
Unsupervised Learning for Stellar Spectra with Deep Normalizing Flows

Ioana Ciucă^{*12} Yuan-Sen Ting^{*12}

Abstract

Stellar spectra encode detailed information about the stars. However, most machine learning approaches in stellar spectroscopy focus on supervised learning. We introduce `Mendis`, an unsupervised learning method, which adopts normalizing flows consisting of Neural Spline Flows and GLOW to describe the complex distribution of spectral space. A key advantage of `Mendis` is that we can describe the conditional distribution of spectra, conditioning on stellar parameters, to unveil the underlying structures of the spectra further. In particular, our study demonstrates that `Mendis` can robustly capture the pixel correlations in the spectra leading to the possibility of detecting unknown atomic transitions from stellar spectra. The probabilistic nature of `Mendis` also enables a rigorous determination of outliers in extensive spectroscopic surveys without the need to measure elemental abundances through existing analysis pipelines beforehand.

1. Introduction

As probes of stellar nucleosynthesis, stellar spectra provide an unparalleled testbed for atomic physics in the era of high-resolution stellar spectra from surveys such as GALAH (Buder et al., 2021) and APOGEE (Majewski et al., 2017). They offer a window into identifying new atomic lines, which may otherwise evade the notoriously difficult *ab initio* quantum mechanics calculations (Hasselquist et al., 2016; Cunha et al., 2017). Stellar spectra also reveal the fundamental properties of stars, including their stellar parameters (temperature, surface gravity, and global metallicity) and their chemical composition, which holds the key to unravel-

ling the evolutionary history of our Milky Way galaxy.

Our ever-expanding power to collect spectra has called for more advanced machine learning methods to analyze them. However, most have focused on supervised learning to solve a label determination problem when analyzing spectroscopic data with machine learning (Ness et al., 2015; Ting et al., 2019), which is somewhat limiting. On the one hand, there are only a handful of stars with high-fidelity stellar labels (stellar parameters and elemental abundances) (Jofré et al., 2014). On the other hand, due to its inherent complexity, stellar spectral models elude direct physical modelling, even as more advanced theoretical prescriptions become available (Nordlander et al., 2017). This dilemma has often led to perennial debates regarding the pros and cons of applying “data-driven” or “model-driven” analysis when dealing with stellar label determination. Additionally, label determination through supervised learning projects the data onto some predefined “known” label space, significantly impeding the search for exciting unknown outliers.

These challenges pose the question: can we learn from the spectra themselves without first reducing them to stellar labels. Unfortunately, unsupervised learning has not seen the same surge in applications for stellar spectroscopy. Pioneering works of Price-Jones & Bovy (2018) and de Mijolla et al. (2021) employed methods such as Principal Component Analysis (PCA) and Variational Auto-Encoders (VAE) to extract a representation of stellar spectra. However, these methods are limited in a few key areas. For example, the PCA framework lacks the flexibility to model the data, such as modeling conditional distributions, which are of paramount importance, as we will demonstrate in this study. The posterior embedding distribution learned in a VAE often deviates from the prior distribution and is not able to provide exact likelihood evaluation.

In this study, we explore normalising flows, a class of deep generative models, to bridge this gap. We will show that this more principled way of describing the statistical distribution of stellar spectra through unsupervised learning can lead to new opportunities, including correlation identification and out-of-distribution detection.

^{*}Equal contribution ¹Research School of Astronomy & Astrophysics, Australian National University, Cotter Rd., Weston, ACT 2611, Australia ²School of Computing, Australian National University, Acton, ACT 2601, Australia. Correspondence to: Ioana Ciuca <ioana.ciuca@anu.edu.au>, Yuan-Sen Ting <yuan-sen.ting@anu.edu.au>.

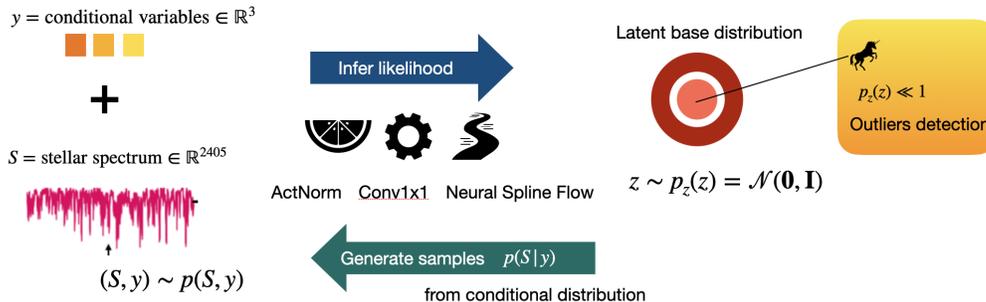


Figure 1. Deep normalizing flows for stellar spectra. Our model (Mendis) turns an ensemble of training spectra into a Gaussian base distribution from which we can perform exact likelihood evaluation and search for out-of-distribution outliers. As normalizing flows are invertible, they allow us to generate sample from conditional distribution and evaluate all moments and correlations from the distribution.

