Consistent and Accurate Estimation of Stellar Parameters from HARPS-N Spectroscopy

Technical University of Denmark

Frederik Boe Hüttel and Line Katrine Harder Clemmensen

DTU Compute · Technical University of Denmark
Kgs. Lyngby, Denmark

DTU Compute
Department of Applied Mathematics and Computer Science

Introduction

- We present methods and results from estimating stellar parameters from the HARPS-N data-pipeline using deep learning models. This approach eliminates the need for spectral pre-processing steps to extract the 1D spectra.
- We quantify the uncertainty in estimations of the stellar parameters. The estimated distribution provides a basis to create data-driven confidence intervals.

Data

The data-set is altered to mimic a complete end-to-end approach

- The use of synthetic spectra provides a unique opportunity to generate a large set of labelled data.
- Addition of the échelle orders to mimimc data comming from the HARPS-N data pipeline.

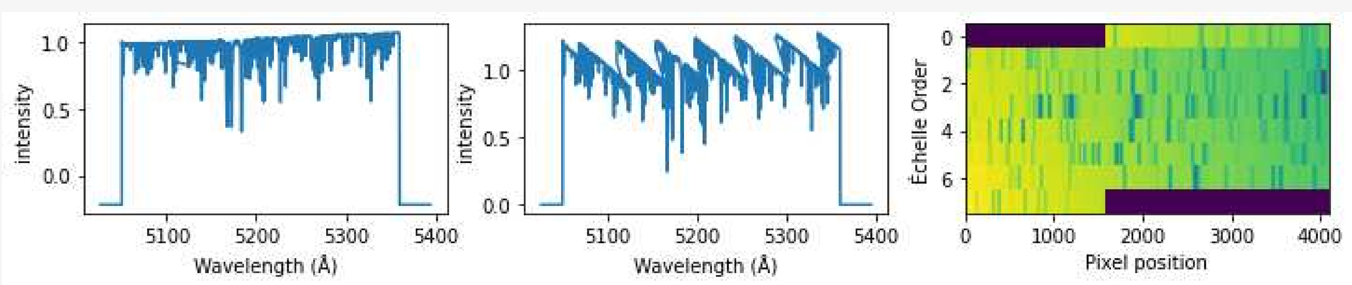


FIGURE 1: *Left:* A sample of a 1D model spectrum *Middle:* A sample of a 1D model spectrum similar to the HARPS-N pipeline, by inclusion of the èchelle orders *Right:* The spectral image of the spectrum in the middle

Model Specification

Heteroscedastic Uncertainty Estimation

We assume that the uncertainty in the stellar parameters y varies with the input $\mathbf{x}$, and we model this uncertainty through a Gaussian distribution.

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

We minimise the negative log-likelihood through the parameters θ using SGD.

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

Attention Model

- Uses an intermediate feature map x_n from a convolutional neural network in combination with a global feature map g to compute an attention map $\alpha_n \in [0, 1]$
- We compute an attention map $\alpha_n \in [0, 1]$, the output of an attention block is the element-wise multiplication of an the input feature-map and the attention map: $\hat{x}_n = a_n \cdot x_n$.

