S/N on network training and estimation.

To isolate the carbon dependence on our synthetic spectra, we restricted our analysis to synthetic spectra corresponding to a giant VMP star,  $T_{\text{eff}} = 5000$  K,  $\log g = 1.00$ ,  $[\text{Fe}/\text{H}] = -2.50$ . Next, 100 spectra were generated by way of cubic spline interpolation across the interval  $-1.0 < [\text{C}/\text{Fe}] < +2.5$ . These normalized spectra were convolved with a blackbody to best approximate a flux-calibrated spectrum.

The flux of each spectrum was randomly rescaled from 50 to 150% to best emulate a random distribution of distance moduli, then injected with Gaussian noise, resulting in 66 batches of the original 100 spectra from  $10 < S/N < 122$ , each batch consisting of a unique global S/N. Synthetic magnitudes for the narrowband J-PLUS filters were computed in the same manner as for the SDSS Reservoir, by convolving the noise-injected,



**Fig. 12.** Results of synthetic carbon training for various S/N implementations. *Left panel:*  $[C/Fe]$  estimates from the network trained on  $S/N = 122$  inputs; the color of the estimates are proportional to the S/N of the validation photometry, where red indicates lower S/N. *Middle panel:*  $[C/Fe]$  residuals for networks of various S/N are shown, as a function of the S/N for the validation sets employed. *Right panel:* median residuals for three tiers of validation sets,  $S/N < 28$ ,  $S/N < 36$ , and  $S/N < 65$ , are shown, as a function of the network S/N.

blackbody-calibrated synthetic spectrum with the appropriate filter-response function. Linear scaling was then performed to center each input magnitude distribution.

For simplicity, SPHINX was limited to a single ANN unit for each trial described below. For each trial, one of the 66 synthetic spectra batches was used to train the ANN unit, consisting of six narrowband inputs –  $J0378$ ,  $J0395$ ,  $J0430$ ,  $J0515$ ,  $J0660$ , and  $J0861$  – with a single hidden layer of five hyperbolic tangent neurons and a stochastic gradient-descent optimization. The result was an array of identical networks, each trained on a set of inputs with a specific S/N, which we refer to as the network S/N. These networks of unique S/N were then tested on each batch of noise-injected synthetic spectra, in order to evaluate the performance of each version of the network with an approximation of photometry of varying quality. In doing so, we assigned a median residual to each network and its constituent 66 verification batch runs. The result of this analysis is shown in Fig. 12.

The left panel of Fig. 12 shows the results of the  $S/N = 122$  network for all batches of synthetic spectra. We find a characteristic behavior, namely, that photometry of lower S/N results in a greater dispersion of the predicted  $[C/Fe]$  about the true value. The dispersion is more prevalent in the lower carbonicity regime ( $[C/Fe] < 0.0$ ), as might be expected, since the carbon-sensitive features at low  $[C/Fe]$  are weaker, and so are generally more influenced by the S/N across the spectrum.

The middle panel of Fig. 12 shows the behavior of each network, as a function of the S/N in the photometry in the validation set. First, there is a general trend of a reduction in the spread of the residuals as the S/N of the validation photometry increases. Interestingly, in the input range of  $S/N < 20$ , networks trained on  $S/N = 10$  and  $S/N = 20$  outperform those trained on higher-quality photometry, for which residuals exceeded a standard deviation of 0.30 dex. We conclude that, in general, a network trained on poor photometry will exhibit better performance on poor photometry.

The right panel of Fig. 12 tracks the performance of the network for different regions of the validation S/N, as a function of the network's S/N. For sources of low quality ( $S/N < 28$ ), the median residual is worse overall for all values of the network S/N. The low-quality source residuals tend to increase as the networks are trained on higher-quality photometry. In the region of  $20 \leq S/N \leq 30$ , we observe no significant increase in the median residual of the low-quality source set (median residual  $\sim 0.13$  dex), while the high-quality set relaxes to a roughly consistent value of  $\sim 0.08$  dex. This suggests an optimal region of S/N for network training, approximately  $30 < S/N < 40$ , to ensure

high-confidence performance for sources of both high and low quality, for network applications of this nature.

## 9. Discussion and conclusion

The Stellar Photometric Index Network Explorer, SPHINX, is designed to estimate the effective temperature and metallicity of stellar atmospheres using broad- and intermediate-band optical J-PLUS photometry, with a particular emphasis on the capacity to identify low-metallicity stars. This pipeline attempts to optimize training databases provided to converge its constituent artificial neural network units, as well as the relative weight assigned to these subordinate ANN units for use in science estimates. By doing so, SPHINX has the potential to be quite flexible in its ability to accommodate a variety of photometric data.

