HARPSpec: Using Machine Learning to Find Chemical Compositions of Stars from Their Spectra

Joseph Hand, Ian J.M. Crossfield

University of Kansas, ExoLab

March 23, 2023

Organization

Background

Terminology

- Species: Distinct element or molecule

Terminology

- Species: Distinct element or molecule
- Abundances: Ratio of number of some species to hydrogen
  - Usually expressed on a log-10 scale
  - Here expressed in relation to the Solar abundance

Terminology

- Species: Distinct element or molecule
- Abundances: Ratio of number of some species to hydrogen
    - Usually expressed on a log-10 scale
    - Here expressed in relation to the Solar abundance
- Labels: Quantitative property of a star (e.g. abundances, surface temperature)

Terminology

- Species: Distinct element or molecule
- Abundances: Ratio of number of some species to hydrogen
  - Usually expressed on a log-10 scale
  - Here expressed in relation to the Solar abundance
- Labels: Quantitative property of a star (e.g. abundances, surface temperature)
- Model: Function to be trained in machine learning

Terminology

- Species: Distinct element or molecule
- Abundances: Ratio of number of some species to hydrogen
  - Usually expressed on a log-10 scale
  - Here expressed in relation to the Solar abundance
- Labels: Quantitative property of a star (e.g. abundances, surface temperature)
- Model: Function to be trained in machine learning
- Weight: Machine learning model degree of freedom

What is HARPSpec?

- Will be machine learning too to derive stellar abundances from spectra.

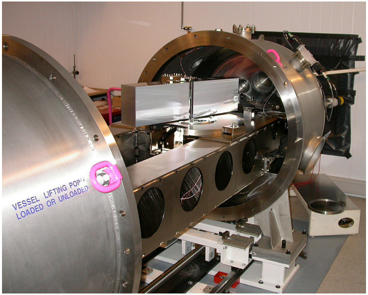


Image courtesy of ESO

What is HARPSpec?

- Will be machine learning too to derive stellar abundances from spectra.
- Will use publicly available spectra from HARPS (pictured)
    - Located at La Silla, Chile
    - High resolution, optical
    - Spectra of $\sim$ 3200 sun-like stars

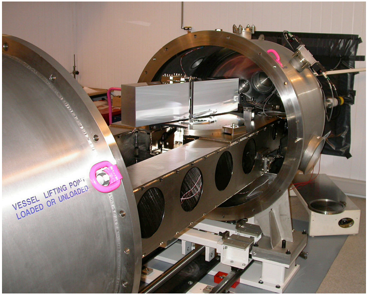


Image courtesy of ESO

What is HARPSpec?

- Will be machine learning too to derive stellar abundances from spectra.
- Will use publicly available spectra from HARPS (pictured)
    - Located at La Silla, Chile
    - High resolution, optical
    - Spectra of $\sim$ 3200 sun-like stars
- Will use the Cannon
    - Supervised machine learning library
    - Fast and efficient

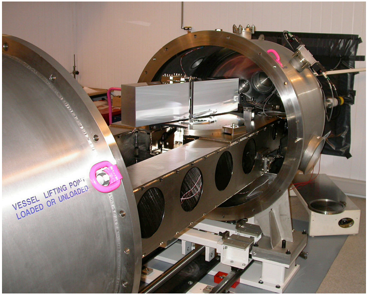


Image courtesy of ESO

Why Stellar Compositions?

Why obtain stellar compositions?

Why Stellar Compositions?

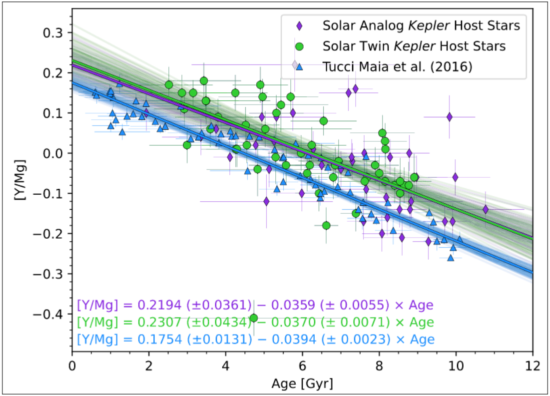
Why obtain stellar compositions?

