Deep learning analyses of synthetic spectral libraries with an application to the Gaia-ESO database

Spencer Bialek,1∗ Sébastien Fabbro,1,2 Kim A. Venn,1 Nripesh Kumar,1,3 Teaghan O’Briain,1 Kwang Moo Yi4

1 Department of Physics and Astronomy, University of Victoria, Victoria, BC, V8W 3P2, Canada
2 National Research Council Herzberg Astronomy & Astrophysics, 4071 West Saanich Road, Victoria, BC, Canada
3 Department of Computer Science, National Institute of Technology, Tiruchirappalli, India
4 Department of Computer Science, University of Victoria, Victoria, BC, V8P 5C2, Canada

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

In the era of stellar spectroscopic surveys, synthetic spectral libraries will form the basis for the derivation of the stellar parameters and chemical abundances. In this paper, four popular synthetic grids (INTRIGOSS, FERRE, AMBRE, and PHOENIX) are used in our deep learning prediction framework (StarNet), and compared in an application to optical spectra from the Gaia-ESO survey. The stellar parameters for temperature, surface gravity, metallicity, radial velocity, rotational velocity, and [α/Fe] are determined simultaneously for FGK type dwarfs and giants. StarNet was modified to mitigate the differences in the sampling between the synthetic grids and the observed spectra, by augmenting the grids with realistic observational signatures, in an attempt to incorporate both modelling and statistical uncertainties as part of the training. When applied to spectra from the Gaia-ESO spectroscopic survey and the Gaia-ESO benchmark stars, the INTRIGOSS-trained StarNet showed the best results with the least scatter. Training with the FERRE synthetic grid produces similarly accurate predictions (followed closely by the AMBRE grid), but over a wider range in stellar parameters and spectroscopic wavelengths. In the future, improvements in the underlying physics that generates these synthetic grids will be necessary for consistent high precision stellar parameters and chemical abundances from machine learning and other sophisticated data analysis tools.

Key words: stars: fundamental parameters – stars: abundances – methods: data analysis – techniques: spectroscopic – surveys

1 INTRODUCTION

Astronomy has entered an era of spectroscopic surveys. The first large scale spectroscopic surveys, pioneering new methods to efficiently observe and determine spectroscopic parameters, include the Sloan Digital Sky Survey (SDSS) Sloan Extension for Galactic Understanding and Exploration (SEGUE) surveys of over 200,000 stars (Yanny et al. 2009; Lee et al. 2011) and the RAdial Velocity Experiment (RAVE) survey of nearly 1 million stars (Steinmetz et al. 2006). Since then, the SDSS Baryon Oscillation Spectroscopic Survey (BOSS) has gathered medium resolution spectra for another ∼250,000 stars (Abolfathi et al. 2018), and the Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST) has collected spectra for ∼1 million stars (Cui et al. 2012; Zhang et al. 2019). In addition, high resolution spectroscopic surveys have begun to provide precise radial velocities, stellar parameters, and exciting results in chemical abundances for over 400,000 stars, e.g., SDSS APOGEE (Holtzman et al. 2018; Zasowski et al. 2019), and GALAH (Buder et al. 2018). Deeper optical high resolution spectroscopic surveys will soon begin at the 4-metre telescopes, including INT/WEAVE (Dalton et al. 2018) and ESO/4MOST (de Jong et al. 2019), and at the 8-metre telescopes, e.g., Subaru/PFS (Tamura et al. 2018).

To prepare for this era of large data sets, methods to consistently and efficiently analyse stellar spectra are being

∗ E-mail: sbialek@uvic.ca. The code used in this analysis is available at https://github.com/Spiffical/StarNet

© 2019 The Authors
explored, particularly with sophisticated data analysis algorithms, e.g., “The Cannon” (Ness et al. 2015; Buder et al. 2018; Zasowski et al. 2019), “The Payne” (Ting et al. 2019; Xiang et al. 2019), and “Matisse” (Recio-Blanco et al. 2006; Kordopatis et al. 2013). We have also been exploring the application of “StarNet”, a convolutional neural network (Fabbro et al. 2018). StarNet was found to reproduce the stellar parameters of benchmark stars at least as well as traditional methods, and it could predict the stellar parameters for the entire APOGEE spectral data set within minutes. Furthermore, StarNet was the first application that could be trained either from data with a priori known stellar labels (data-driven mode) or from a synthetic spectral grid (synthetic mode). Leung & Bovy (2019) improved on the data-driven StarNet implementation by modifying the neural network architecture to track individual abundances, the capability to train on missing or noisy stellar labels, and to estimate prediction uncertainties.

Machine learning methods have now been shown to exceed the performance of traditional methods for spectroscopic analysis, both in terms of time and quality. Machine learning applications are highly versatile, and are an active line of research well beyond astronomical applications, providing a symbiosis where astronomical datasets can both help validate new techniques and also benefit from new analysis methods, e.g., new and clever techniques are being developed to examine the propagation of errors within neural networks (Lakshminarayanan et al. 2017) and generative methods can be used to identify missing physics (O’Brien et al., in prep.).

In this paper, we examine the impacts of training StarNet with a variety of publicly available high resolution, optical synthetic stellar grids. These include INTRIGOSS (Franchini et al. 2018), AMBRE (de Laverny et al. 2012), PHOENIX (Husser et al. 2013), and FERRE (Allende Prieto et al. 2018). These grids of synthetic spectra have been generated using independent model atmospheres and radiative transfer codes (all 1D and in LTE), with a range of atomic and molecular opacities required to describe the stellar photosphere. We also considered exploring other available synthetic grids, but found the wavelength coverage too small (e.g., non-LTE grids from M. Kovalev and M. Bergemann, private communications) or that the stellar parameter range was too small (e.g., optical regions of the APOGEE ASSET grid, by S. Mészáros, private communications).

We describe our continuum normalization scheme and the upgrades to StarNet in Section 2, including a new deep ensembling method that provides estimates of uncertainties in the stellar labels. In Section 3, a description is provided for the data preparation and augmentation of the synthetic grids for training StarNet, which is then used to assess the synthetic gaps. In Section 4, we address the sources of biases in our methods, and provide a validation of StarNet’s predicted uncertainties. In Section 5, FLAMES-UVES spectra from the Gaia-ESO Survey provide a test for StarNet’s performance on observational spectra when trained on the INTRIGOSS grid.

2 METHODS

2.1 Analysis with neural networks

Only a brief description of neural networks is provided here in order to establish the terminology used throughout this paper. For a more complete description of StarNet and our machine learning methodology used, see Fabbro et al. (2018).

Fundamentally, a neural network (NN) is a function which transforms an input to a desired output. The function is composed of many parameters, arranged in layers, which form a highly non-linear combination of the input features, allowing for complex mappings to be represented accurately. StarNet is a convolutional NN, in which a series of learned filters, followed by a series of learned inter-connected nodes, transform a stellar spectrum to a prediction of associated stellar parameters.