Estimates of effective temperature made using J-PLUS Early Data Release were found to be successful across a temperature range of  $4500 < T_{eff} \text{ (K)} < 8500$ , with an uncertainty of  $\pm 91$  K. Comparisons were made to previous calibrations by Lee et al. (2008a) and Fukugita et al. (2011), and in all cases estimates by SPHINX with J-PLUS photometry proved superior to the broadband photometric methods, the J-PLUS analogs to SDSS photometry were used. We emphasize the success of temperature estimates made with SPHINX without the use of a priori knowledge of surface gravity, for which we find no significant influence on temperature determinations within  $2.0 < \log g < 5.0$ .

Photometric estimates of metallicity for stars in the J-PLUS First Data Release within  $4500 < T_{eff} \text{ (K)} < 6200$  indicate sensitivity down to  $[Fe/H]$  approximately  $-3.0$ , with a scatter of  $\sigma([Fe/H]) = 0.25$  dex. However, verification of this sensitivity is limited by the lack of numerous  $[Fe/H] < -3.0$  stars in the footprint of J-PLUS at present. We find that metallicity determinations made with J-PLUS DR1 photometry are not influenced by surface gravity within the range  $1.75 < \log g < 5.0$ . SPHINX is very effective as a means for recovering and correctly identifying significant fractions of low-metallicity stars. SPHINX recovers  $85 \pm 3\%$  of very metal-poor (VMP) targets in the DR1 testing set used, with  $63 \pm 4\%$  of VMP candidates correctly classified. For the trial consisting of synthetic magnitudes from the SDSS Reservoir,  $70 \pm 6\%$  of stars designated as  $[Fe/H] < -3$  were confirmed by the SSPP, while all remaining stars were VMP. Of the 43 EMP stars in the synthetic SDSS trial,  $53 \pm 6\%$  were recovered by SPHINX.

Photometric metallicity estimates were made using J-PLUS photometry of 664 sources associated with globular cluster M15 (NGC 7078). The central value obtained,  $[Fe/H] = -2.32 \pm 0.01$ ,



with a residual spread of 0.29 dex, was commensurate with the value determined from SEGUE-1 medium-resolution spectra of  $-2.32 \pm 0.02$  for the cluster. These estimates essentially match the accepted value from the literature,  $-2.33 \pm 0.02$  (Harris 1996; Carretta et al. 2009). The accuracy of estimates made by SPHINX was, in part, due to the level of optimization made possible by way of adjustable neural network architecture and the ability to tailor the training set used to converge the subordinate ANN units. This enables SPHINX to make parameter estimates for science cases involving stars of a single evolutionary stage, or to generalize to accommodate field stars in a variety of evolutionary stages across large regions of sky.

The development of SPHINX is at present limited to effective temperature and metallicity estimates, but we anticipate the addition of carbon abundance estimates, at least for some portion of the stellar parameter space. With carbon-abundance estimates from SPHINX, the primary sub-classes of carbon-enhanced metal-poor (CEMP) stars, the CEMP-*s* and CEMP-no stars, describing the nature of their neutron-capture element abundance ratios (Beers & Christlieb 2005), could be discerned from photometry alone (Yoon et al. 2016), avoiding the requirement for obtaining far-more time-consuming high-resolution spectroscopic follow-up. In addition, with accurate parallaxes from *Gaia* DR2 (Gaia Collaboration 2016), improved photometric maps of carbon abundances for the inner- and outer-halo regions of the Milky Way are readily obtainable (see spectroscopic-based maps in Lee et al. 2013).

In the future, we plan to provide photometric parameters for a substantial portion of stars in the J-PLUS footprint. However, at present, a pure catalog of stellar sources from J-PLUS is not yet available – present catalogs are contaminated by quasars at faint magnitudes. Low-redshift quasars can be identified and removed by straightforward PSF analysis and the use of color–color diagrams (Caballero et al. 2008). However, high-redshift quasars remain a challenge, as they are optically similar to ultracool dwarfs.

The artificial neural network methodology of SPHINX, in conjunction with J-PLUS photometry, is an ideal tool for selecting low-metallicity targets for spectroscopic follow-up. In addition, reasonable sensitivity has been demonstrated for individual estimates of metallicity down to  $[\text{Fe}/\text{H}]$  approximately  $-3.0$  for stars in the range  $4500 < T_{\text{eff}} \text{ (K)} < 6200$ . We anticipate that future data releases from J-PLUS will enable an expanded sensitivity to low-metallicity stars. Ultimately, we expect to substantially increase the number of known VMP, EMP, UMP, and CEMP stars in the coming years on the basis of such studies.

