cecilia: A Machine Learning Pipeline for Measuring Metal Abundances in Polluted White Dwarfs

Mariona **Badenas-Agusti** (mbadenas@mit.edu; MIT), Andrew Vanderburg (MIT), Javier Viaña (MIT),
Patrick Dufour (Université de Montréal), Simon Blouin (University of Victoria), Siyi Xu (Gemini/NOIRLab),
Lizhou Sha (University of Wisconsin-Madison), Kishalay De (MIT), Susana Hoyos (MIT)

Motivation

Many white dwarfs are "polluted" with icy or rocky material from their progenitor's planetary systems[1,2] (**Fig. 1**). Spectroscopic observations of these polluted white dwarfs provide valuable insights into the geology and chemistry of exoplanetary bodies. However, current spectral analysis techniques (**Fig. 2**) of these stellar remnants involve manual, time-intensive, and iterative work, as well as substantial computational resources and proprietary atmospheric models.[3,4,5,6] These limitations make it difficult to perform large-scale population studies of polluted white dwarfs, therefore hindering our understanding of the properties of their accreted material. Here we present `cecilia`, a machine-learning (ML) spectral modeling code[7,8] designed to measure chemical abundances in polluted white dwarfs.

Methodology

Target: Helium-dominated WDs with 10,000≤T_{eff} ≤ 20,000 K.

Parameters: 27 stellar "labels" (T_{eff}, log(g), 11 elemental abundances* relative to He; 14 elemental abundances† scaled to chondritic values) + Spectrum-specific parameters (RV shift, jitter).

ML Architecture: 1 Autoencoder + 2 Fully Connected Neural Networks. End-to-End Training: 3 weeks (windows of 200Å).

Training Data: About 23,000 labels with their corresponding high-resolution (R~50,000) atmosphere models.[3,5,6]

Optimization: Multi-observation fit of ML-interpolated model spectra to data via non-linear Least Squares Regression (`mpfit`) and a Bayesian MCMC (`edmcmc`).

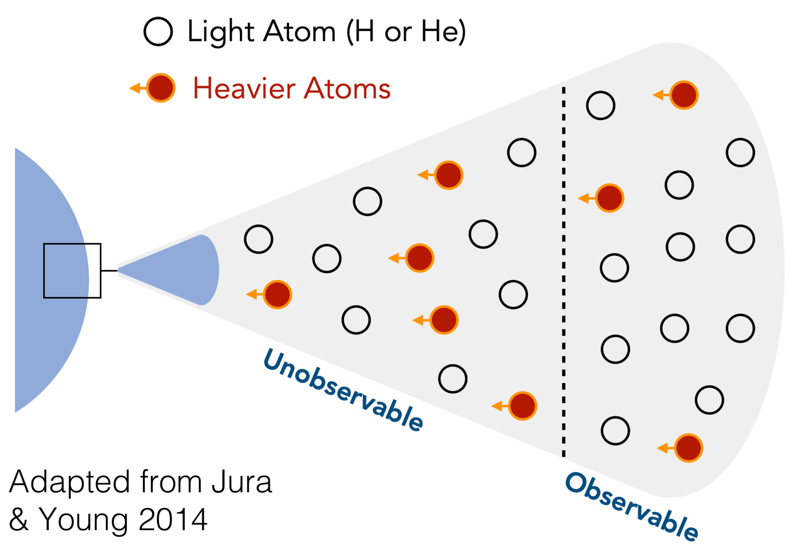
*H, Ca, Mg, Fe, O, Si, Ti, Be, Cr, Mn, Ni

†C, N, Li, Na, Al, P, S, Cl, Ar, K, Sc, V, Co, Cu

Results

- Efficiency of neural-network-powdered interpolation allows for fast generation of polluted white dwarf atmosphere models (≪ 1 sec, 1 GPU).
- Sensitivity study of synthetic and real data yields an accuracy (≲ 0.15 dex) similar to that of conventional methods for up to 10 metals (**Fig. 3**).
- Performed the first ML-based compositional study of 5 previously un-studied He-rich polluted white dwarfs with SDSS and Keck spectra (**Fig. 4**).
- One of these WDs may have accreted an O-rich chondritic body, with potential Si and Fe depletion. Need (UV) data to refine composition (**Fig. 5**).