2. Synthetic Stellar Spectra

Our training data consists of APOGEE-like synthetic high-resolution stellar ($R \sim 22,000$) spectra generated using the Kurucz models (Kurucz & Avrett, 1981; Kurucz, 2005; 2017). Kurucz models solve for a 1D stellar atmosphere with ATLAS2. The atmosphere is then used to generate realistic spectra through a radiative transfer model with SYNTH3, assuming Saha and Boltzmann equilibrium. The spectral simulator allows us to access a varied stellar feature space as input, namely in effective temperature, surface gravity, metallicity, and elemental abundance space consisting of C, N, O, Na, Mg, Al, Si, P, S, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni and Cu. We generate 20,000 spectra by uniformly sampling in temperature between 4300 and 4500 K, surface gravity between 1.9 and 2.2 dex, and elemental abundance between -0.1 and 0.05 dex for all elements apart from nitrogen, which we sample between 0.0 and 0.15 dex. The native resolution of the Kurucz models is of $R = \lambda/\Delta\lambda = 300,000$, where λ is the wavelength. We convolve this model using the mean line spread function from the APOGEE observations (Holtzman et al., 2015). We further assume a Nyquist sampling; we store one pixel per APOGEE resolution element $\Delta\lambda$, leading to a spectrum with 2405 pixels.

To mimic actual observations, we degrade our spectra to a signal-to-noise ratio of 300, corresponding to the level of high-fidelity observed APOGEE spectra. Recall that our goal is to construct a conditional distribution of spectra given stellar parameters. Since the ground truth stellar parameters are often unknown in actual data (but can be estimated from the color-magnitude diagram or the spectra themselves), we also assume a 25 K scatter in effective temperature, 0.1 dex in surface gravity and 0.01 dex in metallicity.

3. Normalizing Flows for Stellar Spectra

In this study, we assume a normalizing flow model to solve an unsupervised¹ task of learning a model that best represents the distribution $p(\mathbf{x})$ from which the training samples $\{\mathbf{x}_i\}$ were drawn. At its core, the normalizing flow aims to find a change of variable f_θ , such that the transformed variable $\mathbf{z} \equiv f_\theta(\mathbf{x})$ is normally distributed, i.e. $p_z(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$, effectively finding an invertible function that ‘‘Gaussianize’’ the distribution (Dinh et al., 2014; 2016; Kingma & Dhariwal, 2018).

The training task of normalizing flows involves optimizing the parameters of f_θ by maximizing the log-likelihood over the training set \mathcal{D} , $\mathcal{L}(\mathcal{D}) = \frac{1}{D} \sum_{\mathbf{x} \in \mathcal{D}} \log p(\mathbf{x})$, where

$$p_{\mathbf{x}}(\mathbf{x}) = p_z(f_\theta(\mathbf{x})) |\det J_f(\mathbf{z})| \quad (1)$$

and J the Jacobian of the transformation.

In our study, we have $\mathbf{x} = (\mathbf{S}, \mathbf{y}) \in \mathbb{R}^{2405+3}$, where \mathbf{S} is the stellar spectrum, and \mathbf{y} three stellar parameters, namely the effective temperature, surface gravity and the metallicity. Learning the joint distribution $p(\mathbf{S}, \mathbf{y})$ allows us to evaluate the conditional distribution $p(\mathbf{S}|\mathbf{y})$ which will prove critical. While in theory it is possible to model the conditional distribution directly as done, for example, in Ting & Weinberg (2021), in practice training the model on this large dimensionality is prohibitively difficult.

