1. How do we improve synthetic models?

Enormous efforts have been made over the past several decades to enhance our models of stellar atmospheres and the spectra they produce. Collecting more atomic data, comparing simulations to reference stars, and removing simplifying assumptions (e.g. local thermodynamic equilibrium, LTE), are common methods of improving the quality of synthetic spectra. Despite all these efforts there still remains a gap between our models and reality, due to instrumental signatures and unknown physics, which leads to imprecise and biased stellar parameter estimates. Based on results from Cycle-GAN (**Figure 1**) and motivated by this problem, we developed **Cycle-StarNet** to bridge this gap in an automated and unsupervised way; synthetic spectra can inherit features unique to the observed domain and help us learn what was missing in our models.

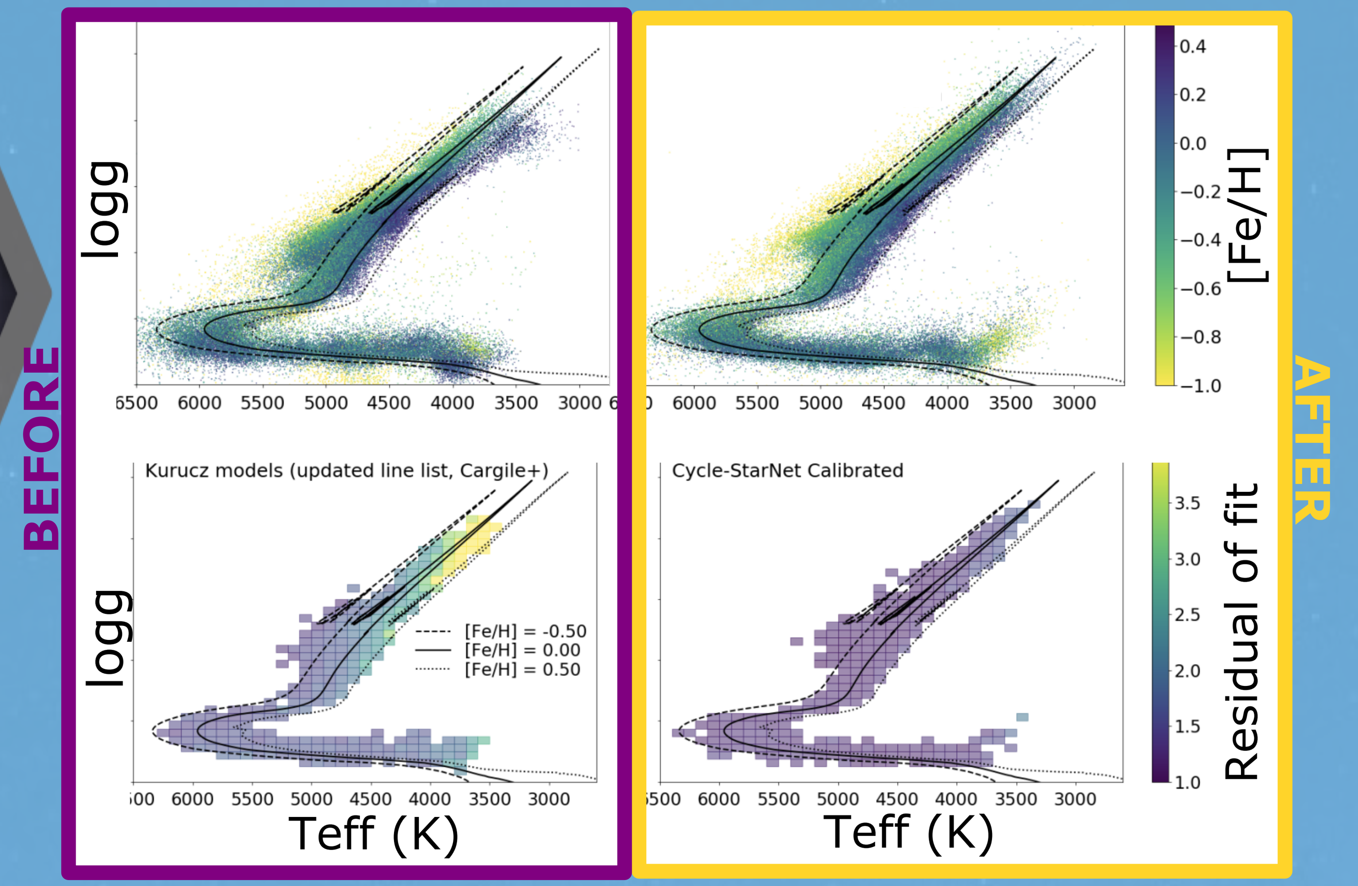

Figure 2: Stellar parameters from a best-fit analysis of synthetic and APOGEE spectra, before and after the synthetic spectra were mapped through Cycle-StarNet to the observed domain. The estimates align better with isochrones, and the residuals between synthetic spectra and real spectra are dramatically reduced, especially for cool giants and dwarfs.