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

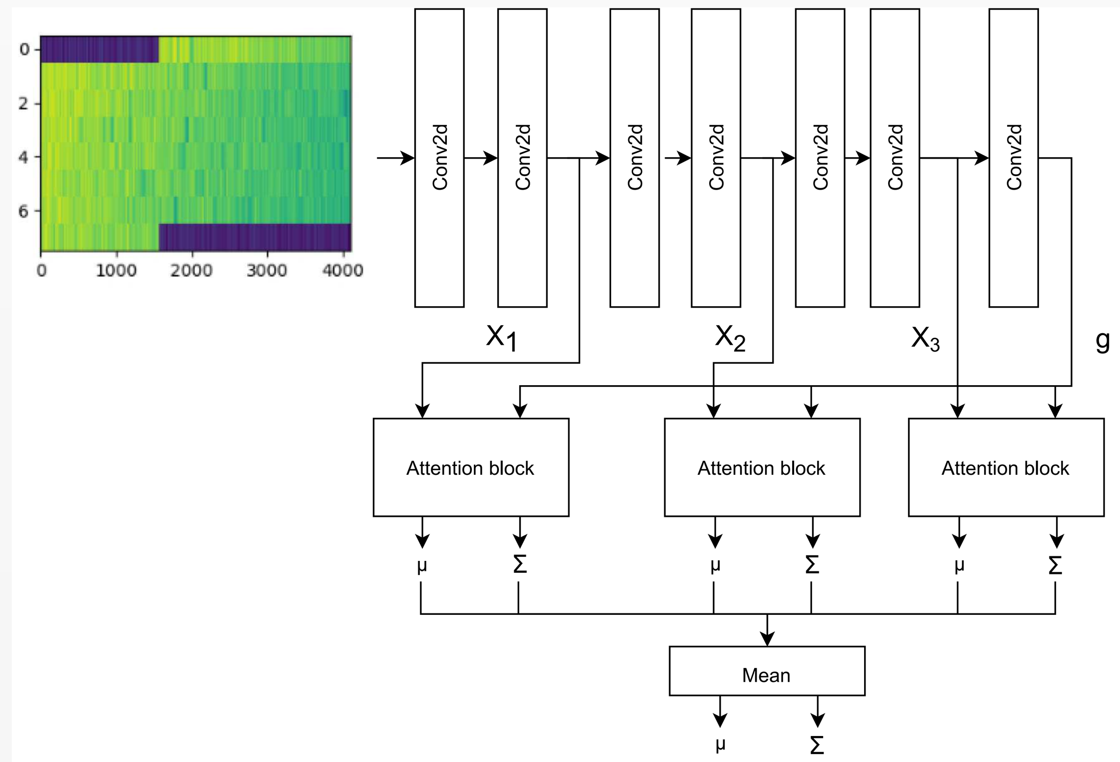


FIGURE 2: Figure showing the overall architecture of the attention model used in this work.

Experiments

Performance metrics

SNR	$\Sigma_\theta(\mathbf{x})$	Model	T_{eff}	log(g)	Z	$V\sin i$
20	Diagonal	Residual-network	76.9	0.138	0.055	0.71
20	Diagonal	Attention-network	73.0	0.135	0.053	0.69
20	Diagonal	DAE Residual-network	72.3	**0.133**	0.052	0.67
20	Diagonal	DAE Attention-network	**70.9**	0.134	**0.049**	**0.57**
20	Full	Residual-network	83.8	0.143	0.060	0.75
20	Full	Attention-network	79.2	0.146	0.055	0.72
20	Full	DAE Residual-network	89.1	0.150	0.060	0.72
20	Full	DAE Attention-network	72.9	0.137	**0.049**	0.58
200	-	Cannon2 [1]	46.8	0.066	0.036	-
200	-	StarNet [2]	31.2	0.053	0.025	-
100	Diagonal	Residual-network†	19.5	0.053	0.026	0.30
100	Diagonal	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 [2].

Based on the results present in Table 1, we here continue with the results from the attention model, with a diagonal covariance matrix

Uncertainty Estimation

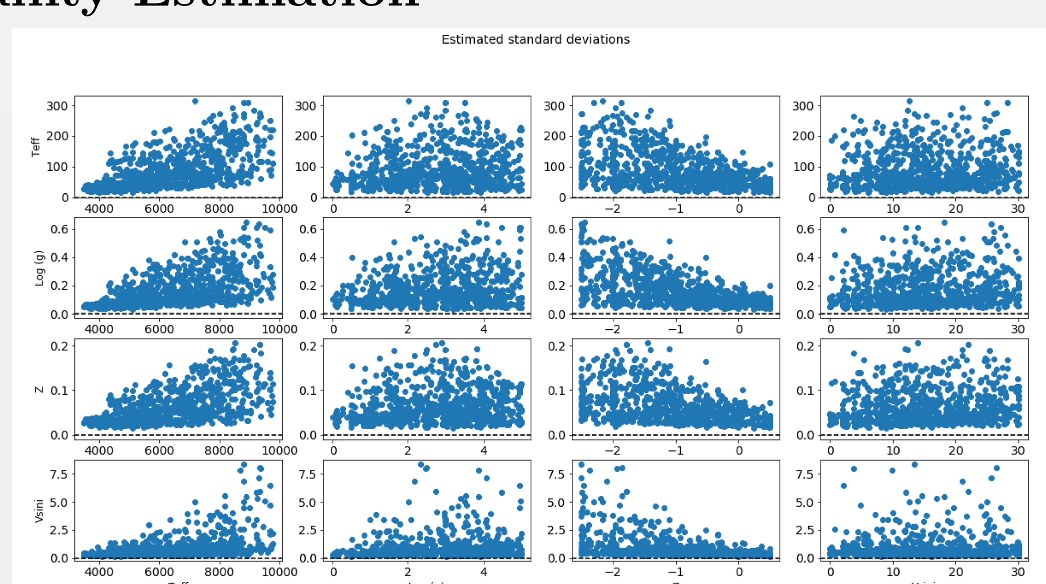


FIGURE 3: The estimated standard deviation from the residuals as a function of the true parameters.