Architecture: Extensive exploration on architectures such as RealNVP (Dinh et al., 2016), SurVAE (Nielsen et al., 2020) and Neural Spline Flows (Durkan et al., 2019) have led to our final architecture, which we dub the name Mendis. Mendis composes of an Activation Normalization (ActNorm) layer and a GLOW (‘Conv1x1’) operation, both introduced in Kingma & Dhariwal (2018), followed by a Neural Spline Flow. The ActNorm layer performs an

¹Since this study also assumes stellar parameters, some might argue that the model should be regarded as weakly supervised. However, as the stellar parameters are only used as conditional variables, we opt to stick to the unsupervised terminology.

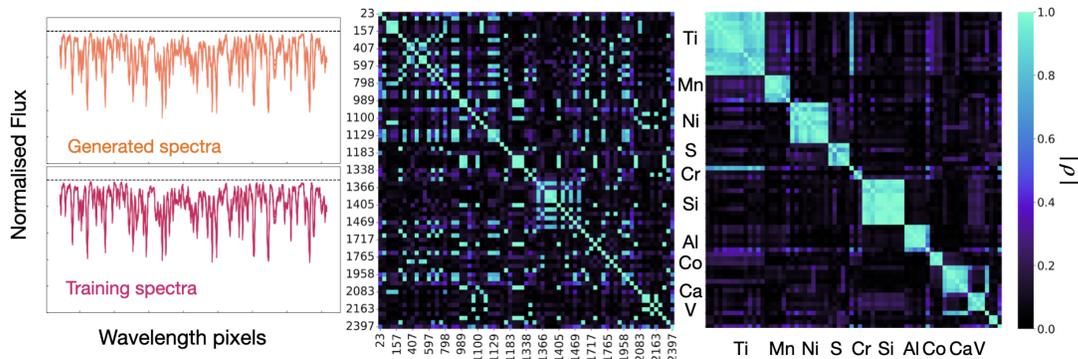


Figure 2. Our generative models produce realistic stellar spectra (upper left panel) indistinguishable from the training spectrum (lower left panel). The probabilistic nature of our models further allows the calculation of the correlation structures between wavelengths and reveals which atomic transitions at different wavelengths are contributed by the same elements. The middle panel shows the correlation matrix sorted by wavelengths, and the right panel by elements. The re-indexing is done by taking the former as an adjacency matrix of a graph.

affine transformation and is used to improve convergence during the training task. GLOW is a block-diagonal linear transformation that essentially rearranges and mixes the dimensions of the input vector before reaching the coupling layer to maximize information sharing across all the input variables. Finally, the Neural Spline Flow employs a monotonic rational-quadratic transform that performs a cubic spline transformation on the input (Durkan et al., 2019). These transforms end up in a tractable Jacobian associated with their lower diagonal matrix and can be inverted in a single pass. Hence they allow for fast likelihood evaluation and sampling, which makes them a powerful choice.

Training approach: A key advantage of normalizing flows, as opposed to other generative models such as Generative Adversarial Networks, is that the maximum likelihood loss function is resilient against mode collapse (Bond-Taylor et al., 2021). However, this comes with the cost. As already alluded to in Equation 1, normalizing flows require neural networks to be invertible and have tractable and efficiently computable Jacobian. This stringent constraint often makes normalizing flows less expressive. As such, we found that an extensive normalizing flow is needed to deal with the considerable dimensions in the spectral space where $x \in \mathbb{R}^{2405+3}$. We choose a normalizing flow with 20 units of ActNorm, GLOW and Neural Spline coupling layer. This architecture choice leads to a final network with approximately a billion parameters.

Our code is made multi-node multi-GPU parallelized across 16 Nvidia A100 GPUs using the Distributed Data-Parallel functionality of PyTorch to train this massive network. We optimize our parameters using Adam (Kingma & Ba, 2014), adopt a learning rate of 5×10^{-6} and train for 3,000 epochs. Training across 16 A100 takes approximately 7 hours. All our codes are made available on Github.

4. New Frontiers in Stellar Spectroscopy

In the following, we present two applications to demonstrate the potential of normalising flows for stellar spectroscopy by harnessing their unique capabilities of describing the underlying probabilistic distribution of the spectra.

Unsupervised Learning of Atomic Features: Although spectra are vectors with thousands of dimensions (pixels at different wavelengths), they are also remarkably simple. Most pixels are highly correlated due to the underlying physics: the absorption features are formed from different atomic transitions, resulting from quantum mechanics. As various elements have distinct atomic transitions at different wavelengths, the underlying correlation structures of spectral are therefore tell-tale signs of contributions from specific elements, but with one caveat. Apart from the elemental abundances, stars with different stellar parameters will also alter the opacity of the stellar photosphere and hence the emergent spectrum. Therefore, if we were to evaluate the correlation matrix from $p(\mathbf{S})$, the correlation matrix is dominated by the contribution of the stellar parameters, obscuring the contributions from individual elements.