2. A new way forward: Automated domain adaptation using machine learning

We developed a new method, **Cycle-StarNet,** which harnesses recent advances in ML to learn from the synthetic and observed domains, simultaneously learning the features both in common with and exclusive to each domain. After being trained, synthetic spectra can be mapped to the observed domain -- a process which augments the spectra with features not present in the synthetic domain, e.g. unknown or poorly modeled absorption features and instrumental broadening. **Figure 2** shows the effects on stellar parameter estimates of APOGEE spectra after the transformation of synthetic spectra to the observed domain, whereas **Figure 3** shows the increased precision in the tracks of chemical of abundance predictions.

Figure 1: What is domain adaptation?

An ML algorithm, Cycle-GAN, was trained to reconstruct images in two domains: horses and zebras. Once trained, it could map (i.e. adapt) one domain to the other, turning a horse into a zebra. In our case we want synthetic spectra to look more like observed spectra, i.e. adapt the synthetic to the observed domain. This process can be manual (e.g. by studying non-LTE effects) or automatic (similar to Cycle-GAN).

Credit: Yuan-Sen Ting

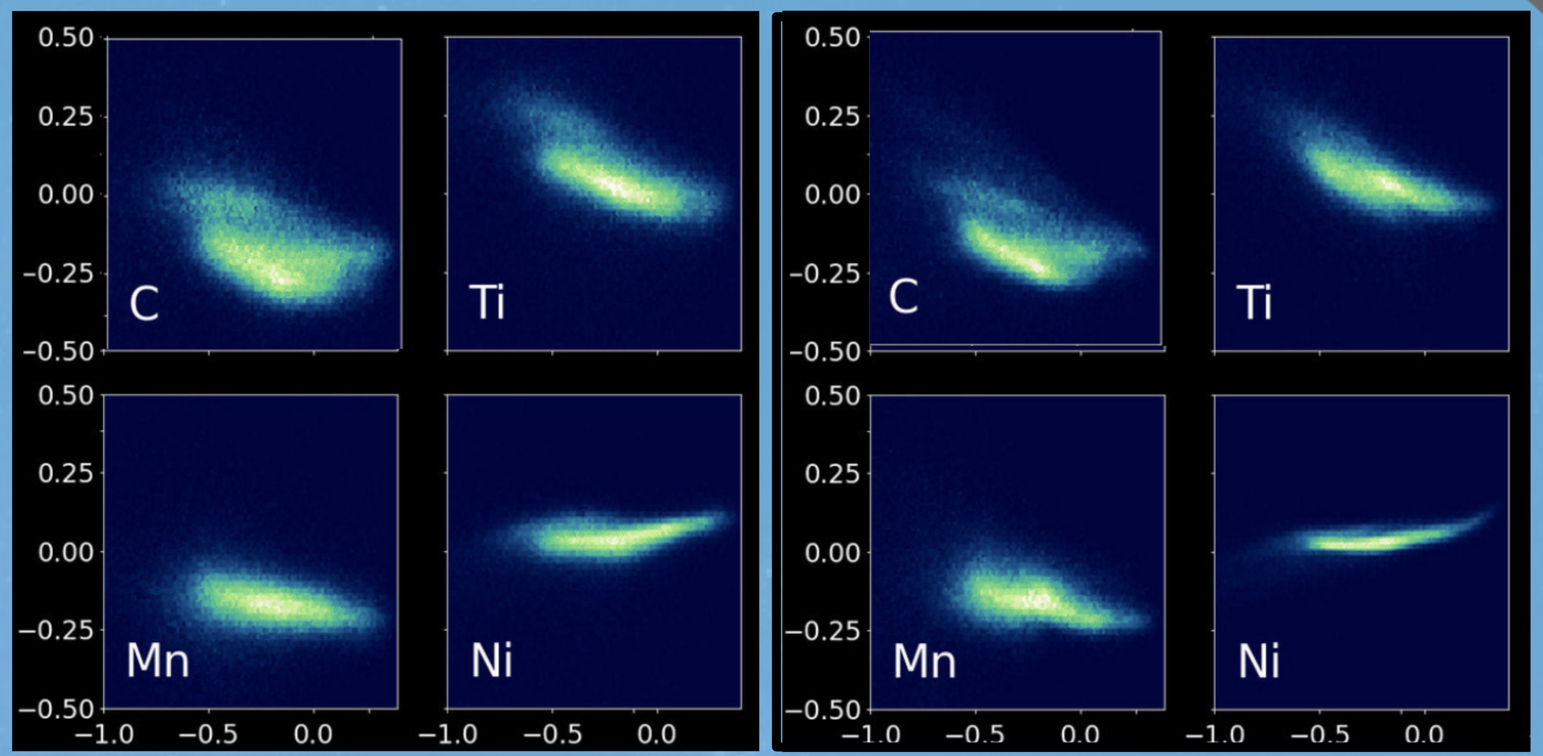

Figure 3: Chemical abundances from a best-fit analysis of synthetic and APOGEE spectra, before and after the synthetic spectra were mapped through Cycle-StarNet to the observed domain. The chemical tracks become more precise, revealing a more nuanced substructure.

