

# An application of deep learning in the analysis of stellar spectra

S. Fabbro,<sup>1</sup>[★](#) K. A. Venn,<sup>2</sup>[★](#) T. O'Briain,<sup>1,2</sup> S. Bialek,<sup>1,2</sup> C. L. Kielty,<sup>2</sup> F. Jahandar<sup>2</sup> and S. Monty<sup>2</sup>

<sup>1</sup>National Research Council of Canada, Herzberg Astronomy & Astrophysics Program, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada

<sup>2</sup>Department of Physics and Astronomy, University of Victoria, Victoria, BC V8W 3P2, Canada

Accepted 2017 December 18. Received 2017 December 18; in original form 2017 August 6

## ABSTRACT

Spectroscopic surveys require fast and efficient analysis methods to maximize their scientific impact. Here, we apply a deep neural network architecture to analyse both SDSS-III APOGEE DR13 and synthetic stellar spectra. When our convolutional neural network model (StarNet) is trained on APOGEE spectra, we show that the stellar parameters (temperature, gravity, and metallicity) are determined with similar precision and accuracy as the APOGEE pipeline. StarNet can also predict stellar parameters when trained on synthetic data, with excellent precision and accuracy for both APOGEE data and synthetic data, over a wide range of signal-to-noise ratios. In addition, the statistical uncertainties in the stellar parameter determinations are comparable to the differences between the APOGEE pipeline results and those determined independently from optical spectra. We compare StarNet to other data-driven methods; for example, StarNet and the Cannon 2 show similar behaviour when trained with the same data sets; however, StarNet performs poorly on small training sets like those used by the original Cannon. The influence of the spectral features on the stellar parameters is examined via partial derivatives of the StarNet model results with respect to the input spectra. While StarNet was developed using the APOGEE observed spectra and corresponding ASSET synthetic data, we suggest that this technique is applicable to other wavelength ranges and other spectral surveys.

**Key words:** methods: numerical – techniques: spectroscopic – surveys – stars: fundamental parameters – infrared: stars.

## 1 INTRODUCTION

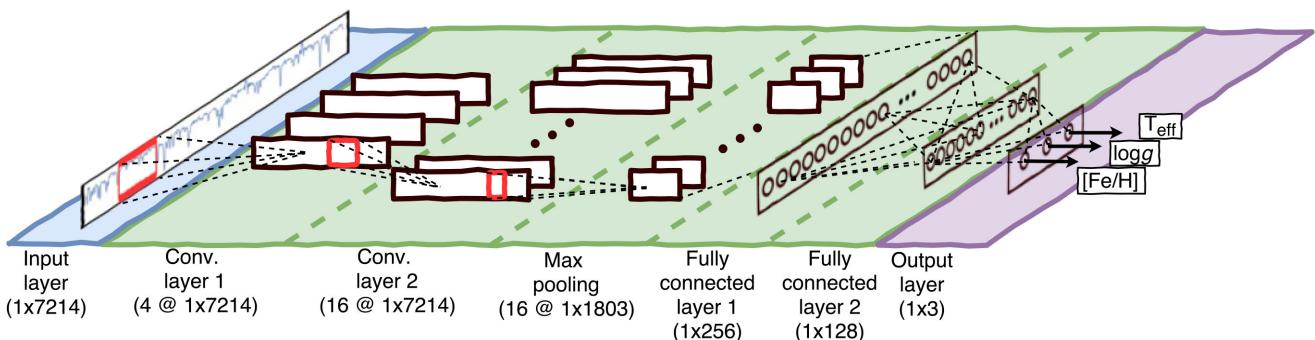
Spectroscopic surveys provide a homogeneous data base of stellar spectra that are ideal for machine learning applications. A variety of techniques have been studied in the past two decades. These range from the SDSS SEGUE stellar parameter pipeline using a decision tree architecture between spectral matching to a synthetic grid and line index measurements (e.g. Allende-Prieto et al. 2008; Lee et al. 2008; Yanny et al. 2009; Lee et al. 2011) to the detailed algorithms employed for analysis of the *Gaia* spectroscopic data (e.g. Smiljanic et al. 2014; Recio-Blanco 2016, 2017; Pancino et al. 2017).

The use of artificial neural networks (NNs) in astrophysical applications has a history going back more than 20 yr, with pioneering research in stellar classification by authors such as Von Hippel et al. (1994) and Singh, Gulati & Gupta (1998). In Bailer-Jones et al. (1997) and Bailer-Jones (2000), an NN was applied to synthetic stellar spectra to predict the effective temperature  $T_{\text{eff}}$ , surface gravity  $\log g$ , and metallicity [Fe/H].

More recently, dramatic improvements have occurred in the usability and performance of algorithms implemented in machine learning software. This, combined with the increase in computing power and the availability of large data sets, has led to the successful implementation of more complex NN architectures. These have proven to be pivotal in difficult image recognition tasks and natural language processing. The earlier attempts at NNs for stellar spectra analyses (e.g. Bailer-Jones 2000; Manteiga et al. 2010) would train the NN on synthetic spectra and test the model on synthetic spectra. Machine learning methods were also used in one of the SEGUE pipelines (Lee et al. 2008), where two NNs were trained: one on synthetic spectra and the other on previous SEGUE parameters.

In this paper, we examine several cases of training on synthetic spectra and predicting stellar parameters on both synthetic and observed spectra, or training and predicting on only observed spectra in a purely data-driven approach. The method we use to follow is the supervised learning approach, where a representative subset of stellar spectra (either synthetic or observed) with known stellar parameters is selected. This subset can be further divided into a reference set, where the NN learns the mapping function from spectra to stellar parameters, and a test set, used for ensuring the accuracy of the predictions. Once the network model is trained (i.e. with fixed

\* E-mail: [sebastien.fabbro@nrc-cnrc.gc.ca](mailto:sebastien.fabbro@nrc-cnrc.gc.ca) (SF); [kvenn@uvic.ca](mailto:kvenn@uvic.ca) (KAV)



**Figure 1.** The current StarNet CNN model composed of seven layers. The first layer is solely the input data; followed by two convolutional layers with 4 and 16 filters (in successive order), then a max pooling layer with a window length of four units, followed by three fully connected layers with 256, 128, and 3 nodes (again, in successive order). The final layer is the output layer.

network parameters), the stellar parameters can be predicted for the rest of the sample. The stellar parameters we consider for this spectral analysis are the effective temperature ( $T_{\text{eff}}$ ), surface gravity ( $\log g$ ), and metallicity ( $[\text{Fe}/\text{H}]$ ).

In this paper, we present StarNet: a convolutional neural network (CNN) model applied to the analysis of stellar spectra. We introduce our machine learning methods in Section 2 and evaluate our model for a set of synthetic data in Section 3. As an exercise of its effectiveness, we apply StarNet to the APOGEE survey in Section 4 (DR13 and earlier data releases when appropriate) and compare the stellar parameters to those from the APOGEE pipeline(s). In Section 5, we discuss the success of our StarNet results to other stellar analyses and confirm that NNs can significantly increase the robustness, efficiency, and scientific impact of spectroscopic surveys.

## 2 MACHINE LEARNING METHODOLOGY

Supervised learning has been shown to be well adapted for continuous variable regression problems. Given a training set in which, for each input spectrum, there are known stellar parameters, a supervised learning model is then capable of approximating a function that transforms the input spectra to the output values. The learned function can then ideally be used to predict the output values of a different data set. The particular form of this function and how it is learned depends on the NN architecture. Summarized below is the CNN that we have implemented for the analysis of stellar spectra; we provide more details about deep NNs and the mathematical operations of our selected architecture in Appendix A.

### 2.1 The StarNet CNN

An NN can be arranged in layers of artificial neurons: an *input layer* (i.e. the input data), a number of *hidden layers*, and an *output layer*. Depending on the architecture of the NN, there may be several hidden layers, each composed of multiple neurons that weight and offset the outputs from the previous layer to compute input values for the following layer. The more hidden layers used in the NN, the *deeper* the model is. The combination of these layers act as a single function, and in the case of StarNet, this function predicts three stellar parameters.

Inspired from recent studies of deep NNs on stellar spectra (Li, Pan & Duan 2017; Wang, Guo & Luo 2017), we have focused our analysis on deep architectures. The CNN selected for StarNet is shown schematically in Fig. 1. This architecture is composed of a combination of fully connected layers and one-dimensional (1D)

convolutional layers. Fully connected layers are the classical NN layers that compute multiple linear combinations of all of the input values to produce an output vector. In the case of a convolutional layer, a series of filters are applied, extracting local information from the previous layer. Throughout the training phase the network learns the filters that are activated most strongly when detecting specific features, thus producing a collection of *feature maps*. Using two successive convolutional layers, the second of the two convolves across the previous layer's feature map, which allows the model to learn higher order features.

The combination of convolutional layers and fully connected layers in our StarNet implementation means that the output parameters are not only affected by individual features in the input spectrum, but also, combinations of features found in different areas of the spectrum are utilized. This technique strengthens the ability of StarNet to generalize its predictions on spectra with a wide range of signal-to-noise ratios (S/Ns) across a larger stellar parameter space. More details of the StarNet model itself are discussed in Section 4.3.

### 2.2 Training and testing of the model

Before training the model, the reference set is split into a training set and a cross-validation set. Training is performed through a series of forward and back propagation iterations, using batches of the training set. The forward propagation is the model function itself: at each layer, weights are applied to all of the input values, and at the output layer, a prediction is computed. These predictions are then compared to the target values through a *loss function*. In our StarNet model, a mean-squared-error (MSE) loss function computes the loss between predictions and target values and a total loss for all training examples is minimized.

Initially, the weights of the model are randomly set and therefore the predictions will be quite poor. To improve these predictions, the model weights are updated following each batch forward propagation. Therefore, the weights are adjusted multiple times per *iteration*. A minimum was usually reached after 20–25 iterations (although this could vary greatly depending on the complexity of the model used). The training of our StarNet model reached convergence in an amount of time that depended on the size of the training set and the model architecture; for  $\sim 41\,000$  spectra (close to 300 million intensity values) and 7.5 million NN parameters, the training converged in 30 min, using a 16 cores virtual machine.

When the training stage reaches convergence, the weight values are frozen. The estimated model is evaluated on a test set of spectra with a wide range of S/Ns. All tests of StarNet are quantified with

**Table 1.** Stellar parameter distribution of the ASSET synthetic spectra grid.

| Class | $T_{\text{eff}}$ | $\log g$ | (M/H) | (C/M) | (N/M) | ( $\alpha$ /M) |
|-------|------------------|----------|-------|-------|-------|----------------|
| GK    | Min 3500         | 0        | -2.5  | -1    | -1    | -1             |
|       | Max 6000         | 5        | 0.5   | 1     | 1     | 1              |
|       | Step 250         | 0.5      | 0.5   | 0.25  | 0.5   | 0.25           |
| F     | Min 5500         | 1        | -2.5  | -1    | -1    | -1             |
|       | Max 8000         | 5        | 0.5   | 1     | 1     | 1              |
|       | Step 250         | 0.5      | 0.5   | 0.25  | 0.55  | 0.25           |

location (mean, median) and spread (standard deviation, mean absolute difference) summary statistics on the residuals with a test set of known parameters.

### 3 SYNTHETIC SPECTRA

As mentioned in Section 1, machine learning techniques have been applied to stellar spectra in the past using synthetic data for training and testing (Bailer-Jones 2000). Synthetic spectra provide a unique application to evaluate the StarNet model, since all details of the spectra are known a priori, and the S/Ns and wavelength regions can be varied. Since we will evaluate StarNet on the APOGEE observed data sets, we have applied StarNet to synthetic spectra generated by the APOGEE collaboration.

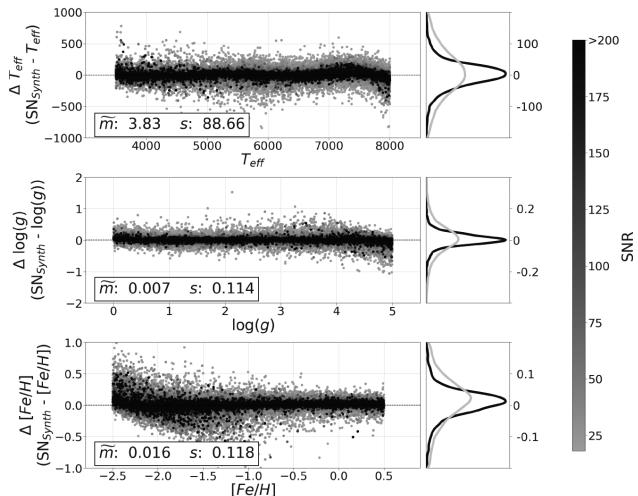
#### 3.1 Training and testing StarNet with the APOGEE ASSET spectra

The synthetic spectra used by the APOGEE consortium have been generated using MARCS and ATLAS9 model atmospheres, as described in Mészáros et al. (2012), and the radiative transfer code ‘ASSET’ (Koesterke et al. 2008). These synthetic spectra were continuum normalized in the same way as the observed APOGEE spectra to facilitate a closer match between the synthetic and observed data sets. Provided in a publicly available 6D spectral grid compressed with Principal Components Analysis (see Table 1 for parameter distribution), we have used these synthetic spectra as a first test of StarNet. To sample a spectrum at any desired location within the synthetic grid, we used a third-order interpolation routine between spectra at the existing grid points.

In all instances of training StarNet on synthetic data, we added Gaussian noise to match  $S/N \approx 20$  up to noiseless spectra. Each spectrum was used several times in a particular training process (though only once per training iteration) and therefore different amounts of noise were added with each iteration to mimic the behaviour of multiple visits to a single object in the APOGEE survey. This method also contributes to generalizing the predictive capabilities of StarNet on lower S/N spectra.

As a first test, we generated a data set of 300 000 synthetic spectra via random sampling of the stellar parameters within the limits of the ASSET grid. Of these spectra, 260 000 were randomly selected as the reference set. A subset of these (224 000 or 86 per cent of the reference set) were used to train StarNet. The remaining 36 000 spectra from the reference set were used to cross-validate the model following each training iteration.

The 40 000 spectra left out of the reference set were used as a test set, and the residuals between the StarNet predictions and the generated parameters are shown in Fig. 2. For all three predicted stellar parameters, the variances of the residual distributions are inversely proportional to the S/N. At  $[Fe/H] < -1.9$  and lower S/N, there is a distinct trend where StarNet appears to overpredict the



**Figure 2.** StarNet prediction residuals with the generated stellar parameters for a test set of 40 000 ASSET synthetic spectra (Koesterke et al. 2008). StarNet was trained with 224 000 synthetic spectra randomly sampled from the ASSET synthetic grid. Projected residual distributions are shown on the right (black for synthetic spectra with  $S/N > 80$ , grey for  $S/N < 60$ ). The median value ( $\tilde{m}$ ) and standard deviation ( $s$ ) are calculated in each panel.

metallicity. In this region, absorption features are not as prominent; therefore, the noise effectively causes the StarNet  $[Fe/H]$  predictions to be similar for all of these stars (i.e.  $[Fe/H] \approx -2.2$ ).