Modeling the training spectra with `Mendis` enables a new window to tackle this problem. More specifically, once we have trained a robust statistical description of $p(\mathbf{S}, \mathbf{y})$, the model allows us to evaluate all moments of the conditional distribution $p(\mathbf{S}|\mathbf{y})$ through repeated sampling from the normalizing flows at a fixed \mathbf{y} . The conditioning will allow us to eliminate all influence on the spectra from the stellar parameters, further revealing only contribution from the chemical composition of the stars. Importantly, this statement holds even if the unsupervised model itself is trained with a sample without any two stars sharing the same \mathbf{y} .

The left panels of Fig. 2 demonstrates how well `Mendis` perform in terms of spectral reconstruction. The bottom

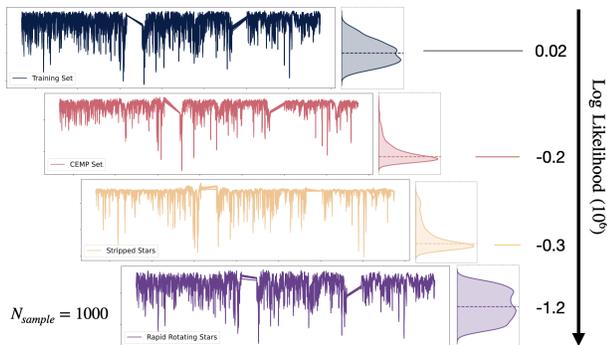


Figure 3. Out-of-distribution detection with *Mendis*. Various outlier classes, such as CEMP stars, stripped stars and rapidly rotating stars are assigned to have a low likelihood by *Mendis* compared to the training data, even though their spectra, as shown in the inset plots, might not look too dissimilar. Without any predetermined stellar labels, *Mendis* provides an automated way to triage outlier objects of interest, sifting through large spectroscopic data.

panel shows ten examples of a training spectrum $\tilde{\mathbf{x}}$, and the top panel shows their closest counterparts determined by computing $\text{argmin}_{\mathbf{z}} |\tilde{\mathbf{x}} - f_{\theta}^{-1}(\mathbf{z})|^2$ from spectra $f_{\theta}^{-1}(\mathbf{z})$ generated from *Mendis*. The two are indistinguishable from each other, demonstrating that our normalizing flow is a faithful surrogate of the training spectra.

The middle panel further demonstrates the correlation matrix evaluated from $p(\mathbf{S}|\mathbf{y})$, where we fix the conditioning stellar parameters \mathbf{y} at an effective temperature of 4300 K, surface gravity 2.1 dex and solar metallicity. The middle panel shows the “native” correlation matrix sorted by the pixel wavelengths. We only show a subset of pixels with strong absorption features. The correlation matrix is sparse and unstructured because atomic transitions from the same elements can occur at vastly different wavelengths. To better visualize the high-correlation clusters, we then turn the correlation matrix into an adjacency matrix of a graph and assume an edge between the two pixels if their correlation is larger than 0.8. We then use the adjacency matrix to re-index the original correlation matrix, essentially searching for connected “trees” within the graph. The re-indexed version of the correlation matrix is shown on the right.

The right panel shows that, by fixing the stellar parameter, *Mendis* reveals the striking underlying correlation from the spectra. Different elements are sorted as block diagonals in the correlation matrix. Notably, the high correlation structure endures even though we have trained using noisy spectra and labels. The result demonstrates that unsupervised learning could unearth unknown atomic transitions through empirical data, even though we are entirely agnostic about their elemental abundances.

Out-of-distribution detection: We have seen many extensive spectroscopic surveys come to fruition in recent years, collecting tens of millions of spectra. With this vast spectro-

scopic data, there are bound to be many unknown unknowns lurking in the data. However, supervised learning tends to project all data to the model “known” space, limiting our ability to find the exciting outliers.