- Better understand stars

Why Stellar Compositions?

Why obtain stellar compositions?

- Better understand stars
  - Evolution

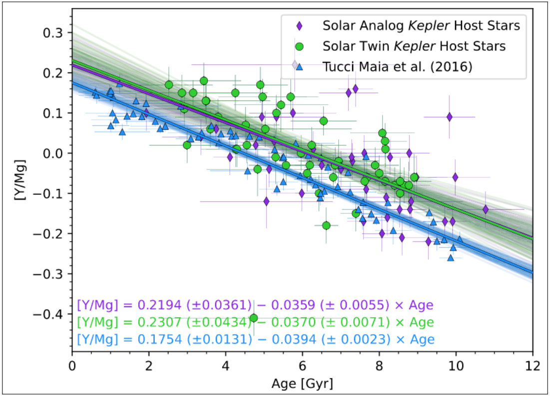


Berger et al. (2022)

Why Stellar Compositions?

Why obtain stellar compositions?

- Better understand stars
  - Evolution
  - Modelling

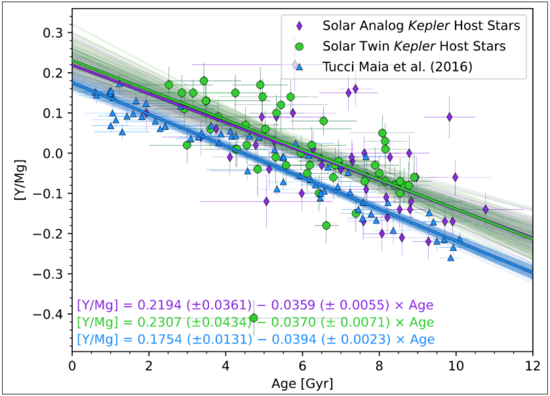


Berger et al. (2022)

Why Stellar Compositions?

Why obtain stellar compositions?

- Better understand stars
  - Evolution
  - Modelling
- Better understand exoplanets

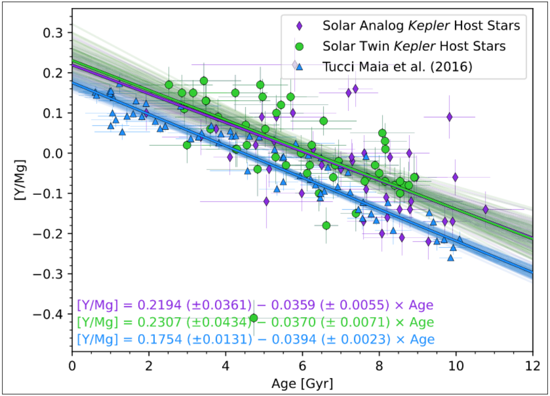


Berger et al. (2022)

Why Stellar Compositions?

Why obtain stellar compositions?

- Better understand stars
  - Evolution
  - Modelling
- Better understand exoplanets
  - Composition

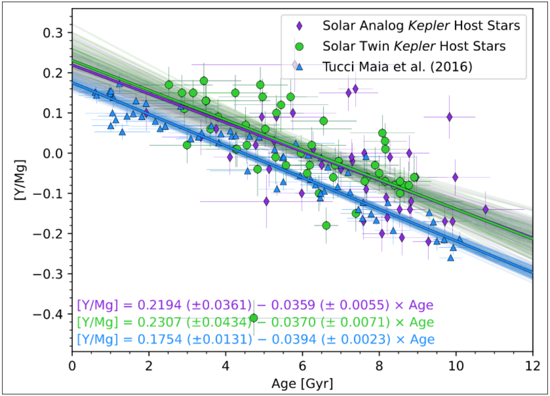


Berger et al. (2022)

Why Stellar Compositions?