#### 3.2 Stellar parameter predictions and precision

To evaluate the errors in the StarNet predictions, it is necessary to understand how the output of the model is affected by the uncertainties in the input spectra ( $\mathbf{x}$ ). Treating the entire network as a vector valued function  $\mathbf{f}(\mathbf{x}, \mathbf{w}, \mathbf{b}) = (f_{T_{\text{eff}}}, f_{\log g}, f_{[Fe/H]})$ , the Jacobian of the StarNet model for a given star and stellar label,  $j$ , is given by

$$\mathbf{J}(\mathbf{x}, \mathbf{w}, \mathbf{b}) = \left( \frac{\partial f_j}{\partial \mathbf{x}_1}, \dots, \frac{\partial f_j}{\partial \mathbf{x}_n} \right) (\mathbf{x}, \mathbf{w}, \mathbf{b}).$$

Here, we assume that each spectrum is accompanied by a corresponding error spectrum,  $\mathbf{e}_x$ . Propagating the error spectrum with the Jacobian of that spectrum results in an approximation of the statistical uncertainty,  $\sigma_{\text{prop}}$ , of the output prediction for each stellar parameter computed as the product:

$$\sigma_{\text{prop}}^2 \approx \mathbf{J}^2 \mathbf{e}_x^2.$$

Besides restricting our approximation to the small error domain, we do not compute the error from the StarNet CNN model weights for practical reason. To compensate for this deficit, an empirical intrinsic scatter,  $\sigma_{\text{int}}$ , is derived from the synthetic test data, which is then added in quadrature – element wise – with the propagated uncertainty. This gives a total estimated uncertainty:

$$\sigma = \sqrt{\sigma_{\text{prop}}^2 + \sigma_{\text{int}}^2}.$$

The intrinsic scatter term is slightly dependent on the region of the parameter space in which the predicted parameter lies, reflecting the capacity of the physical model to match observed spectral features. For instance, this intrinsic scatter term will be larger for stars that are predicted to have low metallicities compared to those that are found to be metal-rich (see Fig. 2). Therefore, we derive this scatter term in bins, within the same range of the ASSET synthetic grid

**Table 2.** Cuts applied to APOGEE DR13 for the training and test set.

| Cut  | Visit spectra          | Combined spectra             |
|--|------------------------|------------------------------|
| No cuts  | 559 484                | 143 482                      |
| ASPCAPFLAG   | 327 594                | 92 983                       |
| STARFLAG   | 142 875                | 45 474                       |
| $4000 \text{ K} < T_{\text{eff}} < 5500 \text{ K}$ | 127 357                | 39 017                       |
| $v_{\text{scatter}} < 1 \text{ km s}^{-1}$         | 120 879                | 37 311                       |
| $[\text{Fe}/\text{H}] > -3$                        | 120 879                | 37 311                       |
| $\log(g) = -9999$                                  | 113 956                | 35 591                       |
| Combined SNR > 200                                 | 53 135                 | Not applicable               |
| Partition into reference and test sets             | Training set<br>41 000 | Cross-validation set<br>3784 |
|  |                        | Test set<br>21 037           |

found in Table 1: for  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  the bin sizes are 450 K, 0.5 dex, and 0.3 dex, respectively.

#### 4 STARNET APPLICATIONS TO APOGEE SPECTRA

The APOGEE survey has been carried out at the 2.5-m Sloan Foundation Telescope in New Mexico, and covers the wavelength range from 1.5 to 1.7  $\mu\text{m}$  in the H band, with spectral resolution  $R = 22\,500$ . Targets are revisited several times until  $S/N \geq 100$  is reached (Majewski et al. 2017). All visits are processed with a data reduction pipeline to 1D and are wavelength calibrated before being combined. The individual visits, unnormalized combined spectra, and normalized combined spectra are all recorded into the publicly-available APOGEE data base (Nidever et al. 2015).

The APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP) is a post-data reduction pipeline tool for advanced science products. This pipeline relies on the non-linear least squares fitting algorithm FERRE (Garcia-Pérez et al. 2016), which compares an observed spectrum to a grid of synthetic spectra generated from detailed stellar model atmospheres and radiative line transfer calculations. For DR12, this synthetic grid was the ASSET grid used for training in the previous section. The fitting process estimates stellar parameters ( $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ) and abundances for 15 different elements (C, N, O, Na, Mg, Al, Si, S, K, Ca, Ti, V, Mn, Fe, and Ni). These results are further calibrated with comparisons to stellar parameters from optical spectral analyses of stars from a variety of star clusters (Cunha et al. 2015; Mészáros et al. 2015); see the Appendix for more details.

##### 4.1 Pre-processing of the input data

As a test, StarNet was also trained using the APOGEE observed spectra and ASPCAP DR13 stellar parameters (see Albareti et al. 2017). The APOGEE error spectra and the estimated errors from ASPCAP were not included as part of the reference set. This effectively limits the accuracy of StarNet to the reference set parameters. To minimize the propagation of the ASPCAP errors in the StarNet results, we limited the range of our data set. For example, Holtzman et al. (2015) caution against using stars cooler than  $T_{\text{eff}} = 4000 \text{ K}$ , where models are less certain, as well as hotter stars where the spectral features are less defined. For these reasons, our data set was reduced to stars between  $4000 \text{ K} \leq T_{\text{eff}} \leq 5500 \text{ K}$ .

The APOGEE team have also flagged specific stars and visits for a variety of other reasons. Stars from the DR13 data set with either a STARFLAG or ASPCAPFLAG were removed from the data set, which includes stars contaminated with persistence (see Nidever

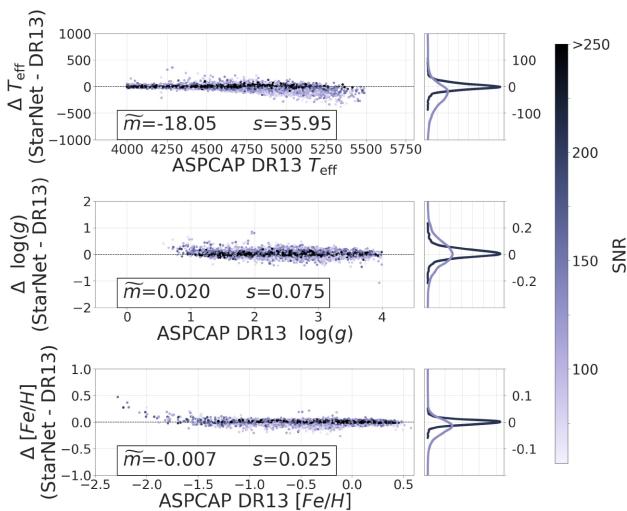
et al. 2015; Jahandar et al. 2017), stars that were flagged for having a bright neighbour, or having estimated parameters near the parameter grid edge of the synthetic models. Furthermore, stars with  $[\text{Fe}/\text{H}] < -3$  were removed, or those with a radial velocity scatter,  $v_{\text{scatter}}$ , greater than  $1 \text{ km s}^{-1}$ . Spectra with a high  $v_{\text{scatter}}$  are possibly binary stars (Nidever et al. 2015), with potentially duplicate spectral lines.

Implementing these restrictions to the *reference set* allowed the NN to be trained on a cleaner data set and learn from more accurate stellar parameters, while placing these restrictions on the *test set* was an attempt to maximize the validity of ASPCAP stellar parameters to be used when comparing with StarNet predictions. All of these cuts to the APOGEE data set are summarized in Table 2.

In addition to the restrictions mentioned above, the reference set only included individual visit spectra from stars where the combined spectrum had  $S/N > 200$ . ASPCAP stellar parameter precision is known to degrade at lower  $S/N$  (Ness et al. 2015; Casey et al. 2016). However, using high  $S/N$  combined spectra as the training input resulted in overfitting the model (see Appendix A) to spectra with high  $S/N$ . To compensate for this issue, StarNet was trained on individual visit spectra, while using the stellar parameters obtained from the combined spectra. This step allowed the model to be trained on lower  $S/N$  spectra with high precision parameters. Limiting our training sample to contain stellar parameters from high  $S/N$  spectra was also practical: currently StarNet is not able to weight the input spectra according to the noise properties of the spectra during training. When combining various data sets, which we will address in future work, it might become more critical, but is less relevant for the current analysis.

The last phase of the data pre-processing was normalizing the spectra. Both the Cannon 2 (Casey et al. 2016) and ASPCAP have implemented independent continuum normalization techniques; in addition, the APOGEE DR13 spectra were further normalized in certain spectral regions. The ASSET synthetic spectra were similarly normalized to facilitate proper matching with the data. It has been suggested (e.g. Casey et al. 2016) that the ASPCAP continuum normalization techniques result in an  $S/N$  dependency. This is a potential limitation of StarNet when it has been trained on synthetic spectra, since these spectra were ASPCAP normalized without the addition of noise beforehand. Any non-linearities in continuum normalization would be excluded from the synthetic spectra data set. If this inconsistent normalization is indeed a problem, it could create an inherent mismatch between a synthetic spectrum and a low  $S/N$  APOGEE spectrum of identical stellar parameters, leading to erroneous stellar parameter estimates.

This potential for a continuum normalization bias motivated us to adopt a simple and linear normalization method in StarNet; for



**Figure 3.** Residuals of StarNet predictions and ASPCAP parameters for a test set of 21,037 combined spectra over a large range of S/N. StarNet was trained on 41 000 individual visit spectra from the APOGEE DR13 data set. As the S/N decreases, small deviations are seen for the hotter stars, stars at the lower end of the surface gravity range, and the most metal-poor stars, which are likely due to a sample size bias in the training set. Projected residual distributions are shown on the right (dark purple for spectra with S/N > 200, light purple for S/N < 100).

*Note.* The median value ( $\tilde{m}$ ) and standard deviation ( $s$ ) are calculated in each panel, as in Fig. 2.

training and testing on APOGEE spectra, the spectra were split into blue, green, and red chips, and each chip was divided by its median flux value. The three chips were then combined to create a single spectrum vector. Given that normalization procedures usually do not need external information, our simple approach is also testing the network capabilities compared to a more physically driven continuum removal. Further analysis is, however, required to determine the full impact of continuum normalization in this NN approach.

## 4.2 Training and testing StarNet with APOGEE spectra

StarNet has been trained and tested on the ASPCAP stellar parameters ( $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ) corresponding to individual visit spectra and combined spectra, respectively. As discussed previously, this included stars from the APOGEE DR13 data set with the cuts outlined in Table 2. There were 17 149 stars with  $S/N > 200$  that met these requirements, of which 2651 were used as part of the test set, and 14 498 stars (containing 44 784 individual visits) were used as the StarNet reference set. This is slightly less than 10 per cent of the total APOGEE DR13 data set. A subset of these individual visits (41 000 or 92 per cent of the reference set) was randomly selected for the training set. The remaining 3784 visit spectra were used to cross-validate the model StarNet following each forward propagation during training. No significant deviations were found between random selections for training and cross-validation samples.

Once trained, StarNet was applied to the test set containing both high and low S/N combined spectra. The StarNet predictions are compared to ASPCAP parameters in Fig. 3. For the high S/N spectra, StarNet predictions show excellent agreement with the ASPCAP DR13 results. For the lower S/N spectra, there are more and larger deviations between StarNet and DR13. For example, at high temperatures ( $T_{\text{eff}} > 5000$  K), StarNet predicts lower effective temperatures than DR13. The quality of the model predictions depends on the

number of stars in the training set that span the parameter space of the test set; therefore, we suggest that these deviations are due to an insufficient number of stars with  $T_{\text{eff}} > 5000$  K (~4 per cent) in our reference set. Similarly, there are few stars at low metallicities,  $[\text{Fe}/\text{H}] < -1.5$  (~0.2 per cent of the reference set), and therefore applying StarNet to the most metal-poor stars also results in larger deviations from the DR13 results. Sample size bias, as seen in these two regions, reinforces the need for large training sets for deep NNs such as StarNet.

Another potential source of error in the predictions at high temperatures and low metallicities lies in the spectra themselves. In these regions, the spectral features are weaker, therefore the NN will struggle to locate the most important features during training. Similarly, the parameters determined by ASPCAP will have larger intrinsic uncertainties. Furthermore, these effects can be amplified when testing on lower S/N spectra.

After training, StarNet was then applied to 148 724 stars in the APOGEE DR13 data base. This final inference step is very fast: taking about 1 min of CPU time for the whole data set. The short amount of time required to make these predictions is a huge advantage for the rapid data reduction of large spectral surveys.

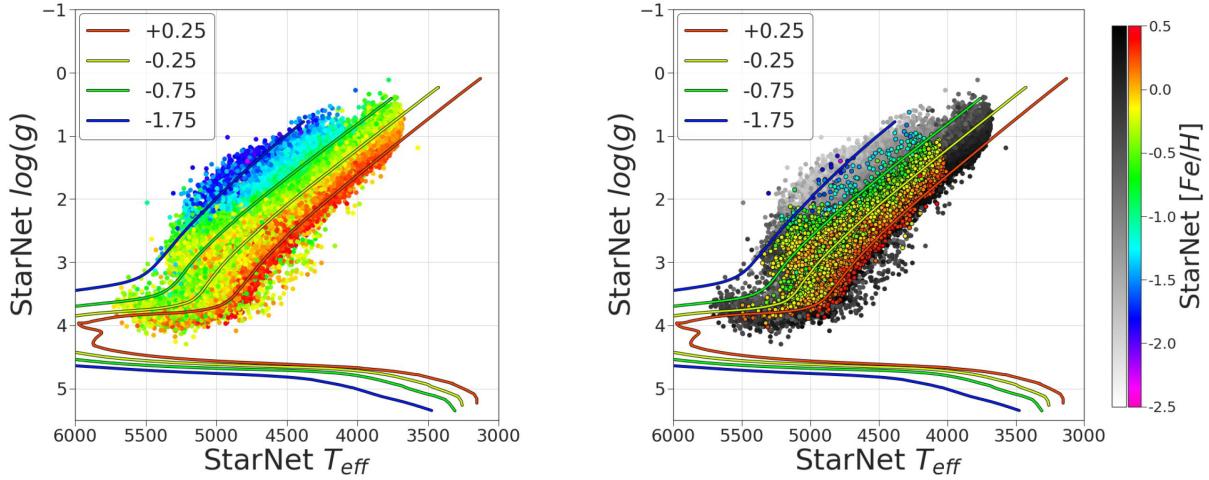
StarNet predictions for 99 211 spectra are shown in Fig. 4, where stars with ASPCAP  $T_{\text{eff}}$ ,  $\log g$ , or  $[\text{Fe}/\text{H}]$  values of -9999 were removed along with those that are presumed to be  $M$  dwarf stars (see Section 5.2). StarNet predictions are compared to stellar isochrones to show the expected trend in the parameter relationships; four isochrones with different metallicities ( $[\text{Fe}/\text{H}] = 0.25, -0.25, -0.75$ , and  $-1.75$ ) were generated from the Dartmouth Stellar Evolution Database (Dotter et al. 2008), with an age of 5 Gyr and  $[\alpha/\text{Fe}] = 0$ . Fig. 4 also shows the StarNet reference set of 14 498 stars, highlighted to show the spread in their abundances across the parameter space.