Mendis, as an unsupervised method, provides a new opportunity to sift through high-dimensional information. It identifies unlabelled unique objects without requiring label determination beforehand. As *Mendis* describes the distribution of spectra $p(\mathbf{S})$, the likelihood for any object can be simply evaluated according to Equation (1). Specifically, for any new datum $\tilde{\mathbf{S}}$, the likelihood $p(\tilde{\mathbf{S}})$ evaluates if $\tilde{\mathbf{S}}$ is within the distribution $p(\mathbf{S})$ spanned by the training set, or if it is something that the model has not encountered.

As a proof of concept, here we built three classes of some of the more sought after “known unknowns” in the study of galaxy evolution (Carollo et al., 2014; Göteborg et al., 2019, e.g.,). Particularly, we generate from the Kurucz models spectra from carbon-enhanced metal-poor (CEMP) stars ($-2.0 < [\text{Fe}/\text{H}] < -1.5$ and $1 < [\text{C}/\text{Fe}] < 2$), rapid-rotating stars ($v_{\text{macro}} > 20$ km/s), and stripped stars (effective temperature = 4500 – 7000 K, surface gravity = 4.5 – 6). Fig. 3 shows that directly using only the spectra, *Mendis* successfully discriminates between the training set and these extreme-valued test samples. *Mendis* assigns a low likelihood to these outlying test samples, signaling that they are objects of interest and merit follow-ups.

While we study known unknowns in this case study, we emphasize that *Mendis* is completely agnostic about stellar labels. It simply assigns a low likelihood to spectra that do not look like anything else in the training data and can be generalized to true unknown unknowns. Our results demonstrate that *Mendis* can serve as a generic spectral broker for any spectroscopic surveys, independent to existing label determination pipelines, to triage for outliers.

5. Prospects and Future Directions

A common criticism of unsupervised learning is that it often lacks interpretability. This study shows that this needs not to be the case. Unsupervised learning with normalizing flows retains its interpretability, as demonstrated by *Mendis*’s ability to identify pixel correlations. Besides, *Mendis* also provides a more principled way to perform out-of-distribution detections. But more generally, the statistical description of the distribution of spectra through *Mendis* has a wide range of applications beyond the few case studies explored here. For example, *Mendis* can serve as a bridge for supervised and unsupervised learning – the learned distribution can serve as the prior distribution for domain adaptation to close the model-data synthetic gap (O’Brian et al., 2021). Additionally, the distribution can serve as the basis for semi-supervision and few-shots learning with a limited

number of stars with high-fidelity labels.

Although *Mendis* holds many promises, it has much room for improvement. Notably, since spectra lack obvious physical symmetry, currently, a “brute-force” large normalizing flow with a billion parameters is chosen to depict the spectral distribution robustly. Recent normalizing flow attempts in cosmology have taught us that embedding physical knowledge (in cosmology, spherical symmetry) can significantly simplify the architecture (Dai & Seljak, 2022). Since most information in spectra is encoded in their pixel correlations, we posit that preprocessing the data with self-attention modules to pre-learn the pixel correlations might lead to a more compact architecture. Despite all the challenges, deep normalizing flows pave the way to fully harnessing information from the ongoing and forthcoming spectroscopic surveys (4MOST, SDSS-V), allowing us to finally “reach the stars.”