Why obtain stellar compositions?

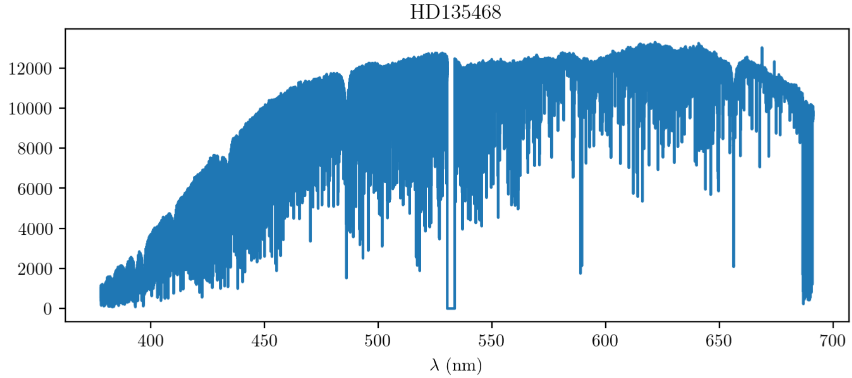
- Better understand stars
  - Evolution
  - Modelling
- Better understand exoplanets
  - Composition
  - Formation



Berger et al. (2022)

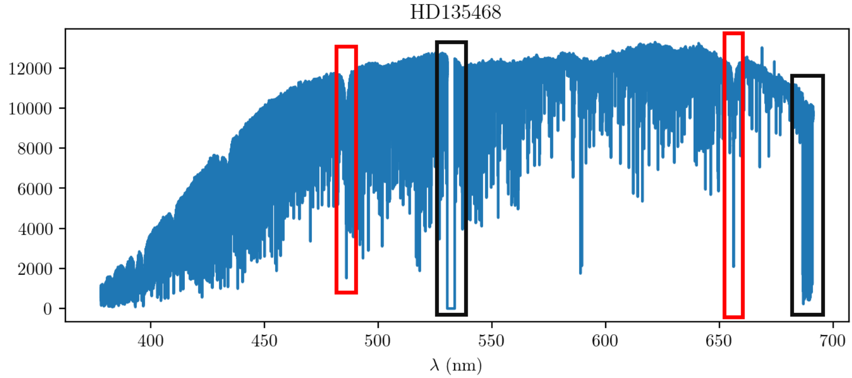
KU ExoLab

Relationship Between Spectra and Abundances



- Overall shape due to instrument sensitivity

Relationship Between Spectra and Abundances

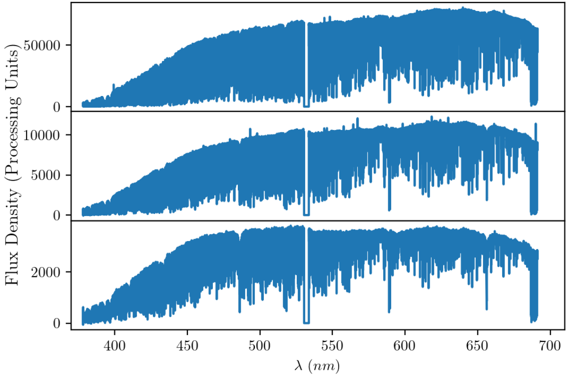


- Overall shape due to instrument sensitivity
- Absorption lines (Red)
- A few artifacts (Black)