- We conclude that the estimated uncertainty is depend on $\mathbf{x}$, which shows that the assumption of heteroscedasic variance across the input is valid.
- We find that that the estimated distributions approximate the Gaussian theoretical values, making us conclude that the estimated standard deviations can create data-driven Gaussian confidence intervals

Model	$\epsilon < \Sigma_\theta(\mathbf{x})$	$\epsilon < 2\Sigma_\theta(\mathbf{x})$
Gaussian	68.2%	95.1%
Residual-network	79.9%	98.4%
DAE-Residual network	77.9%	97.9%
Attention-network	65.5%	93.4%

TABLE 2: Table showing percentages of observations that are within $\mu \pm \Sigma_\theta(\mathbf{x})$ and $\mu \pm 2\Sigma_\theta(\mathbf{x})$. The models are trained with SNR ≈ 20.

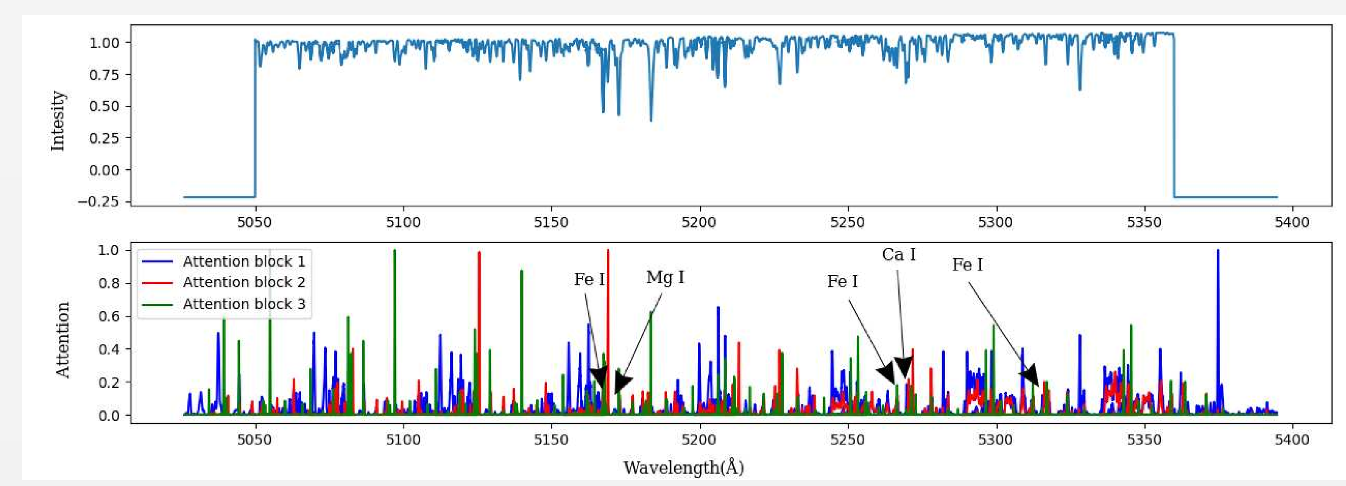
Test on HARPS-N observation

Model	T_{eff}	log(g)	Z	$V\sin i$
HARPS-N	5750	4.44	0	2
Residual-net	5791.6 ± 140.1	4.72 ± 0.28	0.035 ± 0.15	0.762 ± 1.76
Attention-net	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.

Visualisation of attention map

The magnesium b have spectra lines at 5172 A and is often used by traditional methods when estimating the stellar parameters. We find that he attention-network is attending to this element, based on the high activation of the attention feature map α at this absorption line



Conclusions & Outlook

- We have focused on a data-driven estimation of stellar parameters based on the spectral signal directly from the HARPS-N pipeline.
- The estimation of a multivariate Gaussian also lays the groundwork for future research ideas to explore ideas of full Bayesian approaches such as MCMC or variational inference methods.
- The attention models provide a way to reason about the importance of the different composite elements of a spectrum.

References

[1] 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.

[2] 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.