To ensure the NN does not over- or under-fit the data, typically the full data set is split into a training, validation, and test set. The training set is used to directly influence the parameters of the NN, and the validation set is used to periodically check the performance of the NN on a separate data set. Both of these sets are utilized during the training of the NN, in which data is iteratively sent through the NN, the parameters of the NN are nudged in a direction which minimizes the output of the loss function (for regression problems, the loss is typically the residual between the prediction and expected output), and in this study, the training is stopped when performance on the validation set ceases to improve. Since both the training and validation sets influence the final trained NN, the test set is used to quantify the final performance for an independent data set.

For a training set of 90,000 spectra, each with ~40,000 flux values, the training time for StarNet rarely exceeds three hours using a single Tesla V100 GPU. With a final trained model, predictions for a set of thousands of spectra can take a matter of seconds.

2.2 Modifications to StarNet

2.2.1 Uncertainty Predictions

To derive predictive uncertainties we have adapted the method of deep ensembling, in which an ensemble of StarNet NNs with different initialization are trained, as outlined in Lakshminarayanan et al. (2017). Each NN predicts the mean and variance which, after averaging, is associated to the predictive uncertainty of each stellar parameter. This simple scheme has been shown to have good coverage in a variety of applications (Ovadia et al. 2019) and it is easy to implement, as only two modifications to an existing NN are required:

(i) Instead of the mean squared error being used as a loss function, a proper scoring rule which includes the variance, $\sigma^2_y$, is used. In this case, the negative log-likelihood criterion is minimized:

$$-\log p(y|x, \theta) = \frac{\log \sigma^2_y(x)}{2} + \frac{(y - \mu_\theta(x))^2}{2\sigma^2_y(x)}$$

where $x$ and $y$ are respectively the inputs and targets, and $\mu_\theta(x)$ is the predicted mean (note that this is the
mean of one model’s prediction, since we are treating the target values as samples from a Gaussian distribution)

(ii) The last layer of the NN is changed such that – in addition to its regular linear output, \( \mu_\theta(x) \), needed for a regression problem – it outputs another linear value, \( \sigma_\theta(x) \), needed for determining the variance of its predictions.

Once the ensemble of NNs is trained, the final prediction, \( \mu^*(x) \), and final variance, \( \sigma^2(x) \), can be obtained by combining the outputs from each model as you would for a mixture of uniformly-weighted Gaussian distributions. Explicitly, \( \mu^*(x) \) is given by the average of the predicted means of each NN, and the final variance is determined via the following equation:

\[
\sigma^2(x) = M^{-1} \sum_{m=1}^{M} (\sigma^2_{\theta_m} + \mu^2_{\theta_m}(x)) - \mu^2(x) \tag{2}
\]

where \( M \) is the number of NNs used in the ensemble, typically 5-7.

The method of deep ensembling is a powerful upgrade to the StarNet architecture for its ability to quantify how closely the spectra in a test set resemble the spectra used to train the model. The uncertainty not only covers the finite sample training size, but also some of the out-of-distribution uncertainties. In contrast with the Monte-Carlo dropout method for uncertainty predictions, it does not perturb the network architecture as much (Ovadia et al. 2019). Furthermore, since each model can be trained in parallel, an ensemble of networks takes no longer to train than one model.

2.3 Augmenting and pre-processing the data

Synthetic and observed spectra typically have vastly different shapes due to instrumental effects and other signatures that uniquely affect the observed spectra. Special care is required to ensure both sets of spectra are standardized to minimize this synthetic gap. There are several steps involved in this process, including both pre-processing the spectra (matching the resolution of the spectra, re-sampling the spectra to a common wavelength grid, and removing the continuum) and augmenting the spectra (adding noise, effects of rotational and radial velocity, and zeroing flux values to mimic bad pixels). Augmenting data is a popular method used in machine learning experiments, serving a dual purpose of increasing both the robustness of the NN to variations existing in reality (which are not necessarily represented in a vanilla training set) and the size of a training dataset; spectral grids usually contain several thousand templates, however typically more data is required for training a deep NN that can make accurate predictions.

With all of this in mind, the synthetic spectra used for training StarNet were adapted for application to VLT/UVES spectra, by having the following modifications applied (in order):

(i) **Resolution matching**: spectra were convolved to a resolution of \( R \approx 47,000 \), the resolution of the UVES spectra

(ii) **Rotational velocity**: randomly chosen with the constraint \( 0 < v_{\text{rot}} < 70 \text{ km/s} \)

(iii) **Radial velocity**: randomly chosen with the constraint \( |v_{\text{rad}}| < 200 \text{ km/s} \)

(iv) **Sampling matching**: the wavelength grid was re-sampled onto the UVES wavelength grid

(v) **Gaussian noise**: randomly chosen with the constraint \( \sigma < 7\% \) flux value, corresponding to S/N > 14

(vi) **Continuum removal**: using the method described in Section 2.3.1

(vii) **Zeroing flux values**: a maximum of 10% of a synthetic spectrum is randomly given a flux value of zero

(viii) **Masking tellurics**: all telluric lines\(^1\) are given a value of zero.

---

\(^1\) Telluric lines from the Keck-MAKEE pipeline, available online at https://tinyurl.com/y45f8px

**Figure 1.** The results of our continuum fitting procedure for a random sample of INTRIGOSS synthetic spectra (left column) and FLAMES-UVES spectra (right column). The red line indicates the estimated continuum, and for the INTRIGOSS spectra the blue dashed line indicates the true continuum. The complex cyclical shape of the FLAMES-UVES spectra eludes simple fits of polynomials.
Section 2.3.1 Continuum removal

Special attention is required for good estimates of the stellar continuum in a spectroscopic analysis. Any method used for estimating the continuum should be invariant to both the shape and the signal-to-noise (S/N) of the spectrum to prevent the introduction of noise-dependent biases into the parameter estimations.

Several methods involve polynomial fits, with some groups selecting high order polynomial fits to the entire spectrum, and others fitting a lower order polynomial to a set of identified ‘continuum pixels’ (Casey et al. 2016). Other popular methods involve splitting the spectrum into short segments of equal length and estimating the continuum of each segment (e.g., García Pérez et al. 2016; Ness et al. 2015). The segment methods perform well in cases where the spectral shape varies significantly over the wavelength range, possibly due to different detectors.

In this paper, a method based on segmenting the spectra was adopted: with each segment of 10 Angstroms, the known strong absorption features are masked, then iteratively the median is found and flux points are rejected above and below when discrepant by 2 and 0.5 standard deviations, respectively, until convergence is achieved. This ‘asymmetric sigma clipping’ more aggressively rejects absorption features in order to find the true continuum. Once the continuum has been estimated in each segment, a cubic spline is fit to the segments. Figure 1 shows the ability of this method to fit both the complex shape of VLT/UVES spectra and the synthetic INTRIGOSS spectra.

