Consistent and accurate estimation of stellar parameters from HARPS-N Spectroscopy

Frederik Boe Hüttel
DTU Compute
Technical University of Denmark
Richard Petersens plads 31,
2800 Lyngby, Denmark
fbohy@dtu.dk

Line Katrine Harder Clemmensen
DTU Compute
Technical University of Denmark
Richard Petersens plads 31,
2800 Lyngby, Denmark
lkhc@dtu.dk

Abstract

Consistent and accurate estimation of stellar parameters is of great importance for information retrieval in astrophysical research. The parameters span a wide range from effective temperature to rotational velocity. We propose to estimate the stellar parameters directly from spectral signals coming from the HARPS-N spectrograph pipeline before any spectrum-processing steps are applied to extract the 1D spectrum. We propose an attention-based model to estimate the stellar parameters, which estimate both mean and uncertainty of the stellar parameters through estimation of the parameters of a Gaussian distribution. The estimated distributions create a basis to generate data-driven Gaussian confidence intervals for the estimated stellar parameters. We show that residual networks and attention-based models can estimate the stellar parameters with high accuracy for low Signal-to-noise ratio (SNR) compared to previous methods. With an observation of the Sun from the HARPS-N spectrograph, we show that the models can estimate stellar parameters from real observational data.

1 Introduction

Currently, there exists huge variation in the different techniques used to estimate stellar parameters, ranging from decision tree architectures to tailor-made algorithms made for specific astrophysical surveys [1, 2, 3, 4, 5]. Previous research that has applied artificial neural networks to estimate stellar parameters have focused on *effective temperature* (T_{eff}), *surface gravity* ($\log g$), and *metalicity* (Z) and they apply their models on the extracted 1D stellar spectrum of the CCD spectral image [6, 7, 8]. The method used for the extraction of the 1D spectrum introduces biases and assumption into the spectrum resulting in biased estimation of stellar parameters, which leads to different research groups obtaining different results when observing the same stars [9]. We argue that one should strive to use an end-to-end deep learning approach, which can estimate appropriate pre-processing steps inherent in the modelling. The closest thing we get to an end-to-end approach is using the data from the HARPS-N spectrograph pipeline [10]. In this paper, we propose to use the 2D spectral signal coming from the HARPS-N pipeline to estimate the stellar parameters and we present methods for doing so. The main contributions of our approach are:

- The elimination of spectral pre-processing to extract the 1D spectra, as we apply our deep learning models directly to the 2D signal from the HARPS-N pipeline.
- Improved prediction accuracies for the effective temperature, surface gravity, and metallicity compared to other data-driven approaches as-well-as inclusion of stellar rotational velocity estimation.
- Quantification of uncertainty in estimations of the stellar parameters. The estimated uncertainty provides a basis to create data-driven confidence intervals.

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

- An attention-based model which attends to the underlying elements of an input spectrum.

2 Data

Previous related research on data-driven stellar parameters estimation have generated synthetic spectra to train and estimate the performance of machine learning models [6, 7, 8]. The use of synthetic spectra provides a unique opportunity to generate a large set of labelled data, as the details of the spectra are known a priori and the SNR can be varied to mimic different telescope exposure times [7, 8]. A drawback of this approach is that the trained weights will be biased towards the physical model generating the data, as there exists a *synthetic gap* between the feature distributions of the synthetic and the observed spectra [8]. In this work, we sample the synthetic spectra from a grid of model atmospheres using the ATLAS9 code [11]. The original code is described in detail in [12] and was updated to include new opacity distribution functions as outlined in [13]. The grid was uniformly sampled using 26 T_{eff} ranging from 3500 K to 9750 K with a step-size of 250 K, 11 $\log g$ ranging from 0.0 to 5.0 with a step-size of 0.5, and 7 Z ranging from -2.5 to +0.5 in step-sizes of 0.5. The grid was extended by including 16 different rotational velocities ($V \sin i$) ranging from 0 km/s to 30 km/s with a step-size of 2 km/s and was applied using Gray's methods [14]. To generate synthetic spectra images similar to those coming from the HARPS-N pipeline, we split the spectra up into the échelle orders, and stitch them as an image. We limit the wavelength interval between 5050 to 5350 Å, which corresponds to 8 different èchelle orders. We interpolate the wavelengths of the 1D spectrum to match the 2D wavelengths, such that they correspond to the HARPS-N pipeline. We then add a linear slant across each order to mimic the observations from the HARPS-N pipeline. Initially, we sample the model spectra without any noise, such that we can add Gaussian noise corresponding to different SNR during training [8]. The parameters generated are discrete, but we can interpolate linearly between samples to create new observations, not on the grid, again similar to the procedure in [8].