To analyse the StarNet model self-consistency, predictions were made for both the individual visit spectra and combined spectra for the same objects. Targets with four or more individual visits were selected for this test, and the differences in the residuals were calculated. For each stellar parameter bin, the mean value for 100 objects is shown in Fig. 5 (left-hand panel), whereas there were 230 objects in each S/N bin (right-hand panel). The propagated and total errors shown are discussed in Section 3.2. When comparing StarNet predictions for the individual visits to those for the combined spectra, the results are quite consistent across the majority of the parameter space, though as discussed previously, there is a noticeable increase in the deviations at lower metallicities ( $[\text{Fe}/\text{H}] < -1.2$  dex) and higher temperatures ( $T_{\text{eff}} > 5100$  K). Also as expected, the scatter increases at lower S/N, although only marginally until  $S/N \lesssim 60$ . Even at the lowest values ( $S/N \sim 15$ ) the results are quite consistent. The ability to predict well on lower S/N spectra is largely due to the fact that StarNet was trained on individual visits – rather than combined spectra only – with high-validity stellar parameters.

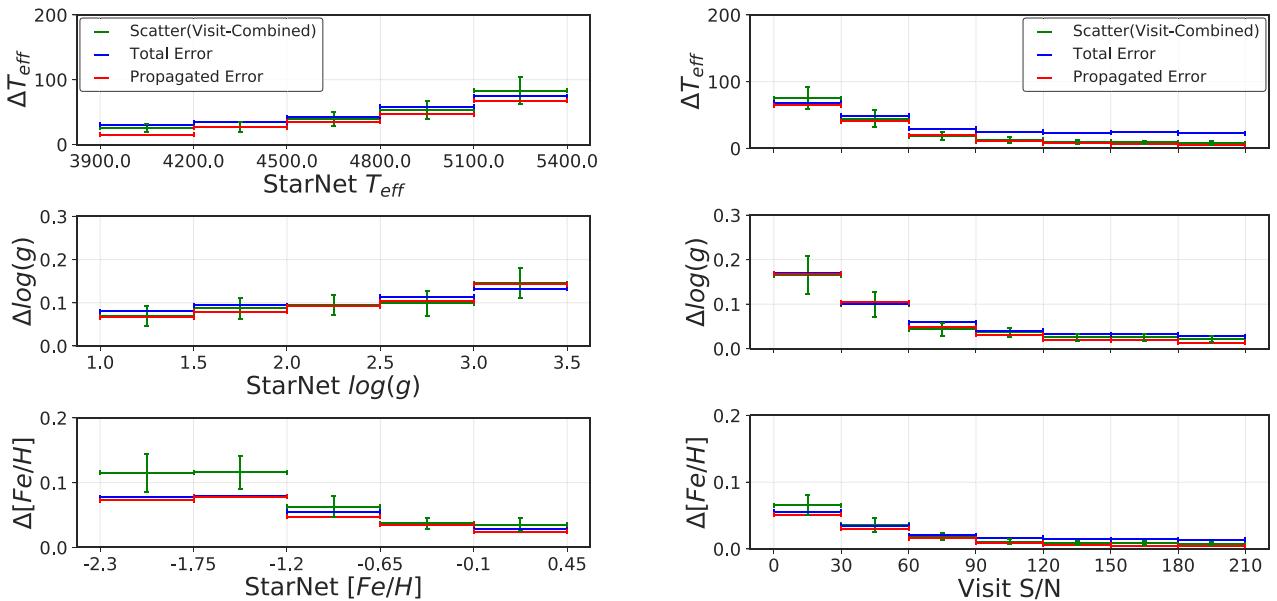
The trends seen in the residuals between predictions for individual visits and combined spectra (i.e. an increase in scatter at lower metallicities and lower S/N) are also reflected in the StarNet propagated errors (see Section 3.2), also shown in Fig. 5. This provides confidence that the error propagation methods used give an adequate estimate of the uncertainties in the StarNet predictions.

## 4.3 Model selection

Deep learning architectures, such as StarNet, typically involve experimentation and tuning of hyper-parameters in order to converge to an adequate model. Some of these hyper-parameters include the



**Figure 4.** Left-hand panel: stellar parameters from StarNet for 99 211 stars, showing  $\log g$  parameters against  $T_{\text{eff}}$  across a wide range of  $[\text{Fe}/\text{H}]$  (see colour bar). Right-hand panel: the StarNet reference set of 14 498 stars highlighted over top of the full sample of 99 211 stars (in grey). Stellar isochrones are shown in both panels (from Dotter et al. 2008), with ages of 5 Gyr and  $[\text{Fe}/\text{H}]$  values found in the upper left corner (corresponding to the colour bar).

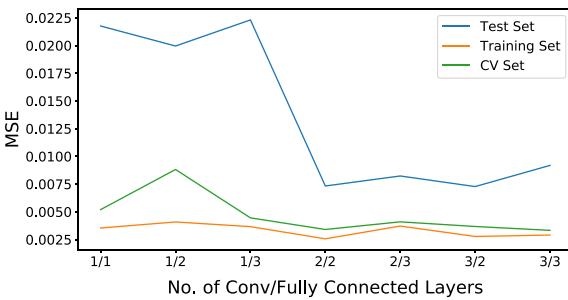


**Figure 5.** Comparison of StarNet predictions on individual visits to predictions for the corresponding combined spectra. Left-hand panel: the comparisons are made to the parameter space, where each parameter bin includes 100 randomly selected spectra. The largest deviations are found at lower metallicities and higher temperatures. Right-hand panel: comparisons are made based on the S/N of the individual visits, where each S/N bin contained 230 spectra. In both plots, the ‘Scatter (Visit-Combined)’ is the standard deviation in the residuals between the predictions for the individual visits and combined spectra for objects with more than four individual visits. The ‘Propagated Errors’ are the error terms due to the error spectrum, whereas the ‘Total Errors’ are the propagated error and ‘Intrinsic Scatter’ terms (see Section 3.2) added in quadrature. StarNet was trained on 41 000 individual visit spectra from the ASPCAP DR13 data set.

number of filters in each convolutional layer; the length of the filters that are convolved across the inputs in each of these convolutional layers; the number of connection nodes in each fully connected layer; the pooling window size in the maxpooling layer (the number of inputs which are compared against each other to find the maximum value); and the learning rate of the gradient descent optimizer. The final selection of hyper-parameters was automated as follows: to select the optimal model architecture, a hyper-parameter optimization was run in two stages. During the first stage, the number of convolutional layers, fully connected layers, filters, and nodes were varied randomly, along with the length of the filters, the pooling window size, and the learning rate. This allowed the optimizer to test different combinations of the number of convolutional and fully

connected layers with a variety of other hyper-parameters to ensure that each combination could reach its maximum predictive potential. The same training and cross-validation sets (as described in Section 4.1) were used during the training of these models, and the same test set (discussed in Section 4.2) was used to test each model. The metric used for model comparisons was the MSE between the target and predicted parameters. To have each parameter weighted equally, the parameters were normalized to have approximately zero mean and unit variance. An example of the best-performing models, evaluated for each combination of layers, is shown in Fig. 6.

It was found that increasing the number of layers improved the model predictions, but that there was a plateau in performance when combining two convolutional layers with two fully connected



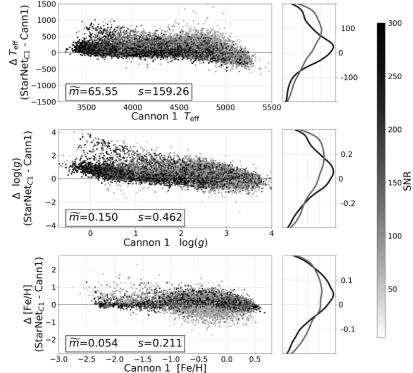
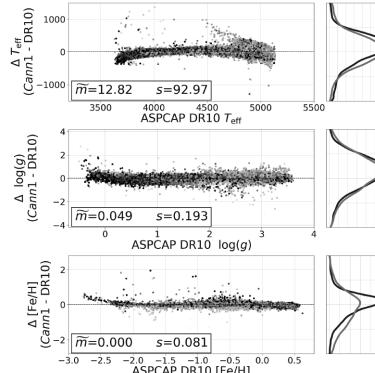
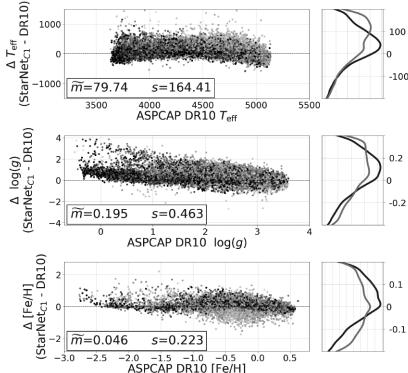
**Figure 6.** The MSE between the normalized target parameters and StarNet predictions are plotted against the different combinations of convolutional and fully connected layers. These results were used to help select the model architecture of StarNet.

layers. The addition of more layers on to this architecture does not clearly improve the prediction results, while fewer layers led to significantly worse MSE.

A second hyper-parameter optimization was then started, where the number of convolutional and fully connected layers were fixed at two each, while the remaining hyper-parameters were varied and optimized using a Tree-structured Parzen Estimator (TPE, Bergstra et al. 2011). The model selected by this second stage of hyper-parameter optimization was similar to that shown in Fig. 1, and this model was used as the starting point for our model architecture selection. Small changes were made to discover possible improvements until the StarNet model architecture performed consistently with adequate results.

While simpler models could perform comparably well to the StarNet architecture when trained on synthetic data, it was decided that a single model architecture should be used to train on both synthetic data and APOGEE spectra.

Earlier studies found that applying NNs to stellar spectra only necessitates at most two hidden layers with fewer nodes than the amount used in StarNet (Bailer-Jones 2000), but these were applied to much smaller data sets. Current machine learning methods – along with large data sets – allow for more complex models to improve performance while still avoiding overfitting. The max pooling layer reduces the degrees of freedom of the model by minimizing the impact of unnecessary weights and therefore simplifying the model function. To speed up convergence for the StarNet model fit, we use the *He Normal* (He et al. 2015) weight initialization,



**Figure 7.** Comparison of StarNet<sub>C1</sub> results with ASPCAP (left-hand panel) and StarNet<sub>C1</sub> results with the Cannon 1 (right-hand panel), as well as comparisons between the Cannon 1 and ASPCAP (centre panel). StarNet<sub>C1</sub> was trained on APOGEE DR10 combined spectra from the same 542 stars that the Cannon 1 used for training. The test set includes combined DR10 spectra that had both ASPCAP and the Cannon 1 predictions. The median value ( $\bar{m}$ ) and standard deviation ( $s$ ) are calculated in each panel, as in Fig. 2. (Note that the range in the plot axes differ from the other plots in this paper).

also permitting for convergence of deeper models and eventually enabling the NN to find more complex features. In short, the *He Normal* initializer, Rectified Linear Unit (ReLU) activation, and ADAM optimizer (see Appendix A) allow for a deeper model to reach convergence more easily and at an accelerated rate, while the max pooling layer simplifies the model rank complexity. The use of a cross-validation set is another technique implemented in StarNet to mitigate overfitting. These techniques and methods are described further in Appendix A.

#### 4.4 Comparisons with the Cannon

The Cannon (Ness et al. 2015) is a data-driven, generative model that shares the same limitations as a supervised learning approach, i.e. both rely on a reference data set. Unlike StarNet, the generative approach uses stellar parameters as the inputs and spectra as the outputs during training. The Cannon uses a quadratic polynomial function to translate the stellar parameters into spectra, and the best-fitting coefficients of the function are found through least-squares fitting. During the test phase, the stellar parameters are determined with another regression step, where the stellar parameters are varied until they produce a spectrum that best matches the observed spectrum.

The Cannon was originally trained on APOGEE DR10 (Mészáros et al. 2013) combined spectra from 542 stars in 19 clusters. These are far fewer training examples than StarNet uses from DR13, and yet the Cannon was able to predict stellar parameters for the DR10 spectra extremely well. In an effort to compare the two techniques, we trained the same StarNet architecture, labelled StarNet<sub>C1</sub>, on the same data set of 542 stars from DR10. StarNet<sub>C1</sub> was then applied to 29 891 combined spectra that had both ASPCAP and the Cannon predictions for stellar parameters. These are compared in Fig. 7, where the limitations of the StarNet<sub>C1</sub> model are clearly seen. The machine learning method used for StarNet requires a large training sample that spans a wide parameter space. Training on 542 combined spectra does not span enough of the parameter range at an accumulated S/N to fit the complex StarNet CNN. This can also be seen in the large residuals when comparing StarNet<sub>C1</sub> predictions and ASPCAP DR10 parameters. Predictions for low S/N spectra show the most significant scatter.

The requirement of large training sets that adequately represent the test set is a limitation of many deep learning methods, such as

**Table 3.** Comparison of the StarNet<sub>C2</sub> and the Cannon 2 results for a test set of 85 341 combined spectra from APOGEE DR12. Metrics used are the MAE and RMSE, with respect to the same stars.

| Metric | Parameter        | StarNet <sub>C2</sub> | Cannon 2 |
|--------|------------------|-----------------------|----------|
| MAE    | $T_{\text{eff}}$ | 31.2                  | 46.8     |
|        | $\log g$         | 0.053                 | 0.066    |
|        | [Fe/H]           | 0.025                 | 0.036    |
| RMSE   | $T_{\text{eff}}$ | 51.2                  | 71.6     |
|        | $\log g$         | 0.081                 | 0.102    |
|        | [Fe/H]           | 0.040                 | 0.053    |

StarNet. However, as larger data sets become available, then these methods can be extremely effective, as discussed in the next section.

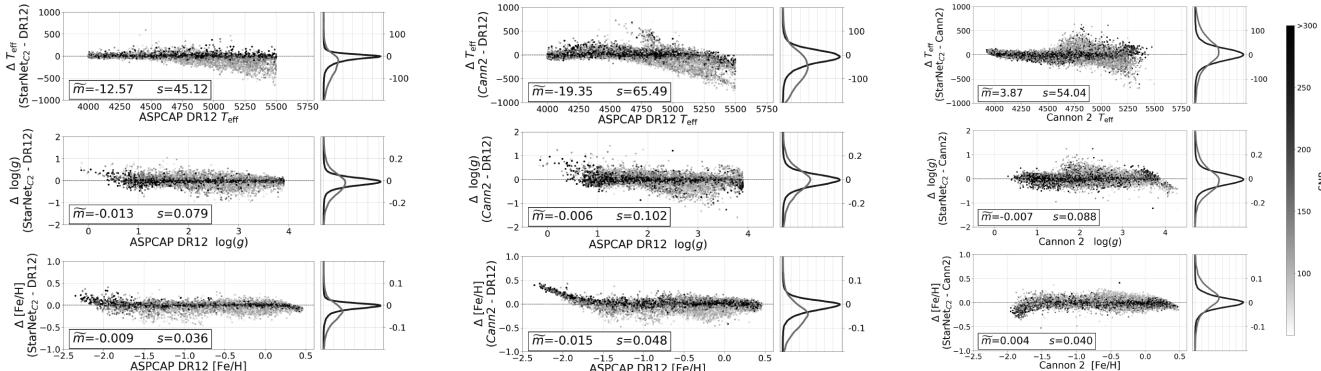
#### 4.5 Comparisons with the Cannon 2

The Cannon 2 (Casey et al. 2016) is a continuation of the Cannon project. In addition to the three stellar parameters predicted by the Cannon, The Cannon 2 predicts 14 additional chemical abundances. The Cannon 2 is also trained on individual visit spectra, but it trains on a larger number of stars than the Cannon. Also, the Cannon 2 adds a regularization term in the function at training, to reduce the likelihood of the model overfitting the training data.