Data

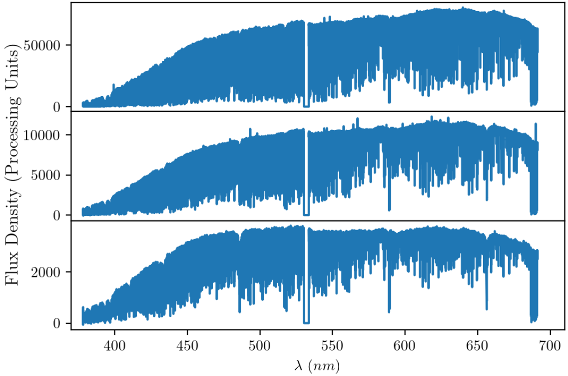
The Spectra

- 3200 spectra from $\sim$ 650 stars



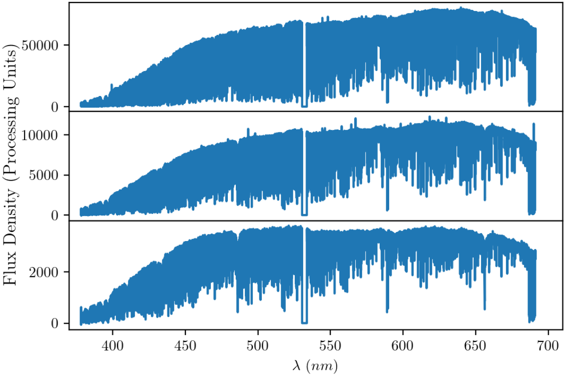
The Spectra

- 3200 spectra from $\sim$ 650 stars
- 5 spectra per star (if available)



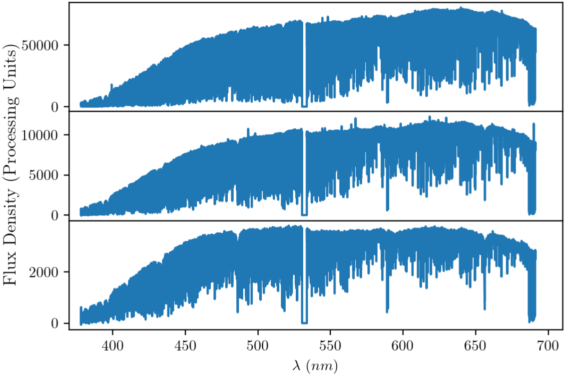
The Spectra

- 3200 spectra from $\sim$ 650 stars
- 5 spectra per star (if available)
- Optical spectra, 400*nm* - 700*nm*



The Spectra

- 3200 spectra from $\sim$ 650 stars
- 5 spectra per star (if available)
- Optical spectra, 400*nm* - 700*nm*
- All spectra are publicly available



The Labels

- 20 labels total

The Labels

- 20 labels total
- T_{eff}, $\log g$, and $v \sin i$

The Labels

- 20 labels total
- T_{eff}, $\log g$, and $v \sin i$
- 17 abundances (Fe, Na, Mg, Al, Si, Ca, V, Mn, Co, O, Ni, C, ScI, TiI, CrI, YII, S)

The Labels

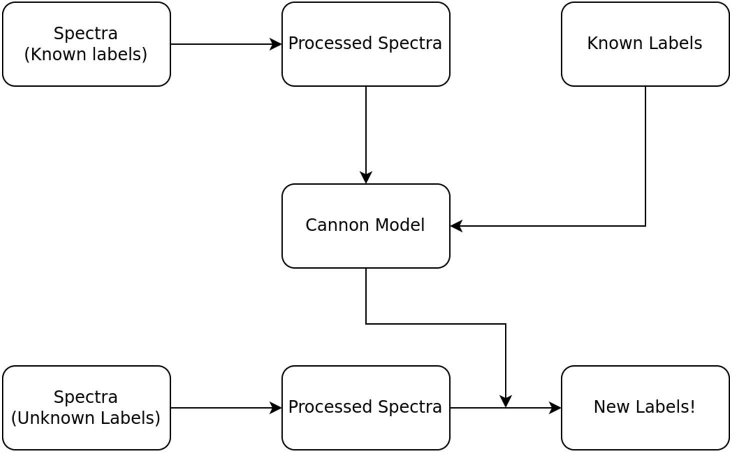
- 20 labels total
- T_{eff}, $\log g$, and $v \sin i$
- 17 abundances (Fe, Na, Mg, Al, Si, Ca, V, Mn, Co, O, Ni, C, ScI, TiI, CrI, YII, S)