A known caveat with the asymmetric sigma clipping method is its noise dependent bias: as the noise levels increase in a spectrum, the found continuum is pushed further towards the ‘noise ceiling’, and thus the estimated continuum is above the true continuum. Figure 2 shows this bias as a function of temperature. It can be seen that in all cases the estimated continuum for a set of synthetic spectra, where the true continuum is known a priori, is higher for a noisy spectrum. Also shown is the trend of spectra with lower temperatures to have a continuum estimate well below the true continuum. This is expected since the majority of a low temperature spectrum lies below the continuum (due to extensive line blanketing), but this is not a problem here since this trend exists in both the synthetic and observed spectra.

Section 4.1 shows how this noise-dependent bias is minimized by simply adding noise at training time, forcing the network to learn the bias correction.

Other continuum estimation techniques were experimented with, e.g. Gaussian smoothing normalization (Ho et al. 2017), but they were found to affect the synthetic spectra differently than the observed spectra and led to more discrepant results.

3 SYNTHETIC SPECTRAL GRIDS

There are numerous grids of synthetic spectra available online (for a summary, see Martins & Coelho 2017), each differing in their spectral parameter and wavelength samplings, and generated from different radiative transfer codes, atomic and molecular line lists, model stellar atmospheres, and comparisons or corrections to observed spectra. These differences have significant impacts on the synthetic spectra, making comparisons between grids inconsistent. With each new grid produced, the quality of the synthetic spectra increases by focusing on the atomic data in the line lists (e.g., see Kurucz 2011), which already include information for many millions of spectral features. To train a machine learning algorithm, it is necessary to carefully consider which grid of synthetic spectra is best to use in a particular spectroscopic analysis.

3.1 The synthetic grids used in this study

The synthetic spectra used in this analysis include the high spectral resolution grids INTRIGOSS, AMBRE, FERRE, and PHOENIX. When StarNet is trained and tested on these grids, they are pre-processed and augmented according to Section 2.3, unless otherwise noted.

The parameter space covered by the grids is summarized in Table 1, and a brief description of each grid follows:

(i) INTRIGOSS: created by Franchini et al. (2018), this grid is a set of high resolution synthetic spectra specifically created for the analysis of F, G, and K type stars in the Gaia-ESO survey. The synthetic spectra were tuned by direct comparison to Gaia-ESO spectra, and in some cases the line list was modified to better match absorption features in the observed spectra without identifying which atom or molecule was the source of the feature. The INTRIGOSS spectra allow the stellar parameters \( T_{\text{eff}} \), \( \log g \), [Fe/H], \( [\alpha/\text{M}] \), and \( v_{\text{micro}} \) to vary within relatively small ranges (see Table 1) and span the wavelength range 483-540 nm only. Although this wavelength range is only a subset of the entire wavelength range of the...
Machine learning for optical stellar spectra

Figure 3. The differences in synthetic spectra when compared to INTRIGOSS, as a function of the three main stellar parameters. For each INTRIGOSS spectrum, spectra with matching parameters from the PHOENIX, AMBRE, and FERRE grids were collected, and the percentage difference between the spectra was calculated. Finally, the average difference across all matched spectra in bins of temperature, surface gravity, and metallicity were determined.

Table 1. The parameter space covered by and sampling of the synthetic spectra grids used in this study.

| T_eff (K) | log g (dex) | [Fe/H] (dex) | [α/M] (dex) | v_micro (km/s) |
|-----------|------------|--------------|-------------|---------------|
| Min. | Max. | Step | Min. | Max. | Step | Min. | Max. | Step | Min. | Max. | Step |
| INTRIGOSS | 3750 | 7000 | 250 | 0.5 | 5.0 | 0.5 | -1.0 | 0.5 | 0.25 | -0.25 | 0.5 | 0.25 | 1 | 2 | 1 |
| FERRE | 3500 | 6000 | 500 | 0 | 5.0 | 1 | -5.0 | 0.5 | 0.5 | - | - | - | 1.5 | 1.5 | - |
| AMBRE | 2500 | 8000 | 250 | -0.5 | 5.5 | 0.5 | -5.0 | 1.0 | 0.25 | -0.4 | 0.4 | 0.2 | 1 | 2 | 1 |
| PHOENIX | 2300 | 7000 | 100 | 0 | 6.0 | 0.5 | -4.0 | -2.0 | 1.0 | - | - | - | 1.2 | 0.2 | 0 | 4 | f(T_eff) |
| 7000 | 15000 | 200 | 0 | 6.0 | 0.5 | -4.0 | -2.0 | 1.0 | - | - | - | 0 | 4 | f(T_eff) |

UVES spectra (480-680 nm, in three settings), it contains important features such as Hβ, the Mgb lines, and numerous metal lines.

(ii) FERRE: this newer grid represents a huge wavelength (120-6500 nm) and parameter range (3500 ≥ T_eff ≥ 30,000 K, 0 ≤ log g ≤ 5, -5 ≤ [Fe/H] ≤ 1), using the newest sources of atomic and molecular data from the literature to model B to early-M type stars at varying resolutions (R ∼ 10,000, 100,000, 300,000). Although not specifically tuned to spectra from any particular survey, the spectra do reproduce the main absorption features when compared to HST UV-optical and APOGEE IR spectra (Allende Prieto et al. 2018). FERRE appears to be the largest general purpose grid of synthetic spectra created to date, though the FERRE authors caution that the grid is, in some ways, already outdated. The full FERRE grid is split into 5 sub-grids with increasing ranges of temperature, and only the first two are used in this study (see Table 1).

(iii) AMBRE: a high resolution (R > 150,000) grid of optical spectra (300-1200 nm) modeling F, G, K, and M type stars, with 4 stellar parameters over a relatively large extent (2500 ≥ T_eff ≥ 8000 K, -0.5 ≥ log g ≥ 5.5, -5 ≥ [M/H] ≥ 1, -0.4 ≥ [α/M] ≥ 0.4). Although it was created several years ago (de Laverny et al. 2012), and thus uses outdated atomic data, it has been used recently, for example, in accurately predicting stellar parameters for Gaia-ESO UVES spectra (Worley et al. 2016).

(iv) PHOENIX: this grid was created as a resource for very high resolution (R > 100,000) stellar spectra spanning ultra-violet to infrared wavelengths (50-5000 nm); Husser et al. (2013) use it to analyse MUSE integral field spectra of stars in the metal-poor globular cluster NGC 6397. It spans a large parameter space (2300 ≥ T_eff ≥ 12,000 K, 0 ≥ log g ≥ 6, -4 ≥ [M/H] ≥ 1, -0.2 ≥ [α/M] ≥ 1.2). It has also been used recently for machine learning applications, e.g., of LAMOST data (Wang et al. 2019).