The training set for the Cannon 2 consisted of stars with  $200 < \text{S/N} < 300$ ,  $[\text{Fe}/\text{H}] > -3$ ,  $[\text{X}/\text{Fe}] < 2$ ,  $[\alpha/\text{Fe}] > -0.1$ ,  $[\text{v}/\text{Fe}] > -0.6$ ,  $\text{v}_{\text{scatter}} < 1 \text{ km s}^{-1}$ , and excluded targets flagged with the ASPCAPFLAG. These cuts removed stars near the edge of the grid of synthetic models, along with other sources of error resulting in poor ASPCAP parameter determinations. However, it did not take into account spectra flagged with persistence. The Cannon 2 training set consisted of visits from 12 681 stars.

To make an adequate comparison between the two models, StarNet<sub>C2</sub> was trained on 39 098 visits from 12 681 stars that met the same restrictions set by the Cannon 2 training set, from APOGEE DR12 (Holtzman et al. 2015). A test set of 85 341 combined spectra with  $4000 \text{ K} < T_{\text{eff}} < 5500 \text{ K}$  and  $\log g < 3.9$  was adopted, with similar restrictions as the Cannon 2 test set.

We compile the root-mean-square error (RMSE) and the mean absolute error (MAE) for an identical test set for both the Cannon 2 and StarNet<sub>C2</sub> in Table 3. This confirms the visual inspection of the residuals in Fig. 8, that StarNet<sub>C2</sub> is capable of predicting



**Figure 8.** Comparison of StarNet<sub>C2</sub> results with ASPCAP (left-hand panel) and StarNet<sub>C2</sub> results with the Cannon 2 predictions (right-hand panel), as well as comparisons between the Cannon 2 and ASPCAP (centre panel). All parameters are based on the DR12 data set. StarNet was trained on individual visits from similar stars as those used for training the Cannon 2. The test set used to compare the two methods was also similar to that used by the Cannon 2. The median value ( $\bar{m}$ ) and standard deviation ( $s$ ) are calculated in each panel, as in Fig. 2.

values closer to the ASPCAP stellar parameters. Also seen in Fig. 8, StarNet<sub>C2</sub> differs from ASPCAP at lower metallicities and hotter temperatures, when the S/N of the spectra is lower. These effects are also seen in the Cannon 2 predictions. In fact, a comparison of StarNet<sub>C2</sub> to the Cannon 2 shows that the effect is larger in the Cannon 2 results.

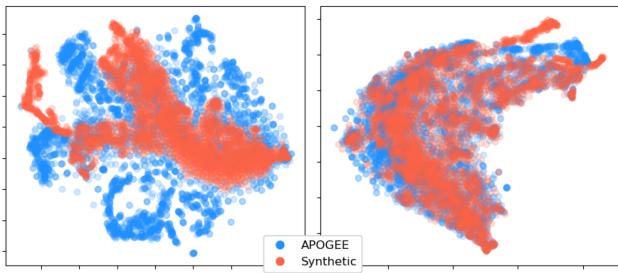
#### 4.6 Training StarNet with synthetic spectra

A crucial next step in the development of an NN approach to analysing stellar spectra is to show that StarNet can also predict stellar parameters without an external pipeline. In this section, we present stellar parameter results for APOGEE data after training StarNet on synthetic spectra only. Our goal is to show that StarNet can operate as a stand-alone data processing pipeline, producing an independent data base of stellar parameters that does not depend on previous ASPCAP pipeline results for training.

##### 4.6.1 Synthetic gap

Differences in feature distributions between synthetic and observed spectra is referred to as the *synthetic gap*. To probe the feasibility of training StarNet on synthetic spectra and accurately predict the parameters from observed spectra, it was necessary to ensure that the synthetic gap was relatively small. Since each spectrum consists of 7214 data points or represents a point in 7214-dimensional space, one would expect a synthetic and observed spectrum with the same parameters to occupy approximately the same region in this space if the gap was indeed small. Visualizing this space is possible only through dimensionality reduction. We employed the t-Distributed Stochastic Neighbour Embedding (t-SNE; Maaten & Hinton 2008), a technique often used in machine learning to find clusters of similar data in a two-dimensional (2D) space. This 2D space is comprised of arbitrary variables – not physical stellar parameters – meant to serve as a lower dimensional representation of the spectra to facilitate visualization.

A subset of 4000 APOGEE DR13 spectra with  $\text{S/N} > 200$  and known ASPCAP parameters were randomly selected, along with 4000 interpolated synthetic ASSET spectra with the same parameters. After applying t-SNE, a distinct separation could be seen between the synthetic and APOGEE spectra (the left image in Fig. 9). Examination of mismatched spectra showed that there were zero-flux values in wavelength bins along the spectra, a method used in



**Figure 9.** t-SNE visualization of the synthetic and APOGEE spectra before zero-flux substitution through nearest neighbour interpolation (left) and after (right).

the APOGEE pipeline to flag bad pixels. Unfortunately, it is not possible to mask these values at the testing stage with the current implementation of StarNet, thus a nearest-neighbour interpolation was carried out to smooth over the zeros. Another round of t-SNE revealed closer agreement between synthetic and APOGEE spectra (the right image in Fig. 9). These zero-flux values for all APOGEE spectra were fixed before predicting their stellar parameters in subsequent StarNet training from synthetic spectra.

While interpolating the zero-point values in the observed spectra results in more accurate predictions, we caution that this may not be an ideal solution. This is because we are modifying the data in a way that may not resemble a spectrum without the zero-point values. An alternative method would be to artificially inject zero-point values in the synthetic spectra during training. This would allow StarNet to distribute weights more evenly across the spectrum, mimicking the training process on observed spectra. We note that this is not an issue when StarNet is trained on real APOGEE spectra that inherently have the zero-flux values.

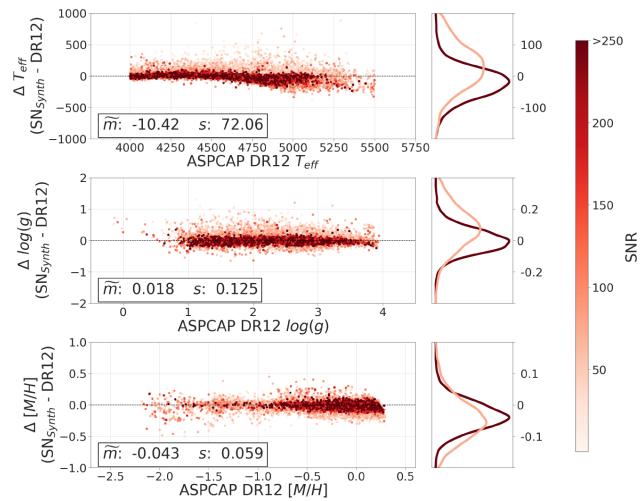
#### 4.6.2 Predictions for APOGEE DR12 spectra

After training StarNet with synthetic spectra from the ASSET code (as described in Section 3), we applied it to a data set of 21 787 combined spectra from APOGEE DR12<sup>1</sup> using the same restrictions found in Table 2. We compare our results to the ASPCAP DR12 parameters (see Fig. 10), since those were also determined using the ASSET spectral grid (unlike the APOGEE DR13 results).

The distribution of the residuals is similar to those seen in Fig. 3, where StarNet was trained on observed APOGEE spectra. Using the method as described in Section 3.2, the intrinsic scatter error terms for spectra with  $S/N > 150$  were calculated to be  $\Delta T_{\text{eff}} = 51 \text{ K}$ ,  $\Delta \log g = 0.06$ , and  $\Delta [\text{Fe}/\text{H}] = 0.08$ . For the entire spectral data set, including the lower S/N spectra, the intrinsic scatter errors are two times larger than when StarNet was trained on ASPCAP parameters. This is likely due to unaccounted systematic effects at training time, e.g. instrumental or extinction effects. An extension of StarNet to include unaccounted synthetic modelling effects is planned for future work.

Interstellar extinction, atmospheric extinction, and instrumental signatures are not simulated in the synthetic spectra. These effects can vary from star to star, generally affecting the bluer regions of a

<sup>1</sup> The ASSET spectra are only available with continuum normalization using the ASPCAP method, therefore we adopted the ASPCAP normalized DR12 spectra to test this implementation of StarNet. The non-ASPCAP normalization scheme described in Section 4.1 was only applied when StarNet was trained on APOGEE DR13 spectra



**Figure 10.** Residuals of StarNet predictions and ASPCAP parameters for a test set of 21 787 combined spectra from APOGEE. Comparisons were made to DR12 for consistency since our tests use the ASSET synthetic grid (not used for DR13). StarNet was trained on 224 000 synthetic spectra randomly sampled from the ASSET synthetic grid by interpolating between grid points. Projected distributions about the mean are shown on the right (dark red for observational spectra with  $S/N > 150$ , light red for  $S/N < 100$ ). The median value ( $\bar{m}$ ) and standard deviation ( $s$ ) are calculated in each panel, as in Fig. 2.

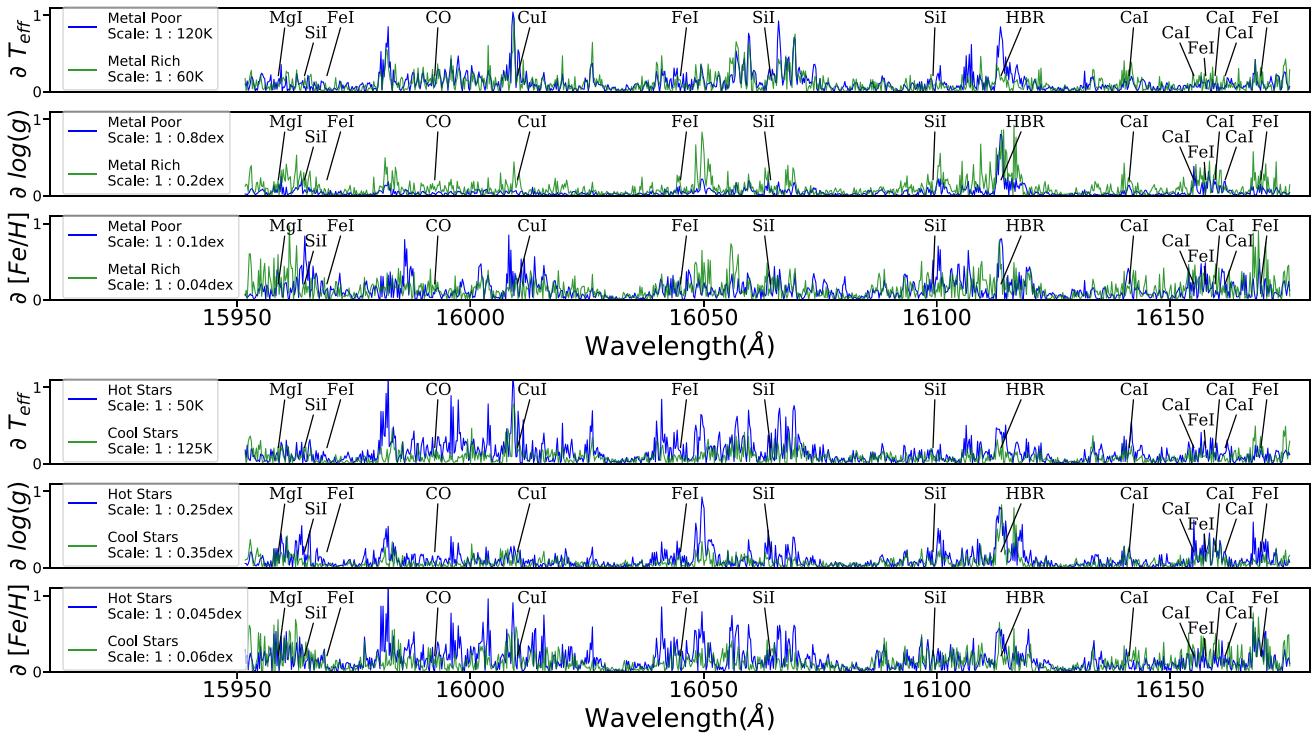
spectrum as a smoothly varying function. Given that the APOGEE spectra cover infrared wavelengths for mostly nearby bright objects ( $H > 12.5$ ), the effects of extinction are not expected to be significant. In Section 4.7, the sensitivity of the StarNet model to various features in the DR13 spectra is discussed.

Adding noise to the synthetic set also at the training stage allowed StarNet to learn that features impacted the stellar parameter estimates, while discounting weak features that would not be detectable in noisy APOGEE spectra. By varying the noise levels, it was found that an  $S/N > 20$  is necessary to decrease the residual intrinsic scatter before saturation. Adding noise to the synthetic training set not only helps to reproduce a more realistic setting but also accounts for some of the uncertainties in the physical and instrumental modelling and decreases overfitting. Since the APOGEE data set consists of mostly high S/N spectra, then adding a simple noise model appears to be sufficient for modelling the APOGEE spectra. However, potential applications on non-APOGEE data (or APOGEE data with more sensitive stellar parameters) may require a more thorough noise model.