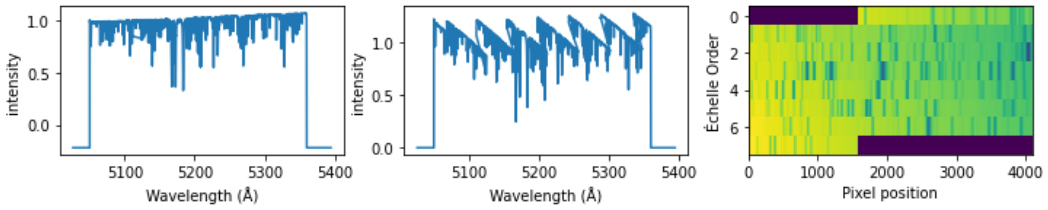

Figure 1: This figure highlights the different approaches. **Left:** A sample of a 1D model spectrum. **Middle:** A sample of a 1D model spectrum similar to the HARPS-N pipeline. **Right:** The spectral image of the spectrum in the middle.

3 Method

In this section, we describe how we to capture heteroscedastic uncertainty in the estimated stellar parameters, and present the attention-based architecture used in this work.

3.1 Heteroscedastic Uncertainty Estimation

Heteroscedastic regression assumes that the uncertainty of observations vary with input x [15]. This uncertainty in the observations can be quantified by the distribution $p(y|x)$, where the expected value is considered the best estimate of the parameters, and the variance of the distribution describes the uncertainty [16]. We then learn the distribution using parameters θ and parameterise the distribution as a Gaussian.

$$p(y|x,\theta) = \mathcal{N}\left(\mu_\theta(\mathbf{x}), \Sigma_\theta(\mathbf{x})^2\right) \tag{1}$$

In order to estimate the parameters in the function in Equation 1, we minimise the negative log-likelihood [15]:

$$\mathcal{L}(\theta) = \frac{N}{2}\log|\Sigma_\theta(x)| + \frac{1}{2}\sum_{i=1}^{N}\left(y_i - \mu_\theta(x_i)\right)^T \Sigma_\theta(x)^{-1}\left(y_i - \mu_\theta(x_i)\right) \tag{2}$$

We only consider the case where Σ is a diagonal matrix, restricting the variance within the single label estimation, instead of a full covariance matrix. We denote the diagonal covariance matrix with the vector $\sigma_\theta(\mathbf{x})$.

3.2 Attention-based model

The soft-attention model used in this work is inspired by the architecture from [17], in combination with the attention blocks presented in [18, 19]. We construct an attention architecture which uses any number of intermediate feature maps x_n from a convolutional neural network in combination with a global feature map g (The last layer in the convolutional neural network), to compute an attention map in attention blocks [17]. The attention map $\alpha_n \in [0, 1]$ is used to identify salient features in the input as the output of an attention block is the element-wise multiplication of the input feature-map and the attention map: $\hat{x}_n = a_n \cdot x_n$. We let F_n denote the number of channels in a given feature map. Formally we can compute the attention map a_n as follows.