Since the INTRIGOSS grid was created specifically for the Gaia-ESO survey and includes a carefully crafted line list and comparisons to both UVES spectra and other synthetic grids, it was chosen as the baseline for our exploration of the impact of the various synthetic grids, and as the primary grid for our analyses of the FLAMES-UVES spectra.

3.2 Comparisons of synthetic grids

To perform a comparison of the synthetic spectral grids, INTRIGOSS was chosen as the baseline. For each INTRIGOSS spectrum, spectra with matching stellar parameters from each grid were selected (if none were found, the INTRIGOSS spectrum was skipped), and the residual of the flux values of each spectrum with respect to the INTRIGOSS spectrum
was calculated and converted to a percentage difference. The average percentage difference was then determined in bins of temperature, surface gravity, and metallicity. As shown in Figure 3, the differences in the spectra are more pronounced at lower temperatures and higher metallicities, i.e., in the grid regions that would be the most sensitive to line blanketing. The FERRE spectra are the most closely matched to the INTRIGOSS spectra, over the widest range in stellar parameters, whereas the PHOENIX spectra are the most dissimilar.

To qualitatively assess how closely the synthetic spectral grids match the Gaia-ESO FLAMES-UVES spectra (discussed further in Section 5), a t-SNE test was carried out to compare the closest matching spectra from each grid to each UVES spectrum. As seen in Figure 4, there is a distinct difference between the raw observed and synthetic spectra; the synthetic gap. However, when the data is augmented as described in Section 2.3 then the synthetic gap is significantly narrowed: the augmented synthetic spectra occupy the same compressed low-dimensional space as the observed UVES spectra.

4 TRAINING STARNET WITH INTRIGOSS

For our first application, StarNet has been trained using the augmented INTRIGOSS spectra, and is referred to as "StarNet-INTRIGOSS". The grid of 7,616 INTRIGOSS spectra were split into a reference set (6,093 spectra) and a test set (1,523 spectra), an 80/20 split. These two datasets were then pre-processed and augmented (as described in Section 2.3) to create datasets several times their size: the 6,093 reference spectra were turned into an augmented reference set of 100,000 spectra (no further improvements were seen with a larger training sample) and the 1,523 test spectra were turned into an augmented test set of 10,000 spectra.

The augmented reference set was then split into a training set (90,000 spectra) and a validation set (10,000 spectra), a 90/10 split. These steps help to mitigate over-fitting during training (further discussed below).

4.1 Addressing method-dependent biases: testing with INTRIGOSS spectra

The performance of StarNet-INTRIGOSS is assessed here using the INTRIGOSS synthetic spectra themselves to first explore the limitations and systematic biases inherent in the method. This is because we know the spectral properties (stellar parameters and continuum) a priori, and we can investigate and mitigate errors or degeneracies before predicting on real spectra. In addition, we want to ensure StarNet does not over-fit to the training data, which would result in both poor interpolation between the synthetic grid points and poor predictions of observed spectra. Both of these issues are discussed below.

4.1.1 Noise-dependent biases in continuum fitting

As discussed in Section 2.3.1, the asymmetric sigma-clipping continuum removal method has a known noise-dependent bias. Figure 2 illustrates this, where the estimated continuum for low S/N spectra can be discrepant by several percent above the true continuum (with an exception at lower temperatures where the stronger absorption features cause much of the spectrum to lie below the continuum). If the estimated continuum is significantly higher than the true continuum, the resulting continuum-normalized spectra will contain artificially lowered flux values. This would lead to...
Figure 5. Residual plots to show noise-dependent biases from the asymmetric sigma clipping continuum removal in the stellar parameter estimations. Two versions of StarNet were trained: one model, StarNet-INTRIGOSS (orange), was trained on 90,000 INTRIGOSS spectra augmented as outlined in Section 2.3, and the other, StarNet-INTRIGOSS noiseless (purple), was trained identically except without the addition of noise to the synthetic spectra prior to continuum removal. Each was tested on 10,000 noisy INTRIGOSS spectra, the median residual at each grid point was calculated, and the results for all spectra with S/N < 80 are shown here. The discrepancies are the most pronounced at lower metallicities, higher surface gravities, and across all rotational velocities.

Deeper absorption features which could mimic a lower temperature or higher metallicity than the true value.

4.1.2 Testing for over and under-fitting with intra-grid synthetic spectra

Along with the published INTRIGOSS grid of spectra, a set of 50 spectra at intra-grid locations was provided by the INTRIGOSS team for testing the ability of a chosen methodology to interpolate between grid points. These intra-grid spectra also provide an excellent test set to confirm that the model for StarNet-INTRIGOSS did not over-fit to the training set nor result in other systematic biases in its predictions.

The predictions from StarNet-INTRIGOSS on the 50 intra-grid spectra are shown in Figure 6. The results are excellent, with no signs of under or over-fitting from the training set. The slight offset of temperature is unexpected, but we note that it is very small, ranging from 1-2 σ (the uncertainty propagated by the NN itself). We also have no information on how the intra-grid spectra were selected and generated, and therefore do not consider this result to be significant. We also note that the intra-grid spectra do not extend below $T_{\text{eff}} = 4500$ K or $\log g < 2.4$, so we cannot confidently evaluate our stellar parameter predictions in those ranges.

Interestingly, the predictions for the radial velocity, $v_{\text{rad}}$, are excellent and do not show significant bias, and have typical uncertainties below 0.5 km s$^{-1}$. This is somewhat surprising, given that convolutional NNs with pooling layers are built to be invariant to small translations.
4.2 Testing StarNet-INTRIGOSS with other synthetic spectral grids

To explore the accuracies and uncertainty estimates from the deep ensembling method, the predictions of StarNet-INTRIGOSS are compared between the INTRIGOSS, FERRE, AMBRE, and PHOENIX grids. These grids have been previously examined by Franchini et al. (2018) in their comparison of seven synthetic grids (see their Figure 7), and in our percentage difference analysis and t-SNE comparisons in Section 3.2 (Figs. 3 and 4). Both analyses show that FERRE is the most similar to INTRIGOSS, while PHOENIX is the least similar.

The validity of the deep ensembling method can be further verified by examining the predictions from within the parameter space used for training, and also beyond those boundaries. As a first test, StarNet-INTRIGOSS is used to predict stellar parameters for test sets of 3,000 augmented INTRIGOSS, AMBRE, FERRE, and PHOENIX spectra which span the same parameter space: the uncertainties are summarized in Figure 7. The uncertainties increase relative to the predictions from the INTRIGOSS spectra at lower temperatures, lower surface gravities, and higher metallicities, i.e., where the synthetic grids were previously shown to deviate the most (see Figure 3). Similarly, the uncertainties in the predictions from the PHOENIX grid are the largest, consistent with the known larger differences between the INTRIGOSS and PHOENIX spectra.