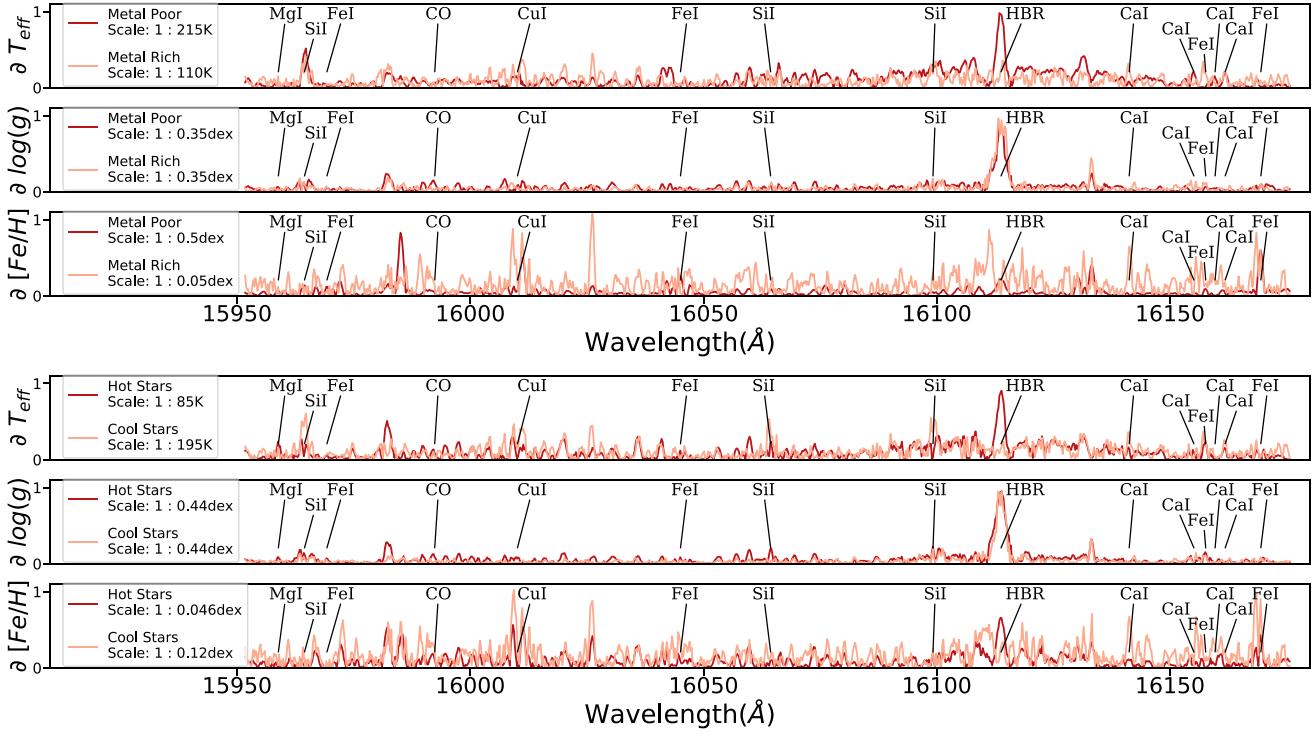
#### 4.7 Partial derivatives

It is possible to examine the learned model to determine which parts of the spectra the NN is weighting when predicting stellar parameters. We do this by calculating the partial derivatives of each of the outputs ( $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ ) with respect to every input flux value of a particular spectrum to obtain the Jacobian (as described in Section 3.2).

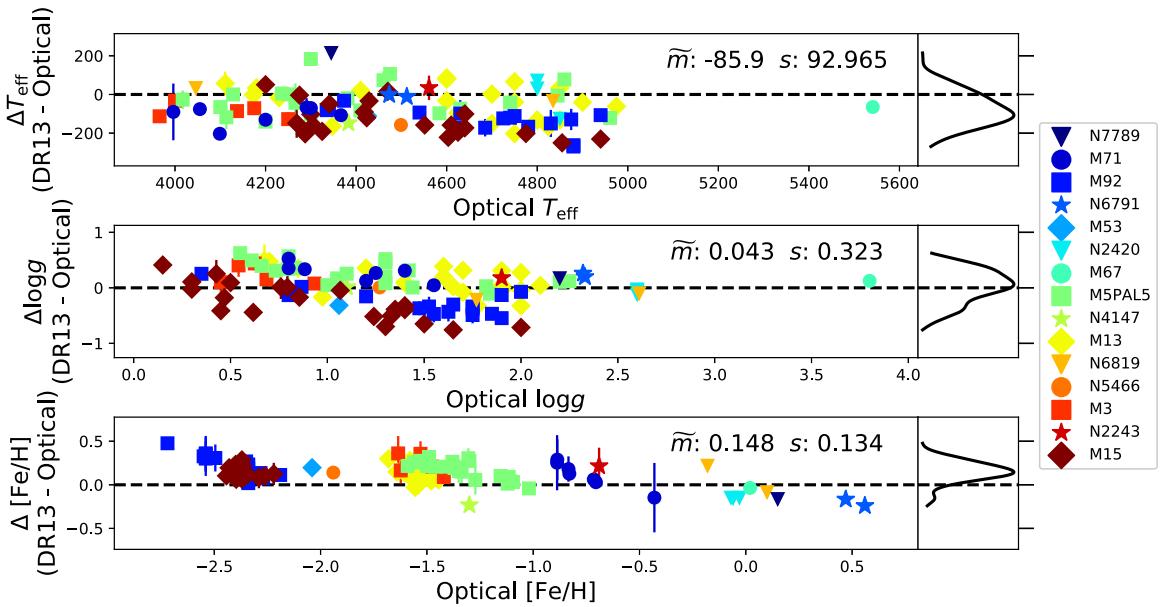
In Figs 11 and 12 – for both StarNet trained on APOGEE DR13 spectra and synthetic spectra, respectively – we show an average of the absolute-valued Jacobian from 2000 stars located in particular ranges of the parameter space. The Jacobian from metal-poor stars were compared against those from metal-rich stars. Comparisons were also made between hot and cool stars. We focus on a subsection of the green chip from 15 950 to 16 180 Å. Some features



**Figure 11.** Partial derivatives of the three stellar output parameters from the StarNet model – trained on APOGEE spectra – with respect to input wavelength bins for a section of the green chip. The partial derivatives of stars from different ranges of the parameter space were compared against each other. Stars with  $[Fe/H] > 0.0$  were compared to those with  $[Fe/H] < -1.2$  (top). Similarly, stars with  $T_{\text{eff}} > 5000$  K were compared to those with  $T_{\text{eff}} < 4300$  K (bottom). An average absolute-valued Jacobian was calculated from 2000 stars in each parameter range. Note the scale differences when comparing the partial derivatives.



**Figure 12.** Partial derivatives of the three stellar output parameters from the StarNet model – trained on synthetic data – with respect to input wavelength bins for a section of the green chip. The partial derivatives of stars from different ranges of the parameter space were compared against each other. Stars with  $[Fe/H] > 0.0$  were compared to those with  $[Fe/H] < -1.2$  (top). Similarly stars with  $T_{\text{eff}} > 5000$  K were compared to those with  $T_{\text{eff}} < 4300$  K (bottom). An average absolute-valued Jacobian was calculated from 2000 stars in each parameter range. Note the scale differences when comparing the partial derivatives.



**Figure 13.** Comparison of ASPCAP DR13 stellar parameters for 102 stars in clusters to their corresponding optical parameters sourced from the literature. When multiple literature sources are available for a particular star, the average of the reported parameters is shown with the error bars defined by the standard deviation of the optical data added in quadrature with the errors produced by StarNet. See Table B1 for a list of the literature references. The median value ( $\tilde{m}$ ) and standard deviation ( $s$ ) are calculated in each panel, as in Fig. 2.

are labelled from the APOGEE input line list. A few notable features include:

(i) Hydrogen bracket (HBR) lines: These lines play a significant role not only in the determination of gravity but also in the determination of temperature in metal-poor stars.

(ii) Atomic metal lines (e.g. FeI, CaI, and CuI): These lines play a significant role in the determination of temperature and metallicity throughout our stellar parameter range.

(iii) Certain atomic metal lines in the synthetic grid that play a role in the determination of temperature, e.g. SiII appears more and CaI appears less significant.

(iv) Unidentified or weaker lines that are not used in the APOGEE window functions and yet play an important role in the determination of all of the stellar parameters from the observed spectra.

(v) Unidentified or weaker lines that play a significant role only in the determination of metallicity for cool stars and metal-rich stars.

None of these features were pre-selected or externally weighted before the NN was trained. We also compared the partial derivatives for subsets of stars with  $S/N < 60$  versus  $S/N > 200$ , but found no notable differences in the predictive power. Similar results are seen in the other wavelength regions across the whole APOGEE spectral range.

## 5 DISCUSSION

### 5.1 Optical benchmark stars

The accuracy of the parameters returned by StarNet is limited by the quality of the training data. When StarNet is trained on the ASPCAP stellar parameters, uncertainties in the ASPCAP pipeline are implicitly propagated throughout the results. One option to examine the fidelity of the APOGEE StarNet results is to compare them to optical analyses. Stellar parameters and abundances for individual

stars derived through optical analysis can have higher precision due to the availability of higher resolution optical spectrographs, access to more spectral features, and higher confidence in the optical stellar atmosphere models (see Sneden et al. 2004; Gilmore et al. 2012; Venn 2012; Aoki et al. 2013; Yong et al. 2013; Howes et al. 2014; Roederer et al. 2014; Lamb et al. 2015).

To investigate the accuracy in the StarNet predictions from training on synthetic spectra, all stars from the APOGEE DR13 data base with the ‘Calibration-Cluster’ or ‘Standard Star’ flag were selected to create a group of benchmark stars with ASPCAP parameters. Parameters determined from optical spectral analyses were then obtained for these stars through direct matching of the APOGEE identifier/2MASS identifier (Skrutskie et al. 2006) to stars in the Pastel 2016 catalogue (Soubiran et al. 2016), and supplemented with additional data from references in Mészáros et al. (2013) and Holtzman et al. (2015). Fig. 13 compares the ASPCAP DR13 stellar parameters to the optically determined parameters for 102 benchmark stars (tabulated in Appendix B, with references).

Examining Fig. 13, there are small offsets between the ASPCAP DR13 and the optical values for all three stellar parameters. Comparing Fig. 13 to 3 – where StarNet is trained on observed spectra and compared to ASPCAP – a similar bias at low [Fe/H] is seen in the residuals. The bias seen in Fig. 3 is likely due to a sample size bias in the StarNet training set, though the bias in Fig. 13 could be due to an insufficient coverage in the synthetic grid. These systematic biases warrant the need for larger reference samples of the parameter space, as well as possible improvements in the coverage of theoretical spectra.

Comparing the residual scatter between the DR13 ASPCAP and the StarNet predictions (Fig. 3) to the scatter in the optical analyses predictions (Fig. 13), we see that the optical analyses have larger scatter by a factor of  $\gtrsim 2$  in  $T_{\text{eff}}$ ,  $\gtrsim 4$  in  $\log g$ , and  $\gtrsim 4$  in [Fe/H]. The larger scatter suggests that our simplified derivation of statistical uncertainties is sufficient until we can calibrate stellar parameters for a larger set of more homogeneous benchmark stars.

Well-calibrated stellar parameters would require a more thorough probabilistic model of both StarNet training and StarNet predictions. Therefore, we consider this very good agreement and do not investigate these intrinsic dispersions further at this time.

A comparison of Figs 10 and 13 reveals that, when StarNet is trained on synthetic spectra and compared to DR12, the scatter is a factor of  $\sim 2$  lower than the residuals between the DR13 and optical analyses. However, due to the different input physics, analysis techniques, spectral resolution, etc. between StarNet, APOGEE, and optical analyses, we cannot directly compare the precision of the stellar parameters returned by StarNet and an optical analysis at this time. Rather, this similar degree of scatter in the residuals indicates that the systematic errors intrinsic to StarNet are comparable to the systematic errors expected from an optical analysis. This indicates that the StarNet model is sufficient for a single survey analysis and can provide a new and exciting tool for the efficient analyses of spectroscopic surveys.

## 5.2 M dwarfs in DR13

While APOGEE has primarily observed red giant and sub-giant stars, DR13 includes spectra for known M dwarfs. These stars are accompanied with a specific Target Flag: ‘APOGEE MDWARP’. Initially, we had not removed these stars from the data set and found obvious discrepancies between the StarNet results and those provided by ASPCAP DR13. On closer inspection, we found 5900 stars that resembled the known M dwarfs, but that had not been flagged as M dwarfs. Those stars were removed from our data sets, using one of the following cuts:

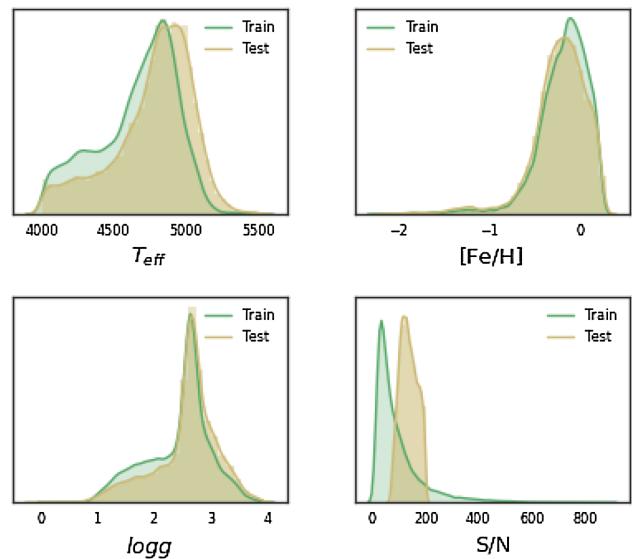
- (i) Metallicity parameters of  $-0.9 < (\text{ASPCAP } [\text{Fe}/\text{H}]) < 1.0$ , and where  $(\text{StarNet } [\text{Fe}/\text{H}] - \text{ASPCAP } [\text{Fe}/\text{H}]) < -0.4$ , or
- (ii) Gravity parameters of  $(\text{ASPCAP } \log g) > 3.6$ , and where  $(\text{StarNet } \log g - \text{ASPCAP } \log g) < -0.3$

The ASSET grid provided with the APOGEE DR12 data release does not include synthetic spectra for M dwarfs, and therefore, StarNet cannot be trained on those stars. Furthermore, StarNet cannot be trained on the observed spectra at this time since there are too few M dwarfs in the APOGEE data set ( $< 1000$  flagged). Our method for flagging M dwarfs is not robust, e.g. some RGB stars may have been removed, while other M dwarfs were not.

## 5.3 NN considerations

One limitation of a deep NN is the necessity of large training sets spanning a wide range of the parameter space. The distribution of the stellar parameters in the training set for StarNet (when trained on observed spectra) are shown in Fig. 14. From these plots, it is evident that there are fewer spectra for stars at high temperatures, low metallicities, and low gravities. As seen in Figs 3 and 7, having fewer training spectra in these regions of the parameter space results in less accurate predictions when testing on stars in those regions. StarNet is also challenged when combinations of limited parameters are scarce in the training set. For example, when StarNet was trained on APOGEE DR13 spectra, there were very few stars with  $4500 \text{ K} \leq T_{\text{eff}} \leq 5200 \text{ K}$ ,  $1 \leq \log g \leq 2$ , and  $[\text{Fe}/\text{H}] \leq -1.6$  in the reference set (see Fig. 4). This issue could be mitigated by adjusting the restrictions set on the StarNet reference set, but may cost in accuracy.

While stellar observations are limited by time and resources, generating synthetic spectra are limited only by computing resources. Being able to produce synthetic spectra that evenly span a parameter space can improve the ability of an NN to predict in regions



**Figure 14.** Distribution of parameters in the StarNet training and test sets when StarNet is trained and tested on APOGEE DR13 spectra. StarNet might not test well on stars occupying the parameter space less populated by the training set.

that may be affected by a sample bias. We see this as one of the strengths of StarNet.