$$\begin{aligned} q^l_{att} &= \boldsymbol{\psi}^T \left(\sigma_1 \left(\boldsymbol{W}_x^T \boldsymbol{x}_n + \boldsymbol{W}_g^T \boldsymbol{g} + \boldsymbol{b}_{xg} \right) \right) + b_\psi \\ \alpha_n &= \sigma_2 \left(q^l_{att} \left(\boldsymbol{x}_n, \boldsymbol{g}; \boldsymbol{\Theta}_{att} \right) \right) \end{aligned} \tag{3}$$

where $\sigma_1(x)$ is the activation function of the neural network, and $\sigma_2(x)$ is the *softmax* operation, such that the attention map sums to one [18]. The set of parameters $\boldsymbol{\Theta}_{att}$ contains the convolutional weights $\boldsymbol{W}_x \in \mathbb{R}^{F_x \times F_{int}}$ and $\boldsymbol{W}_g \in \mathbb{R}^{F_g \times F_{int}}$, which are used to linearly transform the input tensors using a channel-wise $1 \times 1 \times 1$ convolution. The weights $\boldsymbol{\psi} \in \mathbb{R}^{F_{int} \times 1}$ combines the features from all the channels into a 1-channel attention map. In addition we also include a bias term $b_\psi \in \mathbb{R}^{F_{int}}$ in $\boldsymbol{\Theta}_{att}$. The parameters of the attention block (the convolutional layer) are trained using standard back-propagation [18]. In order to ensure that all the attention maps α_n learn meaningful features, we average the outputs of the attention blocks in order to estimate the final $\mu_\theta(\mathbf{x})$ and $\sigma_\theta(\mathbf{x})$. The reader is referred to [17, 18, 19] for more in depth explanation of the architecture and attention blocks.

4 Experiments and results

This section summarises the results obtained on a test set of 5667 synthetic model spectra. We use the Adam [20] variant of stochastic gradient descent with a learning rate of 0.0001 to optimise the models. We train the models for 750 epochs and for the last 50 epochs we decay the learning rate by a γ of 0.1. In order to increase performance for low SNR, we also train a denoising auto-encoder (DAE) [21] to denoise samples before we train a supervised learning model. We train the DAE for 1500 epochs. We use the AdamW [22] variant of SGD with a learning rate of 0.0003; we decay the learning rate the as above. The residual network consists of 5 residual blocks [23] and three fully connected layers. The attention-network contains three attention blocks and 11 convolutional layers. We add Gaussian noise to simulate an SNR ≈ 20. In Table 1 we present the *MAE* performance of the models on the 4 stellar parameters.

SNR	Model	T_{eff}	$\log(g)$	Z	$V \sin i$
20	Residual-network	76.9	0.138	0.055	0.71
20	Attention-network	73.0	0.135	0.053	0.69
20	DAE-Residual-network	72.3	**0.133**	0.052	0.67
20	DAE-Attention-network	**70.9**	0.134	**0.049**	**0.57**
200	Cannon2 [24]	46.8	0.066	0.036	-
200	StarNet [8]	31.2	0.053	0/.025	-
100	Residual-network†	19.5	0.053	0.026	0.30
100	Attention-network†	**12.9**	**0.045**	**0.013**	**0.15**

Table 1: Mean-Absolute error based on the mean prediction from models. † models are trained on a limited data-set to match the parameter ranges presented in previous related work [8]. The limited data-set only contains T_{eff} between 4000K and 6000K, and other parameters stay the same.

From Table 1 we see that using the DAE to denoise samples increase the performance when the SNR ≈ 20. The reader should note that [8, 24] uses a different data-pipeline, so direct comparison should be performed with caution.

4.1 Uncertainty estimation

In order to assess the quality of the estimated standard deviations, we compare the percentages of residuals within 1 or 2 standard deviations as an evaluation of the uncertainty estimation.