Abundances from:

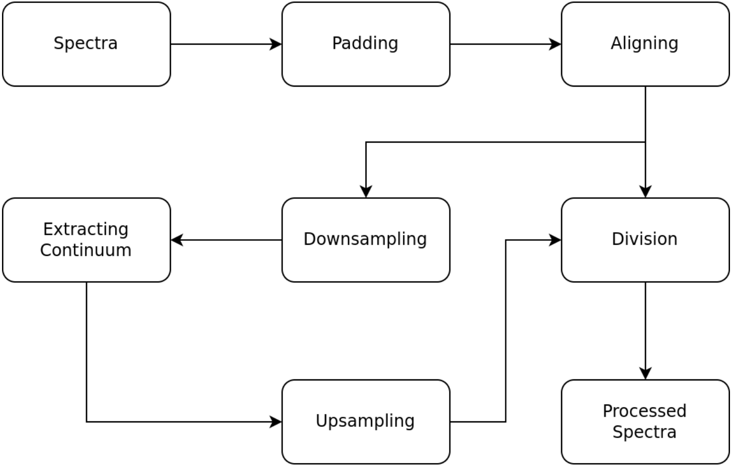
- Adibekyan et al. (2012)
- Bertran et al. (2015)
- Costa Silva et al. (2020)
- Delgado Mena et al. (2010, 2017, 2021)
- Suarez Andres et al. (2017)

HARPSpec

Overview of HARPSpec



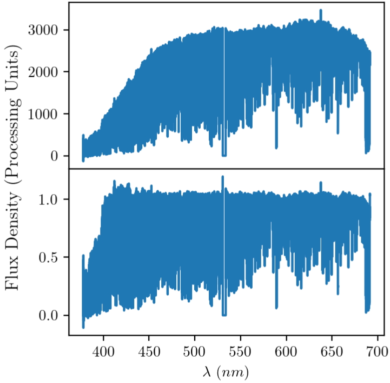
Spectral Processing



Continuum Normalization

Basic technique:

- Identify continuum pixels
- Fit curve to continuum
- Divide spectra by this fit



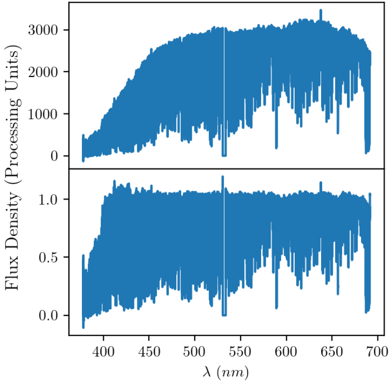
Continuum Normalization

Basic technique:

- Identify continuum pixels
- Fit curve to continuum
- Divide spectra by this fit

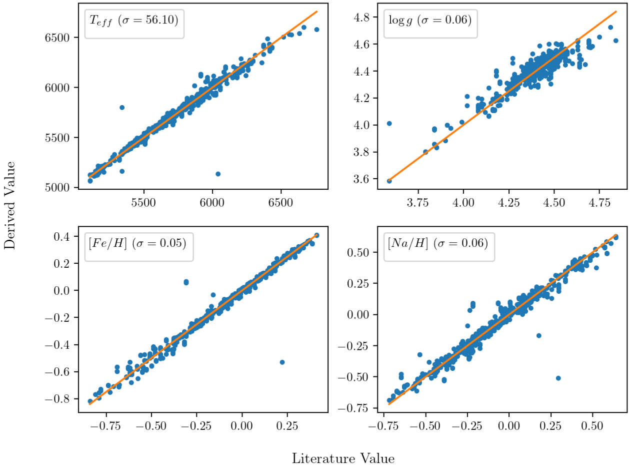
Some details:

- Spectral resolution is to large to do this quickly.
- First two steps done on downsampled spectra

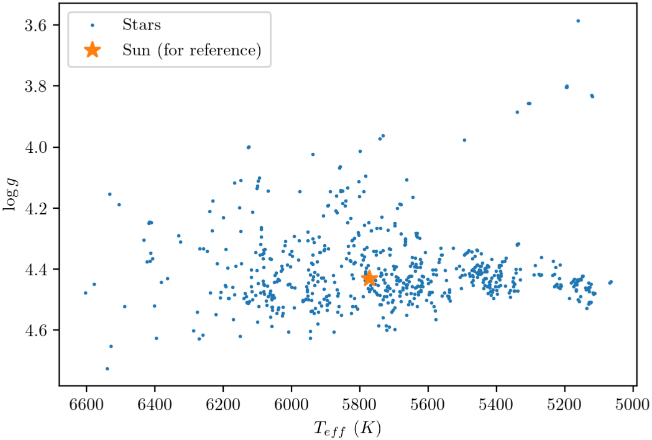


Current Results

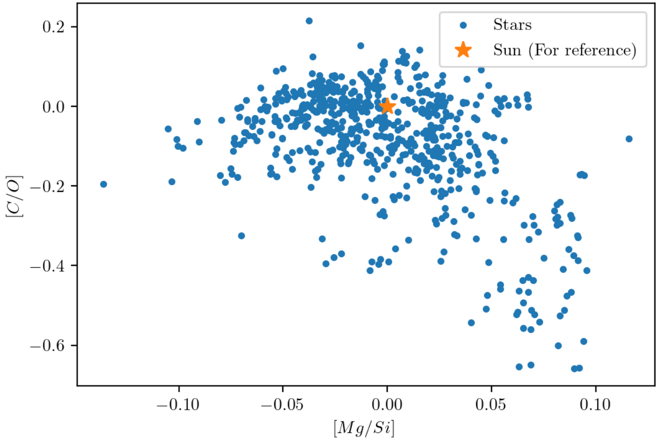
HARPSpec Performance



Temperature vs. Surface Gravity



[C/O] vs. [Mg/Si]


Stars
Sun (For reference)
0.2
0.0
−0.2
−0.4
−0.6
[C/O]
−0.10
−0.05
0.00
0.05
0.10
[Mg/Si]

Summary

Summary

- Much can be learned from stellar abundances

Summary

- Much can be learned from stellar abundances
- Stellar spectra contain imprint of abundances

Summary

- Much can be learned from stellar abundances
- Stellar spectra contain imprint of abundances
- Machine learning can be used to extract them with the Cannon

Summary

- Much can be learned from stellar abundances
- Stellar spectra contain imprint of abundances
- Machine learning can be used to extract them with the Cannon
- HARPSpec will use the Cannon on public HARPS Spectra

Summary

- Much can be learned from stellar abundances
- Stellar spectra contain imprint of abundances
- Machine learning can be used to extract them with the Cannon
- HARPSpec will use the Cannon on public HARPS Spectra
- Promising results so far