In the future, we envision combining observed spectra with synthetically generated spectra to improve the precision in the StarNet predictions. This would increase the number of training spectra over the entire parameter range, while also reducing the effect of the synthetic gap. Of course, this NN application is not limited to the APOGEE wavelength region, but can be extended to other spectroscopic surveys.

Finally, some scepticism exists around machine learning due to its historical inaccessibility and complexity; sometimes being perceived as a black box. New developments have improved usability and made machine learning much more accessible. The training and testing of all of the models used in this paper were performed with the NN library, KERAS (Chollet 2015), which provides a high-level application programme interface to the TENSORFLOW (Abadi et al. 2015) machine intelligence software package. We provide code to reproduce steps of training and testing on the APOGEE DR13 data set with StarNet at <https://github.com/astroai/starnet>.

## 6 CONCLUSION

We have presented a CNN model that is capable of determining the stellar parameters  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  directly from spectra. Our development of this model, StarNet, includes training and testing on real spectra (from the APOGEE data base) and synthetic spectra (from the APOGEE ASSET grid). We have implemented an optimization procedure for hyper-parameter tuning, examining the precision in the stellar parameter predictions when developing the StarNet architecture.

By applying this StarNet model to various data releases from the APOGEE survey, we show that it is capable of estimating stellar parameters from *synthetic spectra*. These parameters and the uncertainties are similar to the ASPCAP results. We also compare StarNet to other *data-driven* methods (such as the Cannon 2) and find similar results when trained on sufficiently large observational data sets.

In our applications on the APOGEE data sets, we also explore the limitations of the training sets; the sample sizes; the stellar parameter ranges; the impact of zero-flux values in real spectra during training and during testing; and the effect of S/N of the spectra on the accuracy of the StarNet predictions. Our results are robust over a large range of S/N values for the APOGEE spectra, including low S/N  $\sim 15$ .

In the future, testing of the intrinsic errors in NN models should be further examined, as well as variations in the synthetic spectral grids; e.g. variations in the model atmospheres adopted when generating the synthetic spectra, changes in the continuum normalization methods, as well as the exploration of different wavelength regions. We are especially interested in applying StarNet to optical spectroscopic surveys and observations.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for helpful comments that have improved this manuscript. This paper was first presented at l’Observatoire de la Côte d’Azur in Nice, France, on 2017 June 1 and again (in part) at the Joint Institute for Nuclear Astrophysics Forging Connections meeting in Lansing, Michigan, on 2017 June 28. The authors wish to thank Dr. Mike Irwin (Institute of Astronomy, University of Cambridge) and Dr. Jo Bovy (University of Toronto) for helpful comments on this work, and Drs. Vanessa Hill, Alejandra Recio-Blanco, and Patrick de Laverny (Observatoire de la Côte d’Azur, France) for many interesting discussions on spectral analyses. SB and TO thank the National Research Council of Canada Herzberg Astronomy and Astrophysics for funding through their co-op programme. K.A.V., C.K., F.J., S.B., T.O., and S.M. acknowledge partial funding for this work through an National Sciences and Engineering Research Council Collaborative Research and Training Experience Program award in *New Technologies for Canadian Observatories*.

## REFERENCES

- Abadi M. et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, available at: <http://tensorflow.org/>
- Albareti F. D. et al., 2017, *ApJS*, 233, 25
- Allende-Prieto C. et al., 2008, *AJ*, 136, 2070
- Aoki W. et al., 2013, *AJ*, 145, 13
- Bailer-Jones C. A., 2000, *A&A*, 357, 197
- Bailer-Jones C. A., Irwin M., Gilmore G., von Hippel T., 1997, *MNRAS*, 292, 157
- Bergstra J. S., Bardenet R., Bengio Y., Kégl B., 2011, in Shawe-Taylor J., Zemel R. S., Bartlett P. L., Pereira F., Weinberger K. Q., eds, *Advances in Neural Information Processing Systems 24*. Curran Associates, Inc., New York, p. 2546
- Bessell M., 1983, *PASP*, 95, 94
- Bragaglia A. et al., 2001, *AJ*, 121, 327
- Brown J. A., Wallerstein G., Oke J., 1991, *AJ*, 101, 1693
- Carretta E., Gratton R., 1997, *A&AS*, 121, 95
- Carretta E., Bragaglia A., Gratton R., 2007, *A&A*, 473, 129
- Carretta E. et al., 2009, *A&A*, 505, 117
- Casey A. R., Hogg D. W., Ness M., Rix H.-W., Ho A. Q., Gilmore G., 2016, *ApJ*, preprint (<arXiv:1603.03040>)
- Cavallo R. M., Nagar N. M., 2000, *AJ*, 120, 1364
- Chen S., Billings S., Grant P., 1990, *Int. J. Control.*, 51, 1191
- Chollet F., 2015, Keras, available at: <https://github.com/fchollet/keras>
- Cohen J., 1978, *ApJ*, 223, 487
- Cohen J., 1979, *ApJ*, 231, 751
- Cohen J., 1980, *ApJ*, 241, 981
- Cunha K. et al., 2015, *ApJ*, 798, L41
- Dotter A., Chaboyer B., Jevremović D., Kostov V., Baron E., Ferguson J. W., 2008, *ApJS*, 178, 89
- Garcia-Pérez A. E. et al., 2016, *AJ*, 151, 144
- Gilmore G. et al., 2012, *Messenger*, 147, 25
- Goodfellow I., Bengio Y., Courville A., 2016, *Deep Learning*. MIT Press, Cambridge, MA
- Gratton R., 1982, *ApJ*, 257, 640
- Gratton R. G., Quarta M. L., Ortolani S., 1986, *A&A*, 169, 208
- Gratton R. G., Contarini G., 1994, *A&A*, 283, 911
- Gratton R., Bragaglia A., Carretta E., Tosi M., 2006, *ApJ*, 642, 462
- Gurney K., 1997, *An Introduction to Neural Networks*. CRC Press, London
- He K., Zhang X., Ren S., Sun J., 2015, *CoRR*, abs/1502.01852
- Holtzman J. A. et al., 2015, *AJ*, 150, 148
- Howes L. M. et al., 2014, *MNRAS*, 445, 4241
- Jahandar F. et al., 2017, *MNRAS*, 470, 4782
- Kingma D., Ba J., 2014, preprint (<arXiv:1412.6980>)
- Koesterke L., Prieto C. A., Lambert D. L., 2008, *ApJ*, 680, 764
- Kraft R. P., Sneden C., Langer G., Prosser C. F., 1992, *AJ*, 104, 645
- Kraft R. P., Sneden C., Langer G., Shetrone M. D., 1993, *AJ*, 106, 1490
- Kraft R. P., Sneden C., Smith G. H., Shetrone M. D., Langer G., Pilachowski C. A., 1997, *AJ*, 113, 279
- Kraft R. P., Peterson R. C., Guhathakurta P., Sneden C., Fulbright J. P., Langer G. E., 1999, *ApJ*, 518, L53
- Lai D. K., Smith G. H., Bolte M., Johnson J. A., Lucatello S., Kraft R. P., Sneden C., 2011, *AJ*, 141, 62
- Lamb M., Venn K., Shetrone M., Sakari C., Pritzl B., 2015, *MNRAS*, 448, 42
- Lee Y. et al., 2008, *AJ*, 136, 2022
- Lee Y. et al., 2011, *AJ*, 141, 90
- Leep E. M., Wallerstein G., Oke J., 1987, *AJ*, 93, 338
- Lehnert M., Bell R., Cohen J., 1991, *ApJ*, 367, 514
- Li X.-R., Pan R.-Y., Duan F.-Q., 2017, *Research in A&A*, 17, 036
- Luck R., Bond H., 1981, *ApJ*, 244, 919
- Maaten L. v. d., Hinton G., 2008, *J. Mach. Learn. Res.*, 9, 2579
- Majewski S. R. et al., 2017, *AJ*, 154, 94
- Manteiga M., Ordóñez D., Dafonte C., Arcay B., 2010, *PASP*, 122, 608
- Mészáros S. et al., 2012, *AJ*, 144, 120
- Mészáros S. et al., 2013, *AJ*, 146, 133
- Mészáros S. et al., 2015, *AJ*, 149, 153
- Minniti D., Olszewski E. W., Rieke M., 1993, *ApJ*, 410, L79
- Molenda-Żakowicz J. et al., 2013, *MNRAS*, stt1095
- Ness M., Hogg D. W., Rix H.-W., Ho A. Y., Zasowski G., 2015, *ApJ*, 808, 16
- Nidever D. L. et al., 2015, *AJ*, 150, 173
- Pancino E., Carrera R., Rossetti E., Gallart C., 2010, *A&A*, 511, A56
- Pancino E. et al., 2017, *A&A*, 598, 5
- Peterson R. C., 1980, *ApJ*, 237, L87
- Pilachowski A., 1985, *PASP*, 97, 801
- Pilachowski C., Leep E., Wallerstein G., 1980, *ApJ*, 236, 508
- Pilachowski C., Sneden C., Wallerstein G., 1983, *ApJS*, 52, 241
- Randich S., Sestito P., Primas F., Pallavicini R., Pasquini L., 2006, *A&A*, 450, 557
- Recio-Blanco A. et al., 2016, *A&A*, 589, 22
- Recio-Blanco A. et al., 2017, *A&A*, 602, 6
- Roederer I. U., Sneden C., 2011, *AJ*, 142, 22
- Roederer I. U., Preston G. W., Thompson I. B., Shectman S. A., Sneden C., Burley G. S., Kelson D. D., 2014, *AJ*, 147, 136
- Sneden C., Bolte M., Stetson P. B., 1998, *AJ*, 115, 1888
- Singh H. P., Gulati R. K., Gupta R., 1998, *MNRAS*, 295, 312
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Smiljanic R. et al., 2014, *A&A*, 570, A122
- Sneden C., Kraft R. P., Prosser C. F., Langer G., 1991, *AJ*, 102, 2001
- Sneden C., Kraft R. P., Prosser C. F., Langer G., 1992, *AJ*, 104, 2121
- Sneden C., Kraft R. P., Prosser C. F., Shetrone M. D., 1994, *AJ*, 107, 1773
- Sneden C., Kraft R. P., Shetrone M. D., Smith G. H., Langer G., Prosser C. F., 1997, *AJ*, 114, 1964
- Sneden C., Pilachowski C. A., Kraft R. P., 2000, *AJ*, 120, 1351
- Sneden C., Kraft R. P., Guhathakurta P., Peterson R. C., Fulbright J. P., 2004, *AJ*, 127, 2162
- Sobeck J. S. et al., 2011, *AJ*, 141, 175

- Soubiran C., Le Campion J.-F., Brouillet N., Chemin L., 2016, *A&A*, 591, A118  
 Venn K. A. et al., 2012, *ApJ*, 715, 25  
 Von Hippel T., Storrie-Lombardi L., Storrie-Lombardi M., Irwin M., 1994, *MNRAS*, 269, 97  
 Wang K., Guo P., Luo A.-L., 2017, *MNRAS*, 465, 4311  
 Yanny B. et al., 2009, *AJ*, 137, 4377  
 Yong D. et al., 2013, *ApJ*, 762, 27

## APPENDIX A: THE STARNET CNN

In this paper, we approach the prediction of stellar parameters from spectra with supervised learning. From a data set consisting of individual stellar spectra and associated physical parameters, we approximate a function to learn the mapping between the two. For other similar spectra, we then assume the trained model is capable of predicting stellar parameters.

Our mapping function is parametrized and organized as an NN. It consists of a collection of artificial neurons or nodes arranged in layers: an *input layer* (in our case the input spectra), a number of *hidden layers*, and an *output layer* (in our case the stellar parameters). Each node is parametrized as a linear function of weights,  $\mathbf{w}$ , applied to an input,  $\mathbf{x}$ , with an additional bias value,  $b$ . A node is then activated by a non-linear function,  $g$ , giving an output of a node to be written as

$$h(\mathbf{x}) = g(\mathbf{w}^T \mathbf{x} + b).$$

Common activation functions include the sigmoid function and the ReLU:

$$g(z) = \max(0, z),$$

which allow the network to adapt to non-linear problems (Chen, Billings & Grant 1990). The ReLU function was used in the implementation for all StarNet layers except for the last output layer.

In a traditional sequential NN architecture, each node is connected to every node from the previous layer as well as every node in the following layer, thereafter referred to as *fully connected* layers. At hidden layer,  $l$ , the output,  $\mathbf{h}^{(l)}$ , is a vector valued function of the previous layer,  $\mathbf{h}^{(l-1)}$ , and is given by

$$\mathbf{h}^{(l)} = g(\mathbf{w}^{(l)} \mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}).$$

The first layer is simply the input  $\mathbf{h}^{(0)}(\mathbf{x}) = \mathbf{x}$ ; in our case, the spectra.

The next two layers of StarNet are *convolutional layers*, which are more adapted to higher dimensional inputs by leveraging local connectivity in the previous layer. In convolutional layers, the weights are applied as filters. The filter slides across the previous layer taking the dot product between the filter weights and sections of the input. For a given filter covering a section,  $s$ , this operation can be summarized as

$$\mathbf{h}_s^{(l)} = g(\mathbf{w}_s^{(l)} \otimes \mathbf{h}^{(l-1)} + \mathbf{b}_s).$$

These filters allow for the extraction of features in the input and learn which features to extract through training. After the convolutional layers in StarNet, we use a max pooling layer. A max pooling layer is a non-linear downsampling technique typically used in CNNs to decrease the number of free parameters and to extract the strongest features from a convolutional layer. In a max pooling layer, a window moves along the feature map generated by each of the filters in the previous convolutional layer – in strides of length equal to the length of the window – extracting the maximum value from each sub-region. These pools of maxima are then passed on

to the following layer. The next two layers in StarNet are fully connected layers.

The combination of all those layers allows for the formation of non-linear combinations of an input vector,  $\mathbf{x}$ , to produce an output vector prediction,  $f(\mathbf{x}; \mathbf{w}, b)$ . For each training sample of spectra,  $\mathbf{x}_t$ , and corresponding known stellar parameters,  $\mathbf{y}_t$ , the NN model weights and biases are estimated by minimizing the *empirical risk* that computes the *loss* between the predictions and targets for a batch of  $T$  training samples, often supplemented with a regularizing function. We ended up adopting a MSE loss function without regularization for StarNet, such that the StarNet empirical risk to be minimized reads:

$$\arg_{\mathbf{w}, b} \min \frac{1}{T} \sum_{t=1}^T (\mathbf{y}_t - f(\mathbf{x}_t; \mathbf{w}, b))^2.$$

The minimization is performed with a stochastic gradient descent (SGD) algorithm. SGD algorithms require the computation of the loss function gradients with respect to the weights and make adjustments to those weights iteratively until reaching a minimum. In our case, the optimization is performed using the ADAM optimizer (Kingma & Ba 2014), an SGD variant using adaptive estimates of the gradient moments to adjust learning rates. Initially, the weights of the model are randomly set and therefore the predictions will be quite poor. Computing the gradients is operated backwards through each sequential layer, a process referred as *back-propagation*, and is the computationally expensive part of the training.