	Gaussian	Residual-network	DAE-Residual network	Attention-network
$\lvert\epsilon\rvert < \sigma_\theta(x)$	68.2%	79.9%	77.9%	65.5%
$\lvert\epsilon\rvert < 2\sigma_\theta(x)$	95.1%	98.4%	97.9%	93.4%

Table 2: Table showing percentages of residuals for the parameters that are within $\mu \pm \sigma$ and $\mu \pm 2\sigma$. The models are trained with SNR ≈ 20

From Table 2 we see that the estimated distributions approximate the Gaussian theoretical values, making us conclude that the estimated standard deviations can be used to create Gaussian confidence intervals.

4.2 HARPS-N observation of the Sun

In order the evaluate the synthetic gap, we assess the models on an observation of the Sun coming from the HARPS-N spectrograph. Due to the extreme apparent brightness of the Sun, these observations obtain high SNR of ≈ 200 [25, 26]. All real observations contain telluric lines (absorption from Earths atmosphere), which is not present in the synthetic data.

SNR	Model	T_{eff}	log(g)	Z	V sin i
200	Ground Truth	5750	4.44	0	2
100	Residual-network	5791.6 ± 140.1	4.72 ± 0.28	0.035 ± 0.15	0.762 ± 1.76
100	Attention-network	5325.2 ± 10.0	2.15 ± 0.04	-0.576 ± 0.01	5.226 ± 0.40

Table 3: Estimated values for the Sun observation. Confidence bands are estimated using Gaussian confidence intervals

From Table 3 we can see that the models can estimate the parameters of the Sun, and use the estimated uncertainty to setup confidence intervals. We see that the attention model is more susceptible to the synthetic gap, whereas the residual network includes dropout in the last dense layers, which allows for some uncertainty in the input features, as e.g. addition of telluric lines in a spectrum.

4.3 Visual inspection of attention

Finally, we visualise the attention feature maps α_n across the 3 different blocks; illustrating the interpretation possibilities in the attention blocks. We also include annotations for some of the elements used in traditional methods for stellar parameter estimation such as the magnesium b lines.

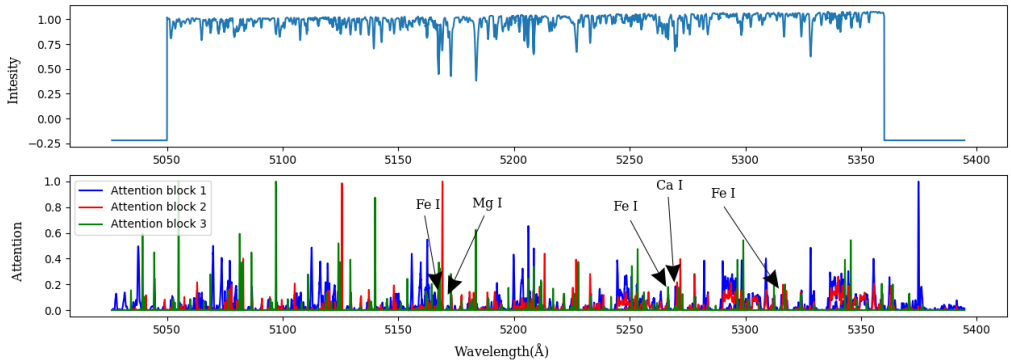

Figure 2: Visualisation of attention α from the 3 different blocks.

5 Conclusion

In this work, we focused on data-driven estimation of stellar parameters based on the spectral signal directly from the HARPS-N pipeline. Our models can estimate uncertainties, and we can use a DAE to enhance model performance for low SNR. We show that the attention model can attend to the underlying elements in a spectrum. We suggest a semi-supervised learning approach, where a real-observations and synthetic spectra could make models more robust to the synthetic gap.

Broader Impact

While it is still essential for physical science to reason about stellar parameters, e.g. by looking at the absorption lines of a spectrum to see the composite elements, we have here focused on an end-to-end deep learning approach and provided methods for estimating the stellar parameters consistently and accurately. We see great opportunities for researchers to apply the methods here to investigate the Universe efficiently and we believe that there are many areas within the physical sciences which can benefit from this research, such as the stellar astrophysical research on large-scale statistics. A specific task which this research contribute to is when the Signal to noise ratios of a sample is lower than anticipated, then the methods provided here can be used to increase the usability of the sample, by removing undesired noise and get an unbiased estimate of the stellar parameters.