Fig. 1



Adapted from Jura & Young 2014

Fig. 2

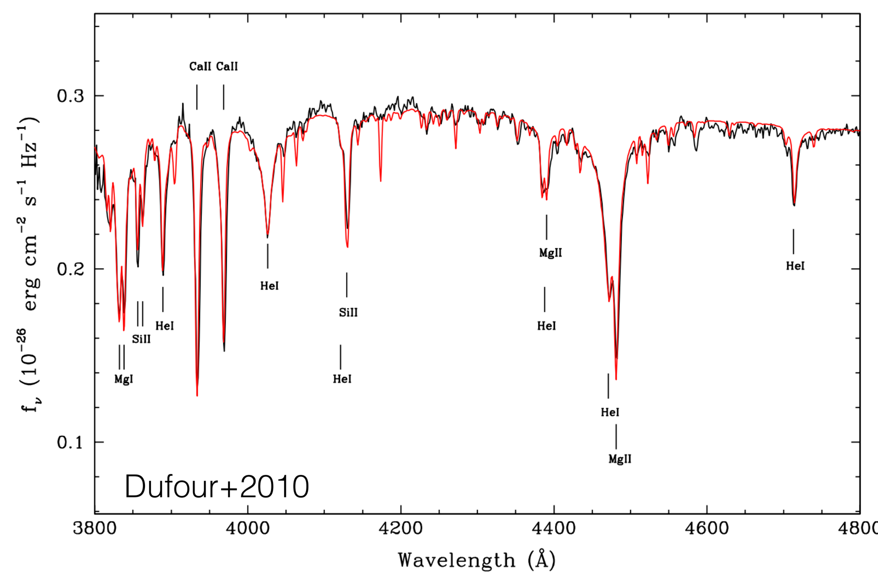
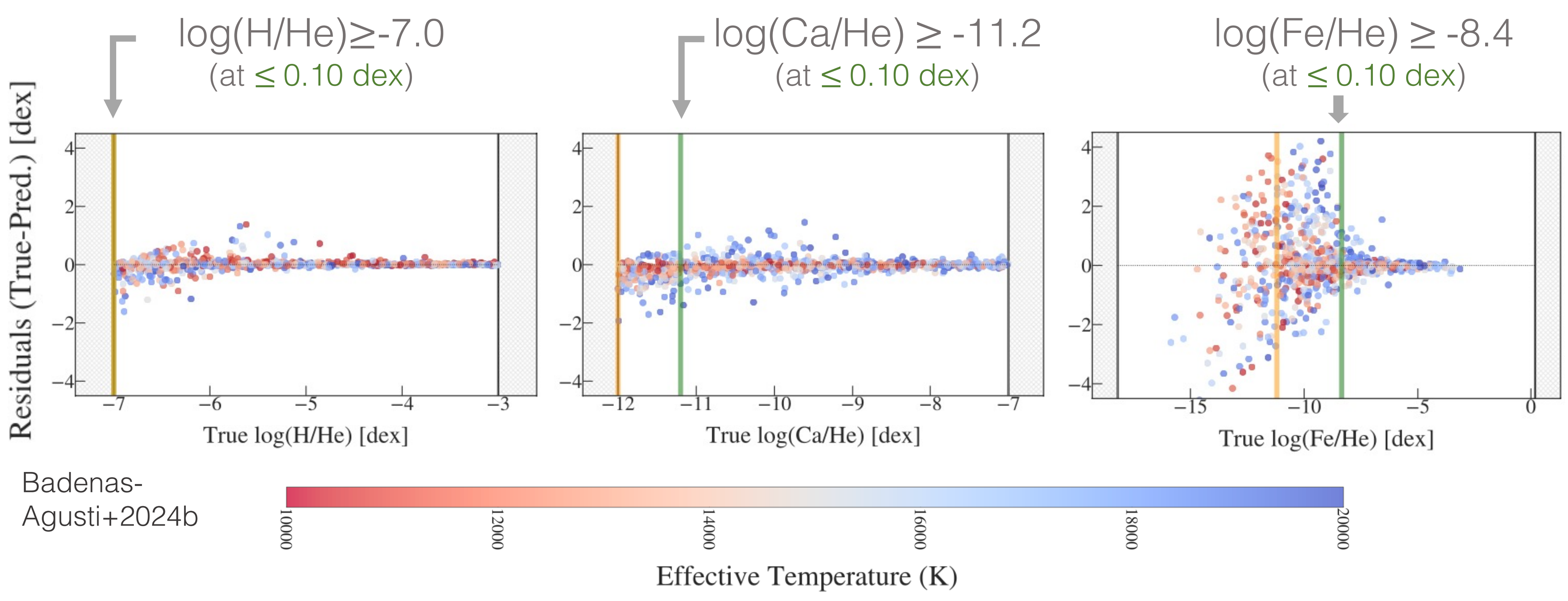


Fig. 3



Badenas-Agusti+2024b

Fig. 4

Badenas-Agusti, in prep.