In the case of StarNet, a cross-validation set was used to test the model following every iteration to evaluate whether or not the model had improved; if improvements were not made after a given number of iterations, the training was stopped. This minimum may differ depending on the complexity of the model architecture as well as various hyper-parameters (discussed in Section 4.3). Following each iteration, the cross-validation set is sent through a single forward propagation, where the outputs are predicted and compared against the target values. This set is not used for training, but only to ensure that the model is not overfitting to the training set. Overfitting occurs when a model learns a function that may be able to compute the outputs for the training set very well, but cannot generalize that function to be applied to a test set that is not included in the training. A cross-validation set is used as a type of middle-ground between the training and test set, and ensures that overfitting does not occur. If the cross-validation predictions do not improve after several iterations, the training will be stopped. Using a cross-validation set to avoid overfitting and tuning hyper-parameters is common practice in machine learning applications (Gurney 1997).

Once a minimum has been reached, a separate test set – not included in training or cross-validation – is selected to ensure that the model performs well on new data. For StarNet, this test set consists of stellar spectra with ‘known’ stellar parameters that are used to compare with StarNet predictions. For more in-depth explanation of deep NNs, we refer the reader to Goodfellow, Bengio & Courville (2016).

## APPENDIX B: OPTICAL BENCHMARK STARS IN CLUSTERS

In this appendix, we tabulate (see Table B1) the stellar parameters for stars in globular clusters that have both APOGEE DR13 results (based on IR spectral analyses) and results in the literature (based on high-resolution optical spectral analyses). Our systematic

**Table B1.** Stellar parameters for stars in globular clusters that have both APOGEE DR13 results and those based on high-resolution optical spectroscopy in the literature.

| 2MASS ID           | Cluster               | APOGEE<br>PERS<br>Flag | ASPCAP DR13 Parameters  |                |                 | Optical Parameters      |                |                 | StarNet <sub>ASPCAP</sub> |                |                 | StarNet <sub>Synth</sub> |                |                 | References             |
|--------------------|-----------------------|------------------------|-------------------------|----------------|-----------------|-------------------------|----------------|-----------------|---------------------------|----------------|-----------------|--------------------------|----------------|-----------------|------------------------|
|                    |                       |                        | T <sub>eff</sub><br>(K) | log g<br>(dex) | [Fe/H]<br>(dex) | T <sub>eff</sub><br>(K) | log g<br>(dex) | [Fe/H]<br>(dex) | T <sub>eff</sub><br>(K)   | log g<br>(dex) | [Fe/H]<br>(dex) | T <sub>eff</sub><br>(K)  | log g<br>(dex) | [Fe/H]<br>(dex) |                        |
|                    | Representative Errors |                        | 92                      | 0.11           | 0.05            | 86                      | 0.24           | 0.17            | 47                        | 0.14           | 0.07            | —                        | —              | —               | —                      |
| 2M16412709+3628002 | M13                   | —                      | 4184                    | 0.81           | -1.54           | 4348                    | 0.98           | -1.54           | 4201                      | 0.96           | -1.48           | 4405                     | 0.99           | -1.61           | h, ac, af, al          |
| 2M16413053+3629434 | M13                   | —                      | 4431                    | 1.55           | -1.39           | 4410                    | 1.2            | -1.44           | 4434                      | 1.49           | -1.34           | 4429                     | 1.47           | -1.42           | ac, af                 |
| 2M16413072+3630075 | M13                   | High                   | 4816                    | 2.21           | -1.29           | 4750                    | 1.9            | -1.5            | 4809                      | 2.2            | -1.27           | 4742                     | 1.9            | -1.36           | k                      |
| 2M16413082+3630130 | M13                   | —                      | 4658                    | 1.45           | -1.58           | 4825                    | 1.77           | -1.55           | 4665                      | 1.74           | -1.56           | 4869                     | 1.98           | -1.6            | k, ag                  |
| 2M16413476+3627596 | M13                   | —                      | 4213                    | 1.2            | -1.33           | 4230                    | 0.85           | -1.44           | 4205                      | 1.14           | -1.3            | 4249                     | 1.02           | -1.34           | ad, af                 |
| 2M16413482+3627197 | M13                   | High                   | 4171                    | 1.17           | -1.34           | 4173                    | 0.7            | -1.49           | 4157                      | 1.12           | -1.3            | 4220                     | 1.0            | -1.37           | j, ad, af              |
| 2M16413684+3629289 | M13                   | —                      | 4547                    | 1.68           | -1.43           | 4750                    | 2.0            | -1.48           | 4582                      | 1.89           | -1.5            | 4869                     | 2.25           | -1.53           | ag                     |
| 2M16413707+3630378 | M13                   | —                      | 4550                    | 1.62           | -1.48           | 4700                    | 1.7            | -1.58           | 4597                      | 1.87           | -1.49           | 4829                     | 2.16           | -1.51           | ag                     |
| 2M16413870+3625380 | M13                   | High                   | 4168                    | 1.17           | -1.31           | 4111                    | 0.67           | -1.56           | 4134                      | 1.09           | -1.29           | 4205                     | 0.95           | -1.33           | f, j, ac, ad, af, al   |
| 2M16413945+3630209 | M13                   | —                      | 4883                    | 2.14           | -1.41           | 4850                    | 2.1            | -1.56           | 4839                      | 2.21           | -1.44           | 4970                     | 2.47           | -1.45           | ag                     |
| 2M16414398+3622338 | M13                   | High                   | 4568                    | 1.49           | -1.47           | 4600                    | 1.4            | -1.55           | 4638                      | 1.62           | -1.45           | 4712                     | 1.63           | -1.47           | k                      |
| 2M16414478+3623273 | M13                   | High                   | 4529                    | 1.67           | -1.47           | 4625                    | 1.65           | -1.6            | 4522                      | 1.57           | -1.49           | 4657                     | 1.58           | -1.59           | k, ag                  |
| 2M16414528+3631068 | M13                   | —                      | 4679                    | 1.8            | -1.41           | 4600                    | 1.6            | -1.57           | 4742                      | 2.0            | -1.34           | 4701                     | 2.07           | -1.35           | k, ag                  |
| 2M16414558+3630328 | M13                   | —                      | 4711                    | 1.86           | -1.43           | 4750                    | 1.9            | -1.57           | 4781                      | 2.19           | -1.35           | 4733                     | 2.18           | -1.37           | k                      |
| 2M16414744+3628284 | M13                   | —                      | 4214                    | 1.21           | -1.34           | 4180                    | 0.8            | -1.46           | 4203                      | 1.14           | -1.32           | 4261                     | 1.14           | -1.34           | ad, af                 |
| 2M16414932+3625264 | M13                   | High                   | 4686                    | 1.99           | -1.3            | 4600                    | 1.6            | -1.58           | 4736                      | 2.05           | -1.3            | 4741                     | 1.92           | -1.27           | k                      |
| 2M16415037+3623417 | M13                   | High                   | 4702                    | 1.9            | -1.4            | 4700                    | 1.9            | -1.54           | 4771                      | 1.94           | -1.35           | 4697                     | 1.76           | -1.37           | ag                     |
| 2M16415160+3629363 | M13                   | High                   | 4914                    | 2.02           | -1.38           | 4975                    | 1.7            | -1.68           | 4782                      | 1.95           | -1.51           | 4923                     | 1.62           | -1.61           | ag                     |
| 2M16415239+3628395 | M13                   | High                   | 4668                    | 1.96           | -1.39           | 4800                    | 2.0            | -1.52           | 4631                      | 1.85           | -1.52           | 4858                     | 2.04           | -1.46           | k                      |
| 2M16415842+3628312 | M13                   | High                   | 4860                    | 2.28           | -1.32           | 4900                    | 2.0            | -1.53           | 4754                      | 2.18           | -1.36           | 4901                     | 2.05           | -1.48           | k                      |
| 2M16415862+3627465 | M13                   | High                   | 4719                    | 1.94           | -1.49           | 4850                    | 1.9            | -1.64           | 4618                      | 1.89           | -1.43           | 5002                     | 2.35           | -1.52           | ag                     |
| 2M13131736+1814463 | M53                   | —                      | 4314                    | 0.74           | -1.84           | 4425                    | 1.06           | -2.04           | 4414                      | 1.04           | -1.79           | 4620                     | 1.19           | -1.9            | ah                     |
| 2M07380627+2136542 | N2420                 | —                      | 4725                    | 2.55           | -0.22           | 4850                    | 2.6            | -0.07           | 4725                      | 2.56           | -0.22           | 4696                     | 2.57           | -0.22           | am                     |
| 2M07381549+2138015 | N2420                 | —                      | 4872                    | 2.57           | -0.21           | 4800                    | 2.6            | -0.06           | 4890                      | 2.57           | -0.21           | 4855                     | 2.71           | -0.18           | am                     |
| 2M07382696+2138244 | N2420                 | —                      | 4832                    | 2.48           | -0.18           | 4800                    | 2.6            | -0.03           | 4825                      | 2.48           | -0.17           | 4824                     | 2.65           | -0.16           | am                     |
| 2M12100405+1832532 | N4147                 | High                   | 4235                    | 1.12           | -1.53           | 4383                    | 1.1            | -1.3            | 4270                      | 0.92           | -1.73           | 4458                     | 0.79           | -1.86           | ak                     |
| 2M13414576+2824597 | M3                    | —                      | 4051                    | 0.84           | -1.46           | 4137                    | 0.69           | -1.62           | 4057                      | 0.89           | -1.46           | 4277                     | 0.76           | -1.52           | j, r, x, y, ac, af, al |
| 2M13421204+2826265 | M3                    | —                      | 3968                    | 0.95           | -1.27           | 4000                    | 0.54           | -1.64           | 3975                      | 0.87           | -1.27           | 4172                     | 0.55           | -1.42           | r, x, y, ac, af, al    |
| 2M13421679+2823479 | M3                    | —                      | 4123                    | 1.01           | -1.33           | 4250                    | 0.93           | -1.42           | 4119                      | 0.95           | -1.32           | 4260                     | 0.97           | -1.4            | j, ac, af              |
| 2M13423922+2827574 | M3                    | Low                    | 4104                    | 1.06           | -1.27           | 4175                    | 0.63           | -1.48           | 4110                      | 1.03           | -1.3            | 4263                     | 0.94           | -1.35           | x, z                   |
| 2M13424150+2819081 | M3                    | —                      | 3853                    | 0.55           | -1.17           | 3966                    | 0.45           | -1.53           | 3910                      | 0.98           | -1.12           | 4110                     | 0.51           | -1.27           | r, x, y, ac, af        |
| 2M14052071+2829419 | N5466                 | High                   | 4341                    | 1.28           | -1.8            | 4499                    | 1.27           | -1.94           | 4472                      | 1.38           | -1.67           | 4846                     | 1.76           | -1.87           | ah                     |
| 2M08514401+1146245 | M67                   | High                   | 5476                    | 3.93           | -0.02           | 5541                    | 3.8            | 0.02            | 5363                      | 3.62           | -0.1            | 5242                     | 3.7            | -0.13           | i                      |
| 2M15174702+0204519 | MSPAL5                | —                      | 4361                    | 1.22           | -1.25           | 4381                    | 1.05           | -1.53           | 4362                      | 1.27           | -1.25           | 4528                     | 1.47           | -1.28           | u                      |
| 2M15180831+0158530 | MSPAL5                | —                      | 4840                    | 1.73           | -1.37           | 4961                    | 1.84           | -1.59           | 4861                      | 1.9            | -1.39           | 5164                     | 2.36           | -1.25           | u                      |
| 2M15180987+0210088 | MSPAL5                | Low                    | 4557                    | 1.45           | -1.3            | 4630                    | 1.44           | -1.56           | 4558                      | 1.53           | -1.32           | 4751                     | 1.77           | -1.32           | u                      |
| 2M15181075+0212356 | MSPAL5                | Low                    | 4698                    | 1.85           | -1.18           | 4740                    | 1.82           | -1.37           | 4710                      | 1.99           | -1.13           | 4843                     | 2.43           | -1.13           | u                      |
| 2M15181867+0204327 | MSPAL5                | Low                    | 3995                    | 1.04           | -1.05           | 4113                    | 0.73           | -1.15           | 3996                      | 1.09           | -1.06           | 4184                     | 0.69           | -1.19           | l, af, ak              |
| 2M15182262+0200305 | MSPAL5                | —                      | 4035                    | 1.05           | -1.03           | 4100                    | 0.66           | -1.3            | 4034                      | 1.11           | -1.02           | 4203                     | 0.94           | -1.12           | u                      |
| 2M15182283+0203097 | MSPAL5                | —                      | 4251                    | 1.35           | -1.11           | 4250                    | 1.1            | -1.12           | 4234                      | 1.3            | -1.11           | 4352                     | 1.43           | -1.09           | l, af                  |
| 2M15182345+0159572 | MSPAL5                | —                      | 4483                    | 1.76           | -1.02           | 4300                    | 1.3            | -1.11           | 4459                      | 1.69           | -1.01           | 4381                     | 1.61           | -1.05           | l, af                  |
| 2M15182591+0205076 | MSPAL5                | —                      | 4198                    | 1.37           | -1.06           | 4300                    | 0.8            | -1.02           | 4175                      | 1.23           | -1.07           | 4283                     | 1.31           | -1.09           | l, af                  |
| 2M15183223+0201341 | MSPAL5                | —                      | 4060                    | 1.05           | -1.06           | 4200                    | 1.0            | -1.09           | 4053                      | 1.13           | -1.06           | 4236                     | 0.98           | -1.13           | l, af                  |
| 2M15183531+0207400 | MSPAL5                | —                      | 4327                    | 1.51           | -1.19           | 4409                    | 1.3            | -1.35           | 4340                      | 1.45           | -1.19           | 4487                     | 1.7            | -1.21           | af, aj, ak             |
| 2M15183738+0206079 | MSPAL5                | Low                    | 4241                    | 1.1            | -1.17           | 4236                    | 0.61           | -1.49           | 4240                      | 1.08           | -1.2            | 4332                     | 1.09           | -1.19           | u                      |
| 2M15183765+0201212 | MSPAL5                | —                      | 4839                    | 2.31           | -1.25           | 4845                    | 2.2            | -1.46           | 4835                      | 2.28           | -1.27           | 5132                     | 2.7            | -1.22           | u                      |
| 2M15184132+0205014 | MSPAL5                | —                      | 4534                    | 1.73           | -1.19           | 4460                    | 1.42           | -1.34           | 4534                      | 1.82           | -1.2            | 4589                     | 1.96           | -1.16           | u, af, aj              |
| 2M15184164+0203533 | MSPAL5                | —                      | 4128                    | 1.19           | -1.07           | 4128                    | 0.83           | -1.11           | 4103                      | 1.09           | -1.07           | 4216                     | 1.05           | -1.13           | l, af, ak              |
| 2M15184346+0203074 | MSPAL5                | —                      | 4581                    | 1.82           | -1.12           | 4475                    | 1.3            | -1.38           | 4615                      | 1.91           | -1.13           | 4540                     | 1.9            | -1.1            | u                      |
| 2M15184374+0208171 | MSPAL5                | Low                    | 4938                    | 2.37           | -1.19           | 4860                    | 2.25           | -1.4            | 4865                      | 2.29           | -1.18           | 5036                     | 3.15           | -1.04           | u                      |
| 2M15184495+0202034 | MSPAL5                | —                      | 3990                    | 1.17           | -1.0            | 4017                    | 0.55           | -1.3            | 3981                      | 1.01           | -1.02           | 4123                     | 0.79           | -1.14           | ak, al                 |
| 2M15184540+0204302 | MSPAL5                | —                      | 4223                    | 1.14           | -1.21           | 4266                    | 0.8            | -1.27           | 4222                      | 1.11           | -1.25           | 4363                     | 1.09           | -1.25           | l, u, af, ak, al       |
| 2M15185515+0214337 | MSPAL5                | —                      | 4487                    | 1.38           | -1.28           | 4584                    | 1.3            | -1.5            | 4489                      | 1.51           | -1.31           | 4669                     | 1.75           | -1.29           | u                      |
| 2M17163427+4407363 | M92                   | —                      | 4638                    | 1.19           | -2.24           | 4750                    | 1.63           | -2.54           | 4791                      | 1.78           | -1.86           | 4771                     | 1.41           | -2.14           | ab, ae                 |
| 2M17165185+4308031 | M92                   | —                      | 4620                    | 1.41           | -2.24           | 4740                    | 1.75           | -2.72           | 4761                      | 1.73           | -1.81           | 4765                     | 1.98           | -2.03           | ae                     |
| 2M17165557+4309277 | M92                   | —                      | 4264                    | 0.6            | -2.19           | 4340                    | 0.35           | -2.27           | 4382                      | 1.05           | -1.93           | 4454                     | 0.74           | -2.3            | x                      |
| 2M17165738+4307236 | M92                   | —                      | 4254                    | 0.67           | -2.15           | 4335                    | 0.8            | -2.28           | 4352                      | 1.03           | -1.95           | 4430                     | 0.81           | -2.28           | n, x, ab, af           |
| 2M17165772+4314115 | M92                   | —                      | 4443                    | 1.11           | -2.21           | 4615                    | 1.48           | -2.34           | 4515                      | 1.34           | -1.89           | 4612                     | 1.41           | -2.24           | ab, ae                 |
| 2M17165883+4315116 | M92                   | —                      | 4513                    | 1.08           | -2.2            | 4685                    | 1.55           | -2.34           | 4746                      | 1.65           | -1.78           | 4654                     | 1.27           | -1.99           | ab, ae                 |
| 2M17165956+4306456 | M92                   | —                      | 4618                    | 1.35           | -2.32           | 4880                    | 1.9            | -2.34           | 4723                      | 1.63           | -1.88           | 4979                     | 1.99           | -2.31           | ab                     |
| 2M17165967+4301058 | M92                   | High                   | 4746                    | 1.77           | -2.11           | 4875                    | 1.9            | -2.34           | 4803                      | 1.87           | -1.73           | 4904                     | 2.32           | -1.92           | ab, ae                 |
| 2M17170033+4311478 | M92                   | —                      | 4680                    | 1.26           | -2.19           | 4830                    | 1.75           | -2.54           | 4677                      | 1.54           | -1.86           | 4817                     | 1.77           | -2.0            | ab, ae                 |
| 2M17170043+4305117 | M92                   | —                      | 4601                    | 1.35           | -2.19           | 4725                    | 1.65           | -2.5            | 4727                      | 1.77           | -1.84           | 4626                     | 1.56           | -2.08           | ab, ae                 |
| 2M17170081+4310251 | M92                   | —                      | 4531                    | 1.19           | -2.18           | 4630                    | 1.52           | -2.54           | 4655                      | 1.6            | -1.88           | 4679                     | 1.52           | -2.08           | ab, ae                 |
| 2M17170647+4306029 | M92                   | —                      | 4613                    | 1.37           | -2.22           | 4780                    | 1.75           | -2.55           | 4679                      | 1.61           | -1.88           | 4812                     | 1.88           | -2.06           | ae                     |
| 2M17170731+4309308 | M92                   | —                      | 4447                    | 1.05           | -2.16           | 4540                    | 1.2            | -2.34           | 4532                      | 1.33           | -1.9            | 4542                     | 1.16           | -2.11           | ab                     |
| 2M17171221+4302209 | M92                   | High                   | 4833                    | 1.93           | -2.08           | 4940                    | 2.0            | -2.35           | 4768                      |                |                 |                          |                |                 |                        |