References

- Bond-Taylor, S., Leach, A., Long, Y., and Willcocks, C. G. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *arXiv e-prints*, art. arXiv:2103.04922, March 2021.
- Buder, S., Sharma, S., Kos, J., Amarsi, A. M., Nordlander, T., Lind, K., Martell, S. L., Asplund, M., Bland-Hawthorn, J., Casey, A. R., de Silva, G. M., D’Orazi, V., Freeman, K. C., Hayden, M. R., Lewis, G. F., Lin, J., Schlesinger, K. J., Simpson, J. D., Stello, D., Zucker, D. B., Zwitter, T., Beeson, K. L., Buck, T., Casagrande, L., Clark, J. T., Čotar, K., da Costa, G. S., de Grijs, R., Feuillet, D., Horner, J., Kafle, P. R., Khanna, S., Kobayashi, C., Liu, F., Montet, B. T., Nandakumar, G., Nataf, D. M., Ness, M. K., Spina, L., Tepper-García, T., Ting, Y.-S., Traven, G., Vogrinčič, R., Wittenmyer, R. A., Wyse, R. F. G., Žerjal, M., and GALAH Collaboration. The GALAH+ survey: Third data release. *MNRAS*, 506(1):150–201, September 2021. doi: 10.1093/mnras/stab1242.
- Carollo, D., Freeman, K., Beers, T. C., Placco, V. M., Tumlinson, J., and Martell, S. L. Carbon-enhanced Metal-poor Stars: CEMP-s and CEMP-no Subclasses in the Halo System of the Milky Way. *ApJ*, 788(2):180, June 2014. doi: 10.1088/0004-637X/788/2/180.
- Cunha, K., Smith, V. V., Hasselquist, S., Souto, D., Shetrone, M. D., Allende Prieto, C., Bizyaev, D., Frinchaboy, P., García-Hernández, D. A., Holtzman, J., Johnson, J. A., Jónsson, H., Majewski, S. R., Mészáros, S., Nidever, D., Pinsonneault, M., Schiavon, R. P., Sobeck, J., Skrutskie, M. F., Zamora, O., Zasowski, G., and Fernández-Trincado, J. G. Adding the s-Process Element Cerium to the APOGEE Survey: Identification and Characterization of Ce II Lines in the H-band Spectral Window. *ApJ*, 844(2):145, August 2017. doi: 10.3847/1538-4357/aa7beb.
- Dai, B. and Seljak, U. Translation and Rotation Equivariant Normalizing Flow (TRENFlow) for Optimal Cosmological Analysis. *arXiv e-prints*, art. arXiv:2202.05282, February 2022.
- de Mijolla, D., Ness, M. K., Viti, S., and Wheeler, A. J. Disentangled Representation Learning for Astronomical Chemical Tagging. *ApJ*, 913(1):12, May 2021. doi: 10.3847/1538-4357/abece1.
- Dinh, L., Krueger, D., and Bengio, Y. NICE: Non-linear Independent Components Estimation. *arXiv e-prints*, art. arXiv:1410.8516, October 2014.
- Dinh, L., Sohl-Dickstein, J., and Bengio, S. Density estimation using Real NVP. *arXiv e-prints*, art. arXiv:1605.08803, May 2016.
- Durkan, C., Bekasov, A., Murray, I., and Papamakarinos, G. Neural Spline Flows. *arXiv e-prints*, art. arXiv:1906.04032, June 2019.
- Götberg, Y., de Mink, S. E., Groh, J. H., Leitherer, C., and Norman, C. The impact of stars stripped in binaries on the integrated spectra of stellar populations. *AAP*, 629:A134, September 2019. doi: 10.1051/0004-6361/201834525.
- Hasselquist, S., Shetrone, M., Cunha, K., Smith, V. V., Holtzman, J., Lawler, J. E., Allende Prieto, C., Beers, T. C., Chojnowski, D., Fernández-Trincado, J. G., García-Hernández, D. A., Hearty, F. R., Majewski, S. R., Pereira, C. B., Placco, V. M., Villanova, S., and Zamora, O. Identification of Neodymium in the Apogee H-Band Spectra. *ApJ*, 833(1):81, December 2016. doi: 10.3847/1538-4357/833/1/81.
- Holtzman, J. A., Shetrone, M., Johnson, J. A., Allende Prieto, C., Anders, F., Andrews, B., Beers, T. C., Bizyaev, D., Blanton, M. R., Bovy, J., Carrera, R., Chojnowski, S. D., Cunha, K., Eisenstein, D. J., Feuillet, D., Frinchaboy, P. M., Galbraith-Frew, J., García Pérez, A. E., García-Hernández, D. A., Hasselquist, S., Hayden, M. R., Hearty, F. R., Ivans, I., Majewski, S. R., Martell, S., Meszaros, S., Muna, D., Nidever, D., Nguyen, D. C., O’Connell, R. W., Pan, K., Pinsonneault, M., Robin, A. C., Schiavon, R. P., Shane, N., Sobeck, J., Smith, V. V., Troup, N., Weinberg, D. H., Wilson, J. C., Wood-Vasey, W. M., Zamora, O., and Zasowski, G. Abundances, Stellar Parameters, and Spectra from the SDSS-III/APOGEE Survey. *AJ*, 150(5):148, November 2015. doi: 10.1088/0004-6256/150/5/148.
- Jofré, P., Heiter, U., Soubiran, C., Blanco-Cuaresma, S., Worley, C. C., Pancino, E., Cantat-Gaudin, T., Magrini,