To test the uncertainties in the predictions in a parameter space beyond the training data set, StarNet-INTRIGOSS was applied to spectra from the full parameter ranges in the AMBRE, FERRE, and PHOENIX grids. Each extend to higher and lower temperatures, and much lower metallicities; the results are shown in Figure 8. As expected, the uncertainties tend to increase when predicting outside of the parameter ranges used for training, as well as when the predictions become more discrepant from their true values.

5 AN APPLICATION TO GAIA-ESO FLAMES-UVES SPECTRA

The Gaia-ESO public spectroscopic survey (GES, Gilmore et al. 2012) is a large survey with the goal of exploring all components of the MW in a complementary way to Gaia. Along with the observed spectral database, an official Gaia-ESO Survey Internal Data Release (GES iDR) is available, containing stellar parameters derived as the weighted average of the results from a set of working groups (each using different methods). The fourth data release (GES iDR4) is used in this study as a comparison for our StarNet predictions (Pancino et al. 2017).

The GES is carried out using FLAMES at the VLT (Pasquini et al. 2002) to obtain high-quality medium-resolution Giraffe spectra for $10^5$ stars and high-resolution UVES spectra for $\sim5000$ stars. Currently, a dataset of 2308 FLAMES-UVES spectra is available, spanning field and cluster stars from the bulge, halo, thick disc and thin disc. The S/N distribution of these stars is shown in Figure 9, where the majority of the stars have S/N < 100.

In addition, the Gaia-ESO survey includes a set of 34 benchmark spectra of well-known bright stars (Blanco-
Figure 7. The uncertainties in the predictions of StarNet-INTRIGOSS for the three main stellar parameters. The test sets are augmented INTRIGOSS, AMBRE, FERRE, and PHOENIX spectra (limited to the INTRIGOSS parameter range), and the median uncertainty in bins of temperature, surface gravity, and metallicity, were calculated. In general, the uncertainties grow w.r.t INTRIGOSS based on how dissimilar the spectra are (see Figure 3 for these trends), especially pronounced at lower temperatures, lower surface gravities, and higher metallicities.

Figure 8. The uncertainties in the predictions of StarNet-INTRIGOSS for the three main stellar parameters. The test sets are augmented AMBRE, FERRE, and PHOENIX spectra (spanning their entire parameter ranges). The first row shows the uncertainties as a function of the specified parameter, whereas the second row shows the uncertainties as a function of the residual between StarNet-INTRIGOSS predictions and truth values of the specified parameter. The grey dashed lines correspond to the limits of the INTRIGOSS grid. As expected, the uncertainties grow both when StarNet predicts outside the ranges of the INTRIGOSS spectra it was trained on, and as the residuals increase.

Cuaresma et al. 2014), available online\(^3\), to be used as a reference. Their parameters $T_{\text{eff}}$ and $\log g$ were determined independent of spectroscopy, using angular diameter measurements and bolometric fluxes (Heiter et al. 2015), and $[\text{Fe/H}]$ was determined from these values (Jofré et al. 2014).

The Gaia-ESO survey has also observed several calibration clusters, including the globular clusters M 15, NGC 104, NGC 1851, NGC 2808, NGC 4372, NGC 4833, NGC 5927, and NGC 6752, and the open clusters M 67, NGC 3532, and NGC 6705. Some of these clusters have metallicities much lower than the INTRIGOSS metallicity grid ($[\text{Fe/H}]$...


\( \text{iii) Metal-poor (MP): } [\text{Fe/H}] > -1 \), so they were removed from this analysis. This leaves five clusters for testing StarNet-INTRIGOSS, including NGC 104, NGC 2808, NGC 3532, NGC 5927, and M 67.

As a first test, we will examine the abilities of StarNet-INTRIGOSS to predict stellar parameters for the GES benchmark stars. This will be followed by testing its predictions for stars in the calibration clusters. Finally, we test the predictions made on the entire sample of FLAMES-UVES spectra in the Gaia-ESO survey.

5.1 StarNet-INTRIGOSS predictions for the GES benchmark stars

Following the procedure in Smiljanic et al. (2014), the benchmark stars were separated into three groups in order to assess the accuracy in different regions of parameter space:

(i) Metal-rich dwarf (MRD): \([\text{Fe/H}] > -1.00\) and \(\log g > 3.5\)

(ii) Metal-rich giant (MRG): \([\text{Fe/H}] > -1.00\) and \(\log g \leq 3.5\)

(iii) Metal-poor (MP): \([\text{Fe/H}] \leq -1.00\)

Shown in Figure 10 are the results of StarNet-INTRIGOSS predictions on seven MRDs, three MRGs, and four MP stars from the set of benchmarks. The metric for evaluating performance, as in Smiljanic et al. (2014), is the average quadratic difference, \(\Delta\), between the predictions and benchmark values, and is small for all three groups of stars (\(\Delta T_{\text{eff}} < 120 \text{ K}, \Delta \log g < 0.47, \) and \(\Delta [\text{Fe/H}] < 0.05\)). Additionally, there exists no significant deviation for any parameters, with the exception of larger results for both \(\log g\) and \(T_{\text{eff}}\) for the MRGs, and an increasing trend at lower \([\text{Fe/H}]\) for the MP stars. We note that the MP stars lay outside the metallicity range of the spectra used for training, so this is not surprising. The benchmark uncertainties for \(v_{\text{sin}}\) are so large that it is difficult to determine if StarNet-INTRIGOSS produced accurate predictions. It is also interesting that in most cases when the published uncertainties for the benchmark parameters are large, so too are the predicted uncertainties of our deep ensembling method.

Altogether the results obtained through tests on the Gaia benchmark stars provide a convincing validation that our method works well across the range of parameters for high \(S/N\) spectra. However, we caution that these comparisons are against a statistically small sample (the benchmark stars) and that systematic errors could potentially appear in larger samples.

5.2 StarNet-INTRIGOSS predictions for the GES calibration clusters

StarNet can predict stellar parameters for FLAMES-UVES spectra, but are those predictions physically realistic? A common method used to assess the fidelity of astrophysical parameters is to compare them to a theoretical understanding of stellar evolution.

The predictions of \(T_{\text{eff}}\) and \(\log g\) of the stars in each cluster from StarNet-INTRIGOSS were compared to the MESA Isochrones and Stellar Tracks (MIST, Choi et al. (2016)), generated by adopting the published metallicities and ages for each cluster from the Harris catalogue (Harris 2010); see Fig. 11. While StarNet appears to predict both higher surface gravities and temperatures for giants than the GES iDR4 values, they remain physically consistent when compared to the isochrones, and are more constrained. It is important to keep in mind that the predictions from StarNet are uncalibrated and that StarNet recovers the stellar parameters \(T_{\text{eff}}, \log g,\) and \([\text{Fe/H}]\) for both dwarfs and giants in a physically consistent manner.