There are important considerations to take into account when using synthetic spectra to train and evaluate models. Models trained with the synthetic spectra, will all be biased as the models inherit the synthetic gap between real observations and synthetic spectra. We, therefore, argue that researchers within this field, should keep this in mind and understand the biases and limitations related to the use of synthetic spectra. We believe future research could apply a combination of real observations and different synthetic spectra to account for this bias.

The potential risk of increasing the amount of deep learning into the astrophysical research areas could lead to a neglect of the physical knowledge required to interpret the spectra, which potentially will lead to a loss of physical knowledge. We, therefore, encourage future research to continue the path towards an end-to-end deep learning method, while acknowledging the importance of the physical composition of the underlying spectra.

Acknowledgements

We would like to thank Lars A. Buchhave and Alexander Dybdahl Ratchke from the National Space Institute at the Technical University of Denmark, whom have been the inspiration for this work and provided the data-set.

References

[1] Young Lee, Timothy Beers, Thirupathi Sivarani, Carlos Prieto, Lars Koesterke, Ronald Wilhelm, Paola Re Fiorentin, Coryn Bailer-Jones, John Norris, Constance Rockosi, Brian Yanny, Heidi Newberg, Kevin Covey, Haotong Zhang, and A-Li Luo. The segue stellar parameter pipeline. i. description and comparison of individual methods. *The Astronomical Journal*, 136:2022, 10 2008.

[2] Carlos Prieto, Thirupathi Sivarani, Timothy Beers, Young Lee, Lars Koesterke, Matthew Shetrone, Christopher Sneden, David Lambert, Ronald Wilhelm, Constance Rockosi, David Lai, Brian Yanny, ii Ivans, Jennifer Johnson, Wako Aoki, Coryn Bailer-Jones, and Paola Re Fiorentin. The segue stellar parameter pipeline. iii. comparison with high-resolution spectroscopy of sdss/segue field stars. *The Astronomical Journal, v.136, 2070-2082 (2008)*, 136, 11 2007.

[3] Brian Yanny, Constance Rockosi, Heidi Newberg, Gillian Knapp, Jennifer Adelman-McCarthy, Bonnie Alcorn, Sahar Allam, Carlos Prieto, Deokkeun An, Kurt Anderson, Scott Anderson, Coryn Bailer-Jones, Steve Bastian, Timothy Beers, Eric Bell, Vasily Belokurov, Dmitry Bizyaev, Norm Blythe, John Bochanski, and Y. Wadadekar. Segue: A spectroscopic survey of 240,000 stars with g = 14-20. *The Astronomical Journal, v.137, 4377-4399 (2009)*, 137, 05 2009.

[4] Young Lee, Timothy Beers, Jeffrey Carlin, Heidi Newberg, Yonghui Hou, Guangwei Li, A-Li Luo, Yue Wu, Ming Yang, Haotong Zhang, Wei Zhang, and Yong Zhang Niaot. Application

of the segue stellar parameter pipeline to lamost stellar spectra. *The Astronomical Journal*, 150:187, 12 2015.

[5] R. Smiljanic, A. J. Korn, M. Bergemann, A. Frasca, L. Magrini, T. Masseron, E. Pancino, G. Ruchti, I. San Roman, L. Sbordone, and et al. Thegaia-eso survey: The analysis of high-resolution uves spectra of fgk-type stars. *Astronomy Astrophysics*, 570:A122, Oct 2014.

[6] Coryn A. L. Bailer-Jones, Mike Irwin, Gerard Gilmore, and Ted von Hippel. Physical parametrization of stellar spectra: the neural network approach. *Monthly Notices of the Royal Astronomical Society*, 292(1):157–166, 11 1997.

[7] Coryn A. L. Bailer-Jones. Stellar parameters from very low resolution spectra and medium band filters te , log g and (m/h) using neural networks. 2000.