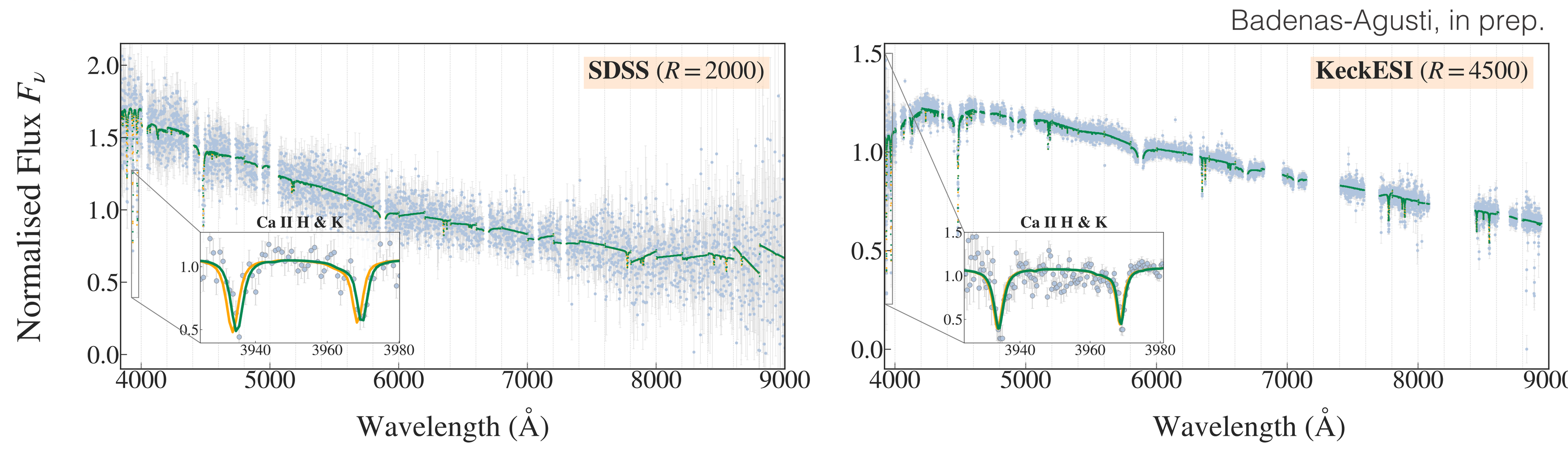
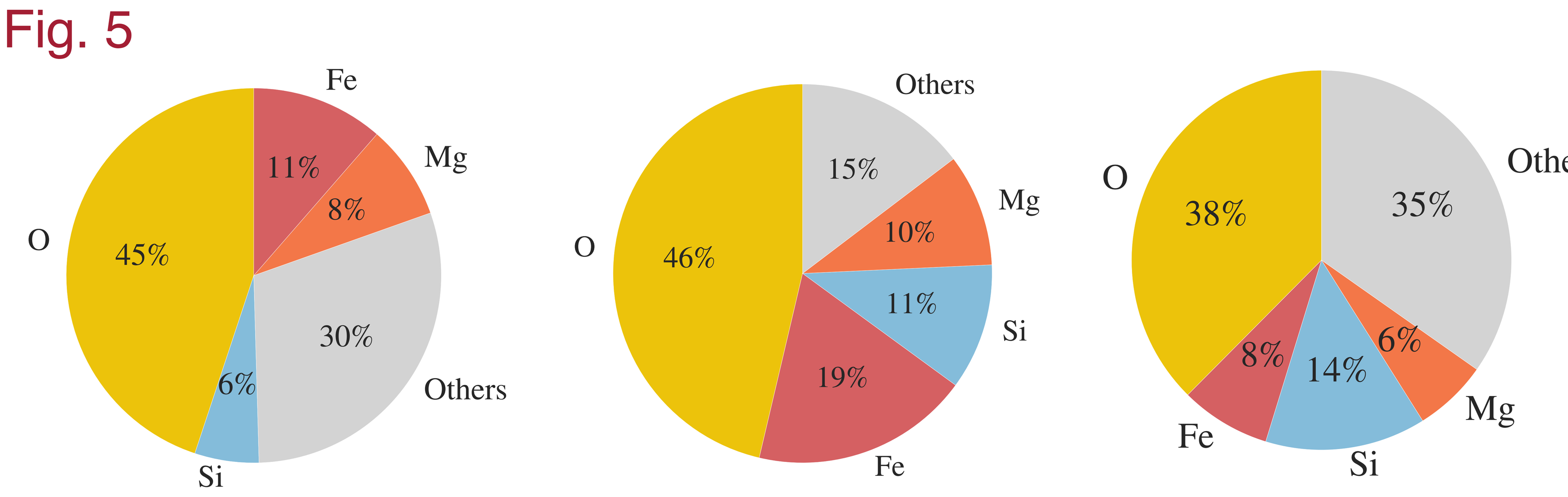


Fig. 5



Polluted WD (Steady-State) (Badenas-Agusti, in prep.)

CI Chondrites (Alexander+2019a,b)

Comet Halley (Jessberger+1988)

Conclusions

- `cecilia` leverages ML to rapidly generate atmosphere models of polluted white dwarfs (≪ 1 sec) and accurately (≲0.15 dex) measure up to 10 metal abundances from real spectroscopic observations.
- Combined with upcoming data from massive multiplexed spectroscopic surveys (e.g. DESI, SDSS-V, 4MOST, WEAVE), `cecilia` can help unlock large studies of white dwarf pollution, hence paving the way to a statistical understanding of extrasolar geochemistry.

References

[1] Jura 2003, ApJ
[2] Jura & Young 2014, Annu Rev Earth Planet Sci, *42*
[3] Dufour+2007, ApJ, *663*, 1291
[4] Koester 2010, Sait, *81*, 921
[5] Blouin+2018a, ApJ, *863*, 184
[6] Blouin+2018b, ApJ, *867*, 161
[7] Ting+2019, ApJ, *879*, 2
[8] Chandra+2020, MNRAS, *497*, 3