As a further check to ensure physically consistent stellar parameters, the metallicity predictions of StarNet-INTRIGOSS were compared directly to the literature values for each cluster. Figure 12 shows the average StarNet \([\text{Fe/H}]\) predictions for the stars in each calibrating cluster, with error bars derived from the standard deviation of the predictions. In the parameter space that INTRIGOSS was trained on (shown by the vertical dashed lines), the metallicity predictions show excellent agreement with cluster values, and even NGC 2808, which is just outside the trained parameter range, is well predicted.

5.3 StarNet-INTRIGOSS predictions for the entire Gaia-ESO Survey (GES iDR4)

The entire sample of FLAMES-UVES spectra was examined with StarNet-INTRIGOSS, with a few cuts made to produce the final sample for predictions: stars were removed if they had NaN values for any parameter in the GES iDR4 catalog and if the uncertainties produced by StarNet for any parameter were abnormally large (\(\sigma T_{\text{eff}} > 65 \text{ K}, \sigma [\text{Fe/H}] > 0.50, \sigma \log g > 0.80, \sigma v_{\text{rot}} > 3 \text{ km/s}, \sigma v_{\text{rad}} > 5 \text{ km/s}\)), decreasing the sample size from 2308 to 2200. The \(T_{\text{eff}}-\log g\) plots for the final sample are shown in Figure 13, where the left panel shows the predictions made by StarNet-INTRIGOSS, and the right panel shows the GES iDR4 catalog values. MIST isochrones for age = 8 Gyr and varying metallicities are overlaid for clarity.

In general StarNet-INTRIGOSS finds slightly larger values for both \(T_{\text{eff}}\) and \(\log g\), especially for giants, as seen for the benchmarks in Figure 10. However, we note that our results are from only a narrow window of the spectrum (483-540 nm), whereas the full GES iDR4 analyses are from the full UVES spectral region (480-680 nm). We also note that
Machine learning for optical stellar spectra

Figure 10. StarNet-INTRIGOSS was used to predict stellar parameters for the Gaia-ESO benchmark stars, and the residuals between predictions and published values are shown here. The stars were split into metal-poor (MP) stars, metal-rich giants (MRGs) and metal-rich dwarfs (MRDs), following the procedure in R. Smiljanic et al. (2014). The average quadratic difference, $\Delta$, between StarNet’s predictions and benchmark values is used to evaluate the accuracy of the predictions.

Figure 11. StarNet-INTRIGOSS predictions of log\(g\) and \(T_{\text{eff}}\) compared with theoretical MIST isochrones with the ages and metallicities shown in light grey text. The cluster metallicities and ages were retrieved from the online updated catalog of Harris (2010) and the WEBDA database. Also plotted are the GES iDR4 stellar parameters for the same stars (except NGC5927 and M67 for which none could be found).

The working groups who contributed to the GES iDR4 were using a different set of synthetic spectra than INTRIGOSS.

Final predictions for the stellar parameters $T_{\text{eff}}$, log\(g\), [Fe/H], and $v_{\text{rad}}$ for the FLAMES-UVES spectra are compared with the GES iDR4 values in Figure 14. The difference in the predictions for $T_{\text{eff}}$ and log\(g\) both show a slight decrease with increasing values when compared with the GES iDR4 results. The predictions for [Fe/H] and $v_{\text{rad}}$ are in excellent agreement. To the best of our knowledge, this is the first time that a NN has been able to accurately predict radial velocities on real spectra.

Finally, the uncertainties on the stellar parameter predictions from StarNet-INTRIGOSS for the full sample of FLAMES-UVES spectra are shown in Figure 15. In the $T_{\text{eff}}$ and log\(g\) plots there are two distinct populations of stars corresponding to dwarfs and giants, where main sequence stars tend to dominate the sample, and where the uncertainties are moderately larger for the giants. The [Fe/H] and [$\alpha$/Fe] uncertainties are very small, suggesting that the INTRIGOSS spectra model the absorption features quite well. Finally, the $v_{\text{rad}}$ uncertainties are comparable to the typical error in the GES iDR4 (0.4 km s$^{-1}$).

6 DISCUSSION

6.1 Exploring StarNet trained on other synthetic grids

In Fig. 8, the differences between the four synthetic grids in this paper were compared according to the predictive StarNet-INTRIGOSS values and uncertainties. We showed
that the differences between AMBRE and FERRE, with respect to INTRIGOSS, were relatively small (both in terms of the uncertainties and the residual of predictions), while the differences between PHOENIX and INTRIGOSS were relatively large. Therefore, if StarNet is trained on the AMBRE or FERRE grids, one might expect similar results when predicting stellar parameters from the FLAMES-UVES spectra; however, more discrepant results may be expected when trained on the PHOENIX spectra.

To test this assumption, the same StarNet architecture was separately trained on 90,000 augmented AMBRE, FERRE, and PHOENIX spectra (spanning the same parameter space as INTRIGOSS). These will be referred to as StarNet-AMBRE, StarNet-FERRE, and StarNet-PHOENIX. The results for the predictions on the GES benchmark stars for each trained StarNet model are summarized in Table 6.1. Overall, training with INTRIGOSS spectra gives better results than training with the other grids, while training with PHOENIX spectra give the most discrepant results. StarNet-INTRIGOSS tests especially well on metallicities that lie in the range of 0.1 - 0.4, provided that the errors are within the uncertainties and residual of predictions, while the FERRE and AMBRE trained StarNet predictions are marginally more precise than those when trained with the PHOENIX grid. Overall, and for general purposes, we recommend StarNet trained with the FERRE grid, StarNet-FERRE, combining both good precision and large parameter extent. Applications for StarNet-FERRE can include the analysis of optical spectral archives, such as for CFHT ESPaDOnS (Donati et al. 2006), and Gemini GRACES (Chene et al. 2014), for precision stellar parameters. The flexibility of StarNet-FERRE also means that it can be trained for lower resolution spectral archives as well, e.g., the SDSS BOSS database (Dawson et al. 2016) or ESO Xshooter library (Vernet et al. 2011). Unfortunately, the current StarNet pipeline requires retraining for each new observational data set and/or for each new synthetic grid library. In the future, this could be accelerated by using transfer learning techniques, e.g., training a very large NN that would cover most cases and would be tuned to specific data sets or spectral parameters.

6.3 Caveats for ML applications

One of the main advantages of the CNN with deep ensembling method developed in this paper is its adaptability to any spectroscopic survey and any grid of synthetic spectra.