**Table B1** – *continued*

| 2MASS ID           | Cluster | APOGEE       | ASPCAP DR13 Parameters  |                   |                 | Optical Parameters      |                   |                 | StarNet <sub>ASPCAP</sub> |                   |                 | StarNet <sub>Synth</sub> |                   |                 | References       |
|--------------------|---------|--------------|-------------------------|-------------------|-----------------|-------------------------|-------------------|-----------------|---------------------------|-------------------|-----------------|--------------------------|-------------------|-----------------|------------------|
|                    |         |              | $T_{\text{eff}}$<br>(K) | $\log g$<br>(dex) | [Fe/H]<br>(dex) | $T_{\text{eff}}$<br>(K) | $\log g$<br>(dex) | [Fe/H]<br>(dex) | $T_{\text{eff}}$<br>(K)   | $\log g$<br>(dex) | [Fe/H]<br>(dex) | $T_{\text{eff}}$<br>(K)  | $\log g$<br>(dex) | [Fe/H]<br>(dex) |                  |
|                    |         | PERS<br>Flag |                         |                   |                 |                         |                   |                 |                           |                   |                 |                          |                   |                 |                  |
| 2M21295562+1210455 | M15     | –            | 4250                    | 0.56              | −2.16           | 4200                    | 0.15              | −2.37           | 4387                      | 1.04              | −1.92           | 4376                     | 0.63              | −2.1            | o, o, ab         |
| 2M21295618+1212337 | M15     | –            | 4327                    | 0.77              | −2.21           | 4418                    | 0.76              | −2.29           | 4489                      | 1.17              | −1.92           | 4456                     | 0.9               | −2.09           | n, n, o, o, af   |
| 2M21295666+1209463 | M15     | High         | 4120                    | 0.28              | −2.34           | 4269                    | 0.3               | −2.44           | 4274                      | 0.81              | −1.94           | 4381                     | 0.66              | −2.33           | m, o, o, ab      |
| 2M21295678+1210269 | M15     | –            | 4134                    | 0.04              | −2.27           | 4325                    | 0.45              | −2.42           | 4310                      | 0.76              | −1.98           | 4349                     | 0.33              | −2.22           | o, o             |
| 2M21295801+1214260 | M15     | –            | 4605                    | 1.29              | −2.26           | 4855                    | 2.0               | −2.37           | 4569                      | 1.32              | −1.95           | 4929                     | 2.03              | −2.19           | ab               |
| 2M21295856+1209214 | M15     | –            | 4197                    | 0.41              | −2.23           | 4300                    | 0.3               | −2.43           | 4373                      | 0.97              | −1.97           | 4392                     | 0.73              | −2.14           | m, o, o, ab      |
| 2M21300033+1210508 | M15     | –            | 4292                    | 0.68              | −2.17           | 4341                    | 0.43              | −2.4            | 4427                      | 1.08              | −1.94           | 4434                     | 0.83              | −2.12           | d, o, o          |
| 2M21300038+1207363 | M15     | –            | 4394                    | 0.72              | −2.26           | 4552                    | 1.24              | −2.4            | 4546                      | 1.16              | −2.0            | 4628                     | 1.05              | −2.22           | d, ab            |
| 2M21300224+1211215 | M15     | Low          | 4084                    | 0.18              | −2.29           | 4288                    | 0.62              | −2.39           | 4255                      | 0.78              | −1.96           | 4371                     | 0.59              | −2.32           | d, o, o, ab      |
| 2M21300274+1210438 | M15     | High         | 4271                    | 0.59              | −2.24           | 4275                    | 0.5               | −2.39           | 4432                      | 1.05              | −1.94           | 4389                     | 0.66              | −2.16           | o, o, ab         |
| 2M21300637+1206592 | M15     | High         | 4430                    | 0.6               | −2.33           | 4625                    | 1.3               | −2.4            | 4640                      | 1.2               | −1.9            | 4515                     | 0.66              | −2.11           | o, o, ab         |
| 2M21300696+1207465 | M15     | –            | 4709                    | 0.85              | −2.3            | 4940                    | 1.5               | −2.37           | 4656                      | 1.02              | −1.93           | 4908                     | 0.83              | −2.14           | ab               |
| 2M21301049+1210061 | M15     | –            | 4486                    | 1.02              | −2.1            | 4470                    | 1.07              | −2.22           | 4668                      | 1.41              | −1.86           | 4562                     | 1.12              | −1.93           | e, o, af         |
| 2M21304412+1211226 | M15     | –            | 4539                    | 1.01              | −2.08           | 4640                    | 1.4               | −2.37           | 4741                      | 1.54              | −1.88           | 4729                     | 1.27              | −1.95           | ab               |
| 2M19533747+1844596 | M71     | High         | 4069                    | 1.33              | −0.7            | 4200                    | 1.2               | −0.83           | 4068                      | 1.34              | −0.69           | 4127                     | 1.1               | −0.8            | p, af            |
| 2M19533757+1847286 | M71     | High         | 3906                    | 1.22              | −0.58           | 3996                    | 0.88              | −0.43           | 3918                      | 1.15              | −0.56           | 3974                     | 0.84              | −0.68           | p, w, af         |
| 2M19533986+1843530 | M71     | High         | 4230                    | 1.52              | −0.68           | 4300                    | 1.25              | −0.7            | 4204                      | 1.42              | −0.66           | 4257                     | 1.27              | −0.74           | p, af            |
| 2M19534750+1846169 | M71     | High         | 4259                    | 1.6               | −0.65           | 4367                    | 1.55              | −0.72           | 4270                      | 1.58              | −0.65           | 4283                     | 1.25              | −0.76           | a, p, af, aj     |
| 2M19534827+1848021 | M71     | High         | 3979                    | 1.33              | −0.6            | 4055                    | 0.8               | −0.89           | 3968                      | 1.19              | −0.59           | 3988                     | 0.79              | −0.71           | p, aa, af        |
| 2M19535064+1849075 | M71     | High         | 4223                    | 1.71              | −0.65           | 4291                    | 1.4               | −0.83           | 4212                      | 1.55              | −0.65           | 4226                     | 1.37              | −0.71           | a, p, af, aj, ak |
| 2M19535325+1846471 | M71     | High         | 3896                    | 1.15              | −0.63           | 4099                    | 0.8               | −0.89           | 3922                      | 1.11              | −0.61           | 3988                     | 0.8               | −0.74           | a, p, aa, af     |
| 2M19205287+3745331 | N6791   | High         | 4500                    | 2.58              | 0.3             | 4512                    | 2.32              | 0.47            | 4520                      | 2.37              | 0.32            | 4590                     | 2.31              | 0.26            | b, t             |
| 2M19205629+3744334 | N6791   | High         | 4472                    | 2.53              | 0.32            | 4473                    | 2.33              | 0.56            | 4456                      | 2.26              | 0.34            | 4581                     | 2.27              | 0.3             | b, t             |
| 2M19411355+4012205 | N6819   | –            | 4802                    | 2.51              | 0.02            | 4835                    | 2.61              | 0.1             | 4805                      | 2.61              | 0.01            | 4783                     | 2.67              | 0.03            | c                |
| 2M19413031+4009005 | N6819   | –            | 4079                    | 1.55              | 0.04            | 4046                    | 1.77              | −0.18           | 4082                      | 1.54              | 0.04            | 4139                     | 1.57              | 0.04            | v                |
| 2M23565751+5645272 | N7789   | –            | 4560                    | 2.37              | −0.01           | 4345                    | 2.2               | 0.15            | 4562                      | 2.36              | −0.01           | 4626                     | 2.45              | 0.05            | ai               |
| 2M06293009-3116587 | N2243   | –            | 4595                    | 2.08              | −0.47           | 4561                    | 1.9               | −0.69           | 4603                      | 2.14              | −0.48           | 4639                     | 2.08              | −0.46           | g, s             |

Notes. 1. Average error from the ASPCAP DR13 parameters.

2. Average standard deviation of the mean from multiple ( $N \sim 3$ ) optical parameter sources.

3. Average statistical errors for StarNet predictions with respect to ASPCAP DR13 parameters.

Source references: (a) Leep, Wallerstein & Oke (1987); (b) Carretta, Bragaglia & Gratton (2007); (c) Bragaglia et al. (2001); (d) Carretta et al. (2009); (e) Minniti, Olszewski & Rieke (1993); (f) Brown, Wallerstein & Oke (1991); (g) Gratton & Contarini (1994); (h) Peterson et al. (1980); (i) Randich et al. (2006); (j) Cavallo & Nagar (2000); (k) Lehner, Bell & Cohen (1991); (l) Sneden et al. (1992); (m) Sobeck et al. (2011); (n) Sneden et al. (1991); (o) Sneden et al. (1997); (p) Sneden et al. (1994); (q) Cohen (1979); (r) Cohen (1978); (s) Gratton (1982); (t) Gratton et al. (2006); (u) Lai et al. (2011); (v) Molend-Żakowicz et al. (2013); (w) Bessell (1983); (x) Sneden, Pilachowski & Kraft (2000); (y) Luck & Bond (1981); (z) Kraft et al. (1999); (aa) Cohen (1980); (ab) Shetrone, Bolte & Stetson (1998); (ac) Kraft et al. (1992); (ad) Kraft et al. (1993); (ae) Roederer & Sneden (2011); (af) Carretta & Gratton (1997); (ag) Kraft et al. (1997); (ah) Lamb et al. (2015); (ai) Pilachowski (1985); (aj) Gratton, Quarta & Ortolani (1986); (ak) Pilachowski, Sneden & Wallerstein (1983); (al) Pilachowski, Leep & Wallerstein (1980); (am) Pancino et al. (2010).

investigation has revealed numerous stars with incomplete optical parameters in the literature (e.g. missing  $T_{\text{eff}}$ ,  $\log g$ , or [Fe/H]), or incomplete IR parameters in the ASPCAP data base (e.g. presumably due to data acquisition issues such as low S/N or persistence). We have removed those stars from our benchmark sample, as well as targets in the APOGEE data base with a high RV scatter between individual visits (which usually indicates a poten-

tial binary system). Fig. 13 compares the stellar parameters for our final sample of optical benchmark stars (as listed below, with the obvious outliers removed). Small systematic offsets in  $T_{\text{eff}}$ ,  $\log g$ , and [Fe/H] are found, though typically within the quoted  $1\sigma$  errors.

This paper has been typeset from a TeX/LaTeX file prepared by the author.