*Acknowledgements.* D.D.W., V.M.P., and T.C.B. acknowledge partial support from Grant No. PHY-1430152 (Physics Frontier Center/JINA Center for the Evolution of the Elements), awarded by the U.S. National Science Foundation. Based on observations made with the (JAST/T80 and JST/T250) telescope(s) at the Observatorio Astrofísico de Javalambre, in Teruel, owned, managed, and operated by the Centro de Estudios de Física del Cosmos de Aragón. We thank the OAJ Data Processing and Archiving Unit (UPAD) for reducing and calibrating the OAJ data used in this work. Funding for the J-PLUS Project has been provided by the Governments of Spain and Aragón through the Fondo de Inversiones de Teruel, the Spanish Ministry of Economy and Competitiveness (MINECO); under grants AYA2016-79425-C3-2-P, AYA2015-66211-C2-1-P, AYA2015-66211-C2-2, AYA2012-30789 and ICTS-2009-14), and European FEDER funding (FCDD10-4E-867, FCDD13-4E-2685). The Brazilian agencies FAPESP and the National Observatory of Brazil have also contributed to this project. S.A. acknowledges support of CNPq, through postdoctoral fellowship (project 300336/2016-0). M.V.C.D. thanks the FAPESP scholarship process number 2014/18632-6. S.R. acknowledges financial support from CAPES, CNPq, and FAPESP. T.M. acknowledges support provided by the Spanish Ministry of Economy and Competitiveness (MINECO), under grant AYA-2017-88254-P. R.L.O. was partially supported by the Brazilian agency CNPq (Universal Grants 459553/2014-3, PQ 302037/2015-2, and PDE 200289/2017-9). C.M.O.

and L.S.J. acknowledge the support from FAPESP and CNPq. Y.S.L. acknowledges support from the National Research Foundation (NRF) of Korea grant funded by the Ministry of Science and ICT (No.2017R1A5A1070354 and NRF-2018R1A2B6003961). R.A.D. acknowledges support from CNPq through BP grant 312307/2015-2, CSIC through grant COOPB20263, FINEP grants REF. 1217/13 – 01.13.0279.00 and REF 0859/10 – 01.10.0663.00 for partial hardware support for the J-PLUS project through the National Observatory of Brazil. P.C. acknowledges support from CNPq 305066/2015-3. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the US Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is [www.sdss.org](http://www.sdss.org). SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

## References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543
- Abolhalima, A., & Frebel, A. 2018, *ApJS*, **238**, 36
- Aguado, D. S., Allende Prieto, C., González Hernández, J. I., & Rebolo, R. 2018, *ApJ*, **854**, L34
- Allende Prieto, C., Sivarani, T., Beers, T. C., et al. 2008, *AJ*, **136**, 2070
- Álvarez, R., & Plez, B. 1998, *A&A*, **330**, 1109
- An, D., Beers, T. C., Johnson, J. A., et al. 2013, *ApJ*, **763**, 65
- An, D., Beers, T. C., Santucci, R. M., et al. 2015, *ApJ*, **813**, L28
- Anthony-Twarog, B. J., Twarog, B. A., Laird, J. B., & Payne, D. 1991, *AJ*, **101**, 1902
- Beers, T. C., & Christlieb, N. 2005, *ARA&A*, **43**, 531
- Beers, T. C., Flynn, K., & Gebhardt, K. 1990, *AJ*, **100**, 32
- Benitez, N., Dupke, R., Moles, M., et al. 2014, ArXiv e-prints [arXiv:1403.5237]
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Boggs, P., & Rogers, J. 1990, *Contemp. Math.*, **112**, 186
- Bonatto, C., & Bica, E. 2007, *MNRAS*, **377**, 1301
- Bonatto, C., Chies-Santos, A. L., Coelho, P. R. T., & J-PLUS Collaboration 2019, *A&A*, **622**, A179
- Bonifacio, P., Caffau, E., Spite, M., et al. 2015, *A&A*, **579**, A28
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, **327**, 977
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
- Brown, L. D., Cai, T. T., & DasGupta, A. 2001, *Stat. Sci.*, **16**, 101
- Caballero, J. A., Burgasser, A. J., & Klement, R. 2008, *A&A*, **488**, 181
- Canterna, R. 1976, *AJ*, **81**, 228
- Carollo, D., Beers, T. C., Lee, Y. S., et al. 2007, *Nature*, **450**, 1020
- Carollo, D., Beers, T. C., Chiba, M., et al. 2010, *ApJ*, **712**, 692
- Carretta, E., Bragaglia, A., Gratton, R., D’Orazi, V., & Lucatello, S. 2009, *A&A*, **508**, 695
- Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., & J-PLUS Collaboration 2019, *A&A*, **622**, A176
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, **142**, 72
- Fabbro, S., Venn, K. A., O’Brian, T., et al. 2018, *MNRAS*, **475**, 2978
- Fitzpatrick, E. L. 1999, *PASP*, **111**, 63
- Frebel, A., Ji, A. P., Ezzeddine, R., et al. 2019, *ApJ*, **871**, 146
- Fukugita, M., Yasuda, N., Doi, M., Gunn, J. E., & York, D. G. 2011, *AJ*, **141**, 47
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, *MNRAS*, **478**, 651
- Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, *A&A*, **486**, 951
- Harris, W. E. 1996, *AJ*, **112**, 1487
- Hon, M., Stello, D., & Yu, J. 2017, *MNRAS*, **469**, 4578
- Hon, M., Stello, D., & Yu, J. 2018, *MNRAS*, **476**, 3233