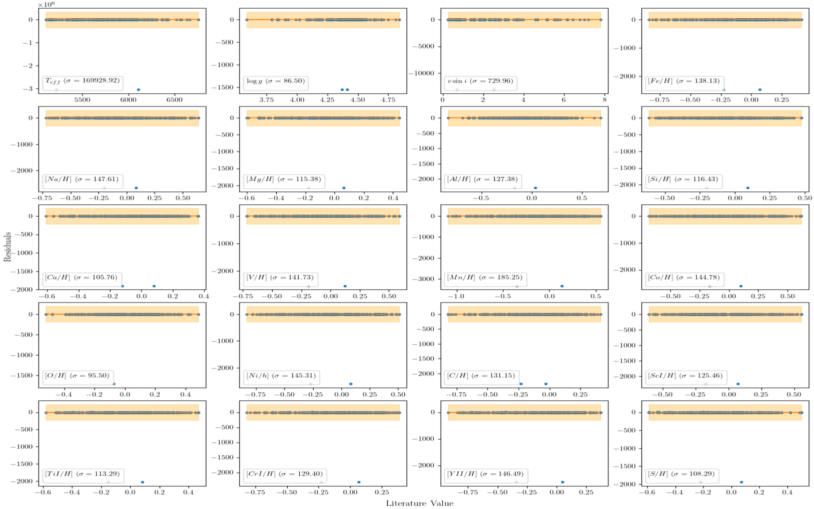
Summary

- Much can be learned from stellar abundances
- Stellar spectra contain imprint of abundances
- Machine learning can be used to extract them with the Cannon
- HARPSpec will use the Cannon on public HARPS Spectra
- Promising results so far
- Many possible ways to improve

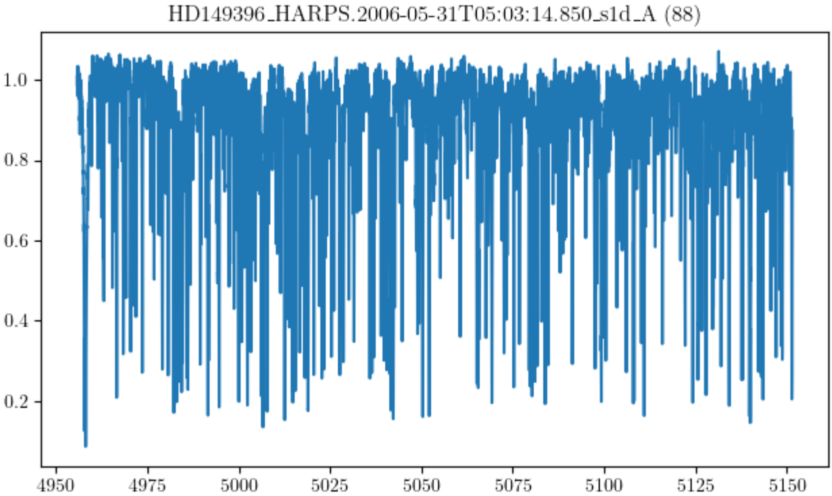
Summary

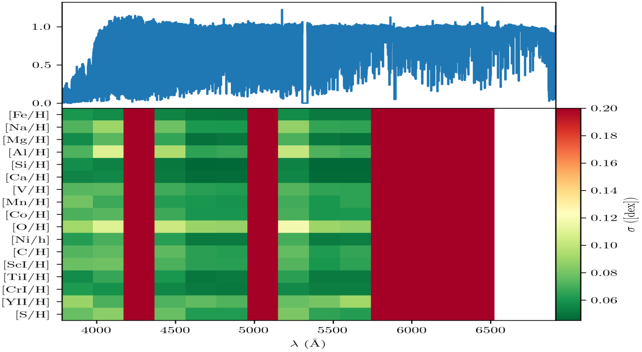
- Much can be learned from stellar abundances
- Stellar spectra contain imprint of abundances
- Machine learning can be used to extract them with the Cannon
- HARPSpec will use the Cannon on public HARPS Spectra
- Promising results so far
- Many possible ways to improve

Questions?



T_{eff} (σ = 169928.92)
log g (σ = 86.50)
v sin i (σ = 729.96)
[Fe/H] (σ = 138.13)
[Na/H] (σ = 147.61)
[Mg/H] (σ = 115.38)
[Al/H] (σ = 127.38)
[Si/H] (σ = 116.43)
[Ca/H] (σ = 105.76)
[V/H] (σ = 141.73)
[Mn/H] (σ = 185.25)
[Co/H] (σ = 144.78)
[O/H] (σ = 95.50)
[Ni/h] (σ = 145.31)
[C/H] (σ = 131.15)
[ScI/H] (σ = 125.46)
[TiI/H] (σ = 113.29)
[CrI/H] (σ = 129.40)
[YII/H] (σ = 146.49)
[S/H] (σ = 108.29)
Residuals
Literature Value