Table 2. StarNet was separately trained on sets of 90,000 augmented spectra from the INTRIGOSS, FERRE, AMBRE, and PHOENIX grids. The results of each trained model when predicting on the Gaia-ESO benchmark stars are shown here.

| MRGs | MPs |
|------|------|
| Mean | Median | Standard Deviation | Min | Max |
| StarNet-INTRIGOSS | 74 | 0.11 | 0.04 | 120 | 0.47 | 0.05 | 23 | 0.19 | 0.05 |
| StarNet-FERRE | 77 | 0.19 | 0.22 | 64 | 0.12 | 0.29 | 63 | 0.17 | 0.19 |
| StarNet-AMBRE | 125 | 0.12 | 0.31 | 100 | 0.17 | 0.36 | 105 | 0.30 | 0.06 |
| StarNet-PHOENIX | 152 | 0.34 | 0.39 | 184 | 0.25 | 0.36 | 79 | 0.11 | 0.17 |

6.2 Recommendations: beyond INTRIGOSS

While training with the INTRIGOSS grid has yielded the most precise results for the GES UVES spectra, the grid is currently limited to significantly smaller temperature, metallicity, and wavelength regimes than FERRE, AMBRE, and PHOENIX. It is inherently less versatile for very metal-poor stars, and for the analysis of observed spectra which do not lay in its very narrow wavelength range. However, the promising results from the INTRIGOSS grid, as shown in this study, do suggest that further work to extend the wavelength coverage of INTRIGOSS is warranted.

For applications outside of the INTRIGOSS parameter and/or wavelength regimes, we have found that StarNet can be trained on any of the other sets of synthetic grids. However, the FERRE and AMBRE trained StarNet predictions are marginally more precise than those when trained with the PHOENIX grid. Overall, and for general purposes, we recommend StarNet trained with the FERRE grid, StarNet-FERRE, combining both good precision and large parameter extent. Applications for StarNet-FERRE can include the analysis of optical spectral archives, such as for CFHT ESPaDOnS (Donati et al. 2006), and Gemini GRACES (Chene et al. 2014), for precision stellar parameters. The flexibility of StarNet-FERRE also means that it can be trained for lower resolution spectral archives as well, e.g., the SDSS BOSS database (Dawson et al. 2016) or ESO Xshooter library (Vernet et al. 2011). Unfortunately, the current StarNet pipeline requires retraining for each new observational data set and/or for each new synthetic grid library. In the future, this could be accelerated by using transfer learning techniques, e.g., training a very large NN that would cover most cases and would be tuned to specific data sets or spectral parameters.
and its ability to predict a consistent set of stellar parameters across surveys, with the same calibration data set. The precision in the method depends on the quality of the synthetic spectra and how closely they match the observed spectra: ideally all synthetic grids would include intra-grid spectra for assessing the interpolation accuracy.

As opposed to training a NN on observed spectra, training on a grid of synthetic spectra has the added benefit of not needing to worry about correlations between stellar parameters being picked up in the training process. For example, when the bulk of a training set of observed spectra has a Mg-Al correlation, then a NN is more likely to falsely assign a Mg-Al correlation to globular cluster stars even if they are known a priori to be anti-correlated (e.g., see the discussion by Leung & Bovy 2019). This problem can be mitigated with domain knowledge, e.g. by windowing the spectra according to spectral features from a particular element, or through an extensive (though potentially prohibitive) array of chemical abundances in the synthetic spectral grids.

There is also the problem of finding rare stars (e.g. carbon-enhanced metal-poor stars, ultra metal-poor stars, stars captured from nearby dwarf satellites, or r-process rich stars, and even spectroscopic binaries; see Venn et al. 2019; Monty et al. 2019; Arentsen et al. 2019; Sakari et al. 2018; Kielty et al. 2017). If a training set does not include a significant proportion of peculiar stars, then predictions on these rare populations will suffer. In machine learning applications, the training set is often the limiting factor, so special care is required to account for out-of-distribution samples. For data-driven methods, this problem is even more difficult to address (tiny sample sizes); however, for synthetic grids, spectra of rare stars can be added a posteriori and the NN re-trained.

In cases where the sample size of a spectroscopic survey is low (in the hundreds or low thousands of spectra), then it might be infeasible to acquire a trained NN which produces accurate results, since the size of the training set could be a limiting factor. This problem is overcome by synthetic spectra: the only limits to the size of a synthetic training set are storage space and the computing time required to produce the spectra.

To extend this analysis to predictions of chemical abundances, spectra could be produced within the parameter range of an existing grid, but not aligned with the grid points (see Ting et al. 2019). Indeed, producing spectra in a grid is quite a rigid and perhaps out-dated strategy as there will inevitably be multiple realizations of the same stellar parameter, resulting in an over abundance of spectra needed for a NN analysis. It is much more economical to produce spectra with randomly varying parameters, especially when considering extending grids to > 10 dimensions.

Training on synthetic spectra allows for a complete model and analysis pipeline to be created before the first light is collected for a spectroscopic survey, meaning as spectra are collected from a telescope their parameters (even radial velocities), along with uncertainties, can be derived in real time. Because our method derives uncertainties, we can also in real-time assess the accuracy of predictions, providing valuable feedback needed to determine how long a star should be observed to achieve a certain level of accuracy.

7 CONCLUSIONS

In this paper, we have presented an updated version of our StarNet convolutional neural network used for the precision analysis of high-resolution stellar spectra. The main update has been the implementation of deep ensembling to estimate realistic uncertainties in the predicted stellar parameters. In addition:

- StarNet has been trained successfully on four independent grids of high-resolution synthetic spectra (INTRIGOSS, FERRE, AMBRE, and PHOENIX), highlighting its versatility.
- Data augmentation is necessary to overcome the synthetic gap, such that different synthetic grids overlap with one another, as well as with observational data from the Gaia-ESO FLAMES-UVES spectroscopic survey.
Figure 14. StarNet was trained on 100,000 augmented INTRIGOSS spectra and tested on 2200 FLAMES-UVES spectra, using parameters from the GES iDR4. In the histogram plots, the dark red and light red lines correspond to distributions of stars with S/N >150 and <100, respectively.

- Data pre-processing included resolution matching (R=47,000), sampling matching (put onto the UVES wavelength grid), and continuum normalization that was consistent between synthetic and observed spectra. The spectra were augmented with a range of rotational and radial velocities, Gaussian noise, and random zero flux values to mimic bad pixels. Finally, regions of known telluric lines were masked in the synthetic and observational data.

- Augmenting the training data with noise before the asymmetric sigma-clipping continuum estimation step was necessary to decrease the biases in predictions.

- Once trained, StarNet was shown to predict stellar parameters for ~2300 FLAMES-UVES optical spectra with high precision compared with traditional methods, and within seconds.

- The precision in StarNet’s predictions for FLAMES-UVES spectra, when compared to Gaia-ESO benchmark stars and calibration clusters, is best when StarNet is trained on the INTRIGOSS grid, as expected since this grid has been specifically tuned for the Gaia-ESO survey in this wavelength region.