Cycle-StarNet

Decreasing the gap between synthetic and real data... automated at last

Spencer Bialek, Teaghan O'Briain, Yuan-Sen Ting, Sebastien Fabbro, Kwang Moo Yi, Kim Venn

3. Conclusions

We have presented an open-source ML framework, **Cycle-StarNet**, that is capable of drastically improving the quality of synthetic data given a large sample of observed data. Estimates of stellar parameters become more realistic, and chemical abundance predictions are more precise. Our method can also be used to discover sources of unknown absorption features, helping to learn unknown physics.

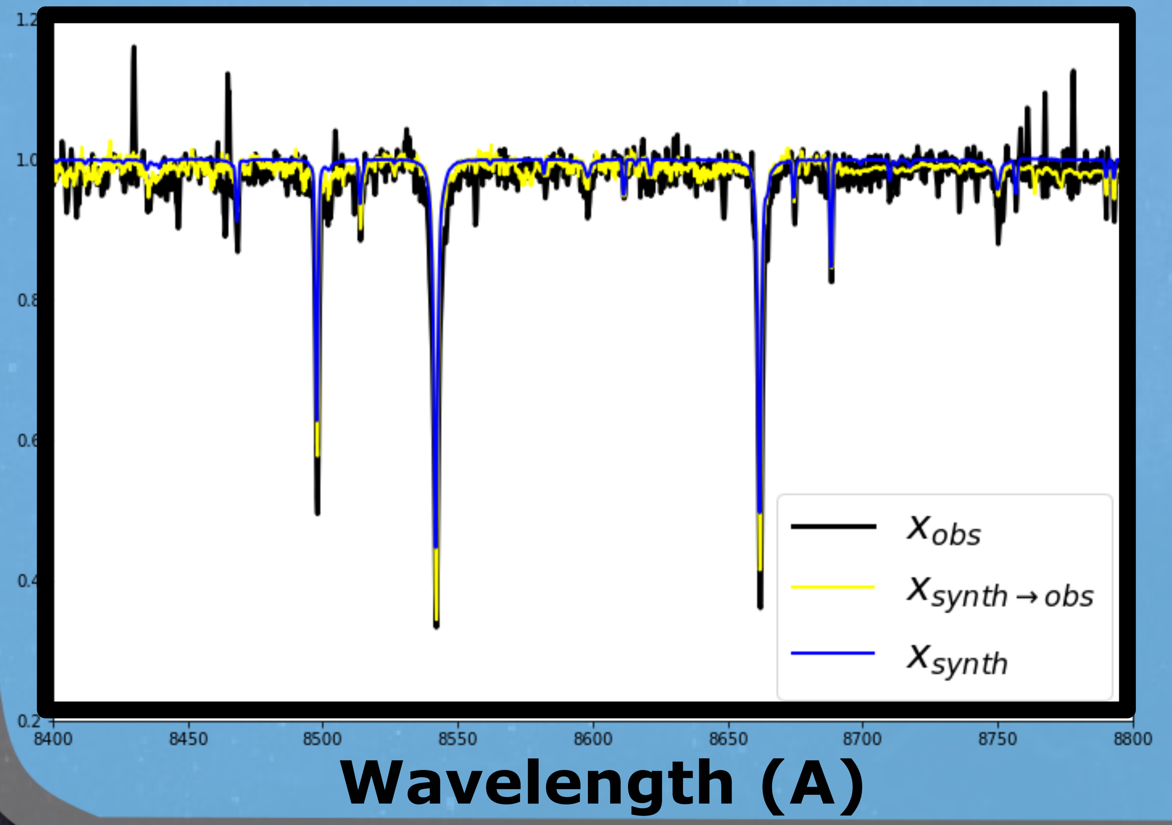

Figure 4: In this test, we show directly how Cycle-StarNet is able to better match a synthetic spectrum to an observed spectrum from the metal-poor Pristine Inner Galactic Survey. When the best-matched synthetic spectrum (blue line) is mapped through Cycle-StarNet (yellow line) the continuum level, absorption features, and noise profile more closely resemble the observed spectrum (black line).

University of Victoria

https://github.com/teaghan/Cycle_SN