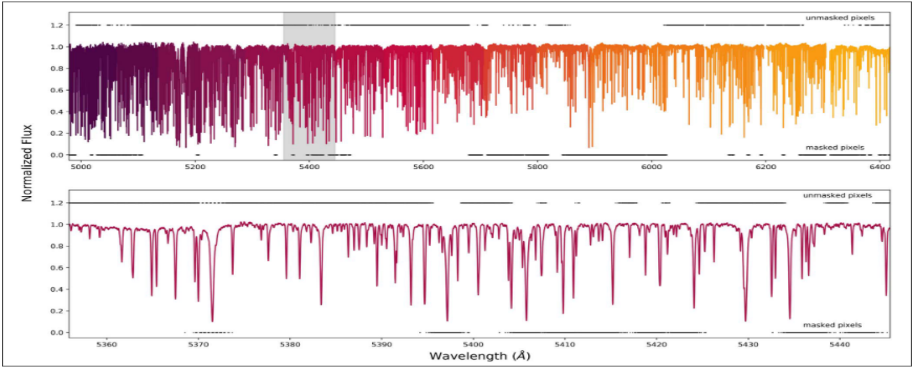

HD149396_HARPS.2006-05-31T05:03:14.850_s1d_A (88)
1.0
0.8
0.6
0.4
0.2
4950
4975
5000
5025
5050
5075
5100
5125
5150


1.0
0.5
0.0
[Fe/H]
[Na/H]
[Mg/H]
[Al/H]
[Si/H]
[Ca/H]
[V/H]
[Mn/H]
[Co/H]
[O/H]
[Ni/h]
[C/H]
[ScI/H]
[TiI/H]
[CrI/H]
[YII/H]
[S/H]
4000
4500
5000
5500
6000
6500
λ (Å)
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
σ ([dex])

Current Issues

- Resource Usage
  - 500GB RAM when training on the full dataset
  - Cutting spectra into 16 parts
  - This is not a permanent solution
  - Possible solutions include:
    - Binning down spectra
    - Getting more RAM
- Stability Issues
  - Occasional math errors with certain spectra
  - Very inaccurate labels for these spectra
  - Still unknown what is going wrong here

Telluric Masking

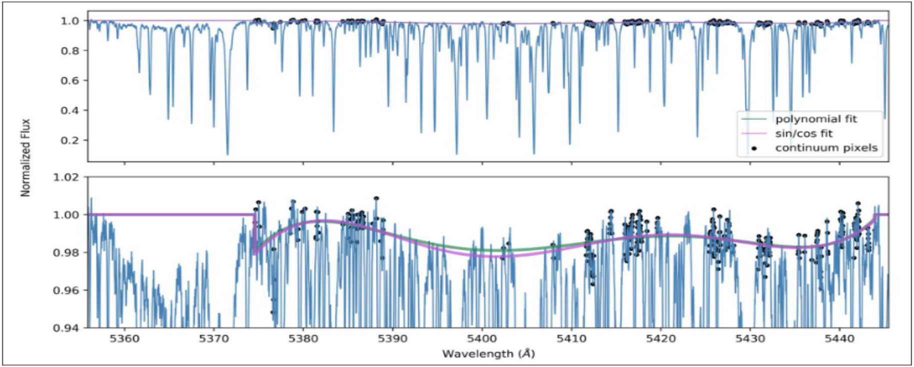


Rice & Brewer (2020)

- Ignore wavelengths affected by Earth's atmosphere
- Absorption lines from Earth's atmosphere confuse the Cannon

KU ExoLab

Data-Driven Continuum Normalization



Rice & Brewer (2020)

- Use the Cannon itself to identify continuum pixels
- Improve continuum-fitting algorithms (e.g. Ness et al. 2015)