- When StarNet is trained with the FERRE synthetic spectral grid, the precision in the results are also excellent (closely followed in precision by training with the AMBRE grid). Due to the limited stellar parameter range and wavelength coverage of INTRIGOSS, we suggest that the FERRE grid is currently the best choice for general
Machine learning for optical stellar spectra

Figure 15. StarNet-INTRIGOS was tested on the Gaia-ESO FLAMES-UVES spectra and shown here are density plots for the uncertainties of StarNet’s predictions.

For the near future, we plan to train StarNet for the analysis of optical spectra from Canadian observational facilities, such as CFHT ESPaDOnS and Gemini GRACES, and to prepare for observational data from the upcoming Gemini GHOST spectrograph. We are also developing new tools for detailed chemical abundances. Our codes are publicly available and simple to adapt to any set of synthetic spectra.

ACKNOWLEDGEMENTS

We thank Jonay González-Hernández, Patrick de Laverny, Alejandra Recio-Blanco, Szabolcs Mészáros, Mikhail Kovalev and Maria Bergemann for many helpful discussions and access to their synthetic spectral grids. We also thank Balaji Lakshminarayanan for helpful feedback in using the deep ensembling method, and Henry Leung for his work and discussions in improving StarNet.

SB and KAV thank the Natural Sciences and Engineering Research Council for funding through the Discovery Grants program and the CREATE program in New Technologies for Canadian Observatories. NK thanks Mitacs for funding through their 2019 Globalink Research Internship program.

REFERENCES

Abolfathi B., Aguado D. S., Aguilar G., Allende Prieto C., Almeida A., Ananna T. T., Anders F. e. a., 2018, ApJS, 235, 42

Allende Prieto C., Koesterke L., Hubeny I., Bautista M. A., Barklem P. S., Nahar S. N., 2018, A&A, 618, A25

Arentsen A., Starkenburg E., Shetrone M. D., Venn K. A., Depagne É., McConnachie A. W., 2019, A&A, 621, A108

Blanco-Cuaresma S., Soubiran C., Jofrè P., Heiter U., 2014, A&A, 566, A98

Buder S., et al., 2018, MNRAS, 478, 4513

Casey A. R., Hogg D. W., Ness M., Rix H.-W., Ho A. Q. Y., Gilmore G., 2016, arXiv e-prints, p. arXiv:1603.03040

Chene A.-N., et al., 2014, in Proc. SPIE. p. 915147 (arXiv:1409.7448), doi:10.1117/12.2057417

Choi J., Dotter A., Cuny C., Cantiello M., Paxton B., Johnson B. D., 2016, The Astrophysical Journal, 823, 102

Cui X.-Q., et al., 2012, Research in Astronomy and Astrophysics, 12, 1197

Dalton G., et al., 2018, Proc. SPIE. p. 107021B, doi:10.1117/12.2312031

Dawson K. S., et al., 2016, AJ, 151, 44

Donati J. F., Catala C., Landstreet J. D., Petit P., 2006, in Casini R., Lites B. W., eds, Astronomical Society of the Pacific Conference Series Vol. 358, Solar Polarization 4. p. 362

Fabbro S., Venn K. A., O’Brien T., Bialek S., Kielty C. L., Jahandar F., Monty S., 2018, MNRAS, 475, 2978

Franchini M., et al., 2018, The Astrophysical Journal, 862, 146

García Pérez A. E., Allende Prieto C., Holtzman J. A., Shetrone M., Mészáros S., Biaxev D. e. a., 2016, AJ, 151, 144

Gilmore G., et al., 2012, The Messenger, 147, 25

Harris W. E., 2010, arXiv preprint arXiv:1012.3224

Heiter U., Jofré P., Gustafsson B., Korn A. J., Soubiran C., Thévenin F., 2015, Astronomy & Astrophysics, 582, A49

Ho A. Y., et al., 2017, The Astrophysical Journal, 836, 5

Holtzman J. A., et al., 2018, AJ, 156, 125

Husser T. O., Wende-von Berg S., Dreizler S., Homeier D., Reiners A., Barman T., Hauschildt P. H., 2013, A&A, 553, A6

Jahandar F., et al., 2017, MNRAS, 470, 4782

Jofré P., et al., 2014, Astronomy & Astrophysics, 564, A133
Kielty C. L., Venn K. A., Loewen N. B., Shetrone M. D., Placco V. M., Jahandar F., Meszaros S., Martell S. L., 2017, MNRAS, 471, 404
Kordopatis G., et al., 2013, AJ, 146, 134
Kurucz R. L., 2011, Canadian Journal of Physics, 89, 417
Lakehminarayanan B., Pritzel A., Blundell C., 2017, in Advances in Neural Information Processing Systems. pp 6402–6413
Lee Y. S., et al., 2011, AJ, 141, 90
Leung H. W., Bovy J., 2019, MNRAS, 483, 3255
Martins L., Coelho P., 2017, Canadian Journal of Physics, 95, 840
Monty S., Venn K. A., Lane J. M. M., Lokhorst D., Yong D., 2019, arXiv e-prints, p. arXiv:1909.11969
Ness M., Hogg D. W., Ho A. Y. Q., Zasowski G., 2015, ApJ, 808, 16
Ovadia Y., et al., 2019, arXiv preprint arXiv:1906.02530
Pancino E., et al., 2017, A&A, 598, A5
Pasquini L., et al., 2002, The Messenger, 110, 1
Recio-Blanco A., Bijaoui A., de Laverny P., 2006, MNRAS, 370, 141
Sakari C. M., et al., 2018, ApJ, 868, 110
Smiljanic R., et al., 2014, Astronomy & astrophysics, 570, A122
Steinmetz M., et al., 2006, AJ, 132, 1645
Tamura N., et al., 2018, in Proc. SPIE. p. 107021C, doi:10.1117/12.2311871
Ting Y.-S., Conroy C., Rix H.-W., Cargile P., 2019, The Astrophysical Journal, 879, 69
Venn K., et al., 2019, arXiv e-prints, p. arXiv:1910.06340
Vernet J., et al., 2011, A&A, 536, A105
Wang R., et al., 2019, PASP, 131, 024505
Worley C., de Laverny P., Recio-Blanco A., Hill V., Bijaoui A., 2016, Astronomy & Astrophysics, 591, A81
Xiang M., et al., 2019, arXiv e-prints, p. arXiv:1908.09727
Yanny B., et al., 2009, AJ, 137, 4377
Zasowski G., et al., 2019, ApJ, 870, 138
Zhang X., Zhao G., Yang C. Q., Wang Q. X., Zuo W. B., 2019, PASP, 131, 094202
de Jong R. S., et al., 2019, The Messenger, 175, 3
de Laverny P., Recio-Blanco A., Worley C. C., Plez B., 2012, A&A, 544, A126

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