- Ibata, R. A., McConnachie, A., Cuillandre, J.-C., et al. 2017a, *ApJ*, **848**, 128
- Ibata, R. A., McConnachie, A., Cuillandre, J.-C., et al. 2017b, *ApJ*, **848**, 129
- Ivezić, Z., Sesar, B., Jurić, M., et al. 2008, *ApJ*, **684**, 287
- Keller, S. C., Schmidt, B. P., Bessell, M. S., et al. 2007, *PASA*, **24**, 1
- Keller, S. C., Bessell, M. S., Frebel, A., et al. 2014, *Nature*, **506**, 463
- Kheirdastan, S., & Bazarghan, M. 2016, *Ap&SS*, **361**, 304
- Kim, E. J., & Brunner, R. J. 2017, *MNRAS*, **464**, 4463
- Kingma, D. P., & Ba, J. 2014, ArXiv e-prints [arXiv:1412.6980]
- Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, *MNRAS*, **385**, 1998
- Le Cun, Y., Kanter, I., & Solla, S. 1991, *Adv. Neural Inf. Process Syst.*, **3**, 918
- Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008a, *AJ*, **136**, 2022
- Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008b, *AJ*, **136**, 2050
- Lee, Y. S., Beers, T. C., Masseron, T., et al. 2013, *AJ*, **146**, 132
- Logroño-García, R., Vilella-Rojo, G., López-Sanjuan, C., & J-PLUS Collaboration 2019, *A&A*, **622**, A180
- López-Sanjuan, C., Vázquez Ramió, H., Varela, J., & J-PLUS Collaboration 2019, *A&A*, **622**, A177
- Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, *AJ*, **118**, 1406
- Molino, A., Costa-Duarte, M. V., Mendes de Oliveira, C., & J-PLUS Collaboration 2019, *A&A*, **622**, A178
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Placco, V. M., Frebel, A., Lee, Y. S., et al. 2015, *ApJ*, **809**, 136
- Placco, V. M., Frebel, A., Beers, T. C., et al. 2016, *ApJ*, **833**, 21
- Plez, B. 2012, Astrophysics Source Code Library [record ascl:1206.013]
- Pryor, C., & Meylan, G. 1993, in *Structure and Dynamics of Globular Clusters*, eds. S. G. Djorgovski & G. Meylan, *ASP Conf. Ser.*, **50**, 357
- Riedmiller, M. 1994, *Comput. Stand. Interfaces*, **16**, 265
- Rosenblatt, F. 1958, *Psychol. Rev.*, **65**
- Ryabchikova, T., Piskunov, N., Kurucz, R. L., et al. 2015, *Phys. Scr.*, **90**, 054005
- Salvadori, S., Ferrara, A., Schneider, R., Scannapieco, E., & Kawata, D. 2010, *MNRAS*, **401**, L5
- San Roman, I., Sánchez-Blázquez, P., Cenarro, A. J., & J-PLUS Collaboration 2019, *A&A*, **622**, A181
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, **737**, 103
- Schlaufman, K. C., & Casey, A. R. 2014, *ApJ*, **797**, 13
- Spite, M., Caffau, E., Bonifacio, P., et al. 2013, *A&A*, **552**, A107
- Starkenburger, E., Martin, N., Youakim, K., et al. 2017, *MNRAS*, **471**, 2587
- Starkenburger, E., Aguado, D. S., Bonifacio, P., et al. 2018, *MNRAS*, **481**, 3838
- Stetson, P. B. 1987, *PASP*, **99**, 191
- Strömberg, B. 1963, *QJRAS*, **4**, 8
- Strömberg, B. 1964, *Astrophys. Norvegica*, **9**, 333
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, *ApJ*, in press, [arXiv:1804.01530]
- Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, *ApJ*, **750**, 99
- Wilson, E. B. 1927, *J. Am. Stat. Assoc.*, **22**, 209
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, **137**, 4377
- Yoon, J., Beers, T. C., Placco, V. M., et al. 2016, *ApJ*, **833**, 20
- York, D. G., Adelman, J., Anderson, Jr. J. E., et al. 2000, *AJ*, **120**, 1579