- L., Bergemann, M., González Hernández, J. I., Hill, V., Lardo, C., de Laverny, P., Lind, K., Masseron, T., Montes, D., Mucciarelli, A., Nordlander, T., Recio Blanco, A., Sobeck, J., Sordo, R., Sousa, S. G., Tabernero, H., Valenari, A., and Van Eck, S. Gaia FGK benchmark stars: Metallicity. *AAP*, 564:A133, April 2014. doi: 10.1051/0004-6361/201322440.
- Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, December 2014.
- Kingma, D. P. and Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL <https://proceedings.neurips.cc/paper/2018/file/d139db6a236200b21cc7f752979132d0-Paper.pdf>.
- Kurucz, R. L. ATLAS12, SYNTHÉ, ATLAS9, WIDTH9, et cetera. *Memorie della Societa Astronomica Italiana Supplementi*, 8:14, January 2005.
- Kurucz, R. L. ATLAS9: Model atmosphere program with opacity distribution functions. Astrophysics Source Code Library, record ascl:1710.017, October 2017.
- Kurucz, R. L. and Avrett, E. H. Solar Spectrum Synthesis. I. A Sample Atlas from 224 to 300 nm. *SAO Special Report*, 391, May 1981.
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Allende Prieto, C., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., Carrera, R., Chojnowski, S. D., Cunha, K., Epstein, C., Fitzgerald, G., García Pérez, A. E., Hearty, F. R., Henderson, C., Holtzman, J. A., Johnson, J. A., Lam, C. R., Lawler, J. E., Maseman, P., Mészáros, S., Nelson, M., Nguyen, D. C., Nidever, D. L., Pinsonneault, M., Shetrone, M., Smee, S., Smith, V. V., Stolberg, T., Skrutskie, M. F., Walker, E., Wilson, J. C., Zasowski, G., Anders, F., Basu, S., Beland, S., Blanton, M. R., Bovy, J., Brownstein, J. R., Carlberg, J., Chaplin, W., Chiappini, C., Eisenstein, D. J., Elsworth, Y., Feuillet, D., Fleming, S. W., Galbraith-Frew, J., García, R. A., García-Hernández, D. A., Gillespie, B. A., Girardi, L., Gunn, J. E., Hasselquist, S., Hayden, M. R., Hekker, S., Ivans, I., Kinemuchi, K., Klaene, M., Mahadevan, S., Mathur, S., Mosser, B., Muna, D., Munn, J. A., Nichol, R. C., O’Connell, R. W., Parejko, J. K., Robin, A. C., Rocha-Pinto, H., Schultheis, M., Serenelli, A. M., Shane, N., Silva Aguirre, V., Sobeck, J. S., Thompson, B., Troup, N. W., Weinberg, D. H., and Zamora, O. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). *AJ*, 154(3):94, September 2017. doi: 10.3847/1538-3881/aa784d.
- Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., and Zasowski, G. The Cannon: A data-driven approach to Stellar Label Determination. *ApJ*, 808(1):16, July 2015. doi: 10.1088/0004-637X/808/1/16.
- Nielsen, D., Jaini, P., Hooeboom, E., Winther, O., and Welling, M. SurVAE Flows: Surjections to Bridge the Gap between VAEs and Flows. *arXiv e-prints*, art. arXiv:2007.02731, July 2020.
- Nordlander, T., Amarsi, A. M., Lind, K., Asplund, M., Barklem, P. S., Casey, A. R., Collet, R., and Leenaarts, J. 3D NLTE analysis of the most iron-deficient star, SMSS0313-6708. *AAP*, 597:A6, January 2017. doi: 10.1051/0004-6361/201629202.
- O’Brian, T., Ting, Y.-S., Fabbro, S., Yi, K. M., Venn, K., and Bialek, S. Cycle-StarNet: Bridging the Gap between Theory and Data by Leveraging Large Data Sets. *ApJ*, 906(2):130, January 2021. doi: 10.3847/1538-4357/abca96.
- Price-Jones, N. and Bovy, J. The dimensionality of stellar chemical space using spectra from the Apache Point Observatory Galactic Evolution Experiment. *MNRAS*, 475(1):1410–1425, March 2018. doi: 10.1093/mnras/stx3198.
- Ting, Y.-S. and Weinberg, D. H. How Many Elements Matter? *arXiv e-prints*, art. arXiv:2102.04992, February 2021.
- Ting, Y.-S., Conroy, C., Rix, H.-W., and Cargile, P. The Payne: Self-consistent ab initio Fitting of Stellar Spectra. *ApJ*, 879(2):69, July 2019. doi: 10.3847/1538-4357/ab2331.