[8] Sebastien Fabbro, Kim Venn, Teaghan O'Briain, Spencer Bialek, Collin Kielty, Farbod Jahandar, and Stephanie Monty. An application of deep neural networks in the analysis of stellar spectra, 2017.

[9] M. Ness, David W. Hogg, H.-W. Rix, Anna. Y. Q. Ho, and G. Zasowski. The cannon: A data-driven approach to stellar label determination. *The Astrophysical Journal*, 808(1):16, Jul 2015.

[10] M. Mayor, F. Pepe, D. Queloz, F. Bouchy, G. Rupprecht, G. Lo Curto, G. Avila, W. Benz, J. L. Bertaux, X. Bonfils, Th. Dall, H. Dekker, B. Delabre, W. Eckert, M. Fleury, A. Gilliotte, D. Gojak, J. C. Guzman, D. Kohler, J. L. Lizon, A. Longinotti, C. Lovis, D. Megevand, L. Pasquini, J. Reyes, J. P. Sivan, D. Sosnowska, R. Soto, S. Udry, A. van Kesteren, L. Weber, and U. Weilenmann. Setting New Standards with HARPS. *The Messenger*, 114:20–24, December 2003.

[11] RL Kurucz and B Bell. Kurucz cd-rom 13. *ATLAS9 stellar atmosphere programs and*, 2:1, 1993.

[12] Robert L Kurucz. Atlas: A computer program for calculating model stellar atmospheres. *SAO Special report*, 309, 1970.

[13] Fiorella Castelli and Robert L Kurucz. New grids of atlas9 model atmospheres. *arXiv preprint astro-ph/0405087*, 2004.

[14] David F Gray. *The observation and analysis of stellar photospheres*. Cambridge University Press, 2005.

[15] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *CoRR*, abs/1703.04977, 2017.

[16] D. A. Nix and A. S. Weigend. Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, volume 1, pages 55–60 vol.1, 1994.

[17] Saumya Jetley, Nicholas A. Lord, Namhoon Lee, and Philip H. S. Torr. Learn to pay attention. *CoRR*, abs/1804.02391, 2018.

[18] Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, Matthew C. H. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. *CoRR*, abs/1804.03999, 2018.

[19] Jo Schlemper, Ozan Oktay, Liang Chen, Jacqueline Matthew, Caroline L. Knight, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention-gated networks for improving ultrasound scan plane detection. *CoRR*, abs/1804.05338, 2018.

[20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

[21] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. In *International Conference on Machine Learning proceedings*. 2008.

[22] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in adam. *CoRR*, abs/1711.05101, 2017.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[24] Andrew Casey, David W. Hogg, Melissa K. Ness, Hans-Walter Rix, Anna Y. Q. Ho, and Gerry F. Gilmore. The cannon 2: A data-driven model of stellar spectra for detailed chemical abundance analyses. 2016.

[25] Xavier Dumusque, Alex Glenday, David F. Phillips, Nicolas Buchschacher, Andrew Collier Cameron, Massimo Cecconi, David Charbonneau, Rosario Cosentino, Adriano Ghedina, David W. Latham, and et al. Harps-n observes the sun as a star. *The Astrophysical Journal*, 814(2):L21, Nov 2015.

[26] N. Langellier, T. W. Milbourne, D. F. Phillips, R. D. Haywood, S. H. Saar, A. Mortier, L. Malavolta, S. Thompson, A. Collier Cameron, X. Dumusque, H. M. Cegla, D. W. Latham, J. Maldonado, C. A. Watson, M. Cecconi, D. Charbonneau, R. Cosentino, A. Ghedina, M. Gonzalez, C-H. Li, M. Lodi, M. López-Morales, G. Micela, E. Molinari, F. Pepe, E. Poretti, K. Rice, D. Sasselov, A. Sozzetti, S. Udry, and R. L. Walsworth. Detection limits of low-mass, long-period exoplanets using gaussian processes applied to harps-n solar rvs, 2020.