AI/ML - applications to stars and stellar population studies

T. Sivarani, Indian Institute of Astrophysics

Outline

- Stellar classification
- Stellar parameterization
- Stellar abundance estimation
- Chemical tagging
- Future direction

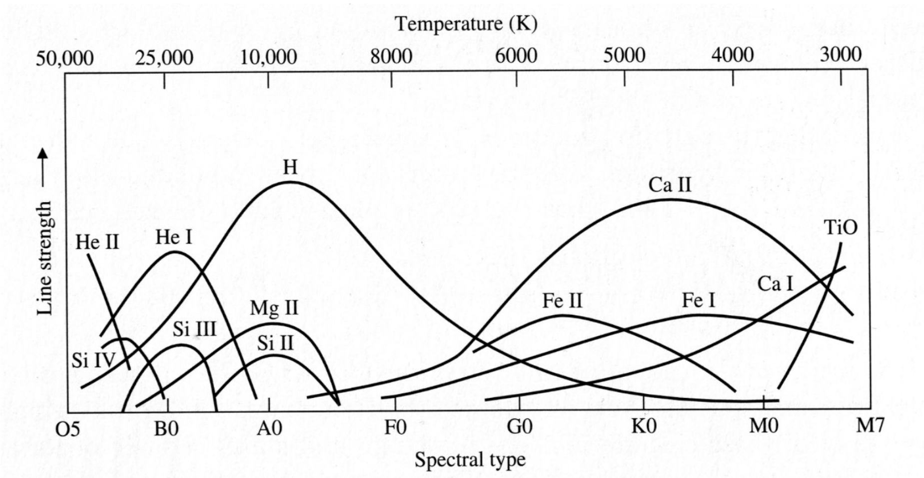
Why are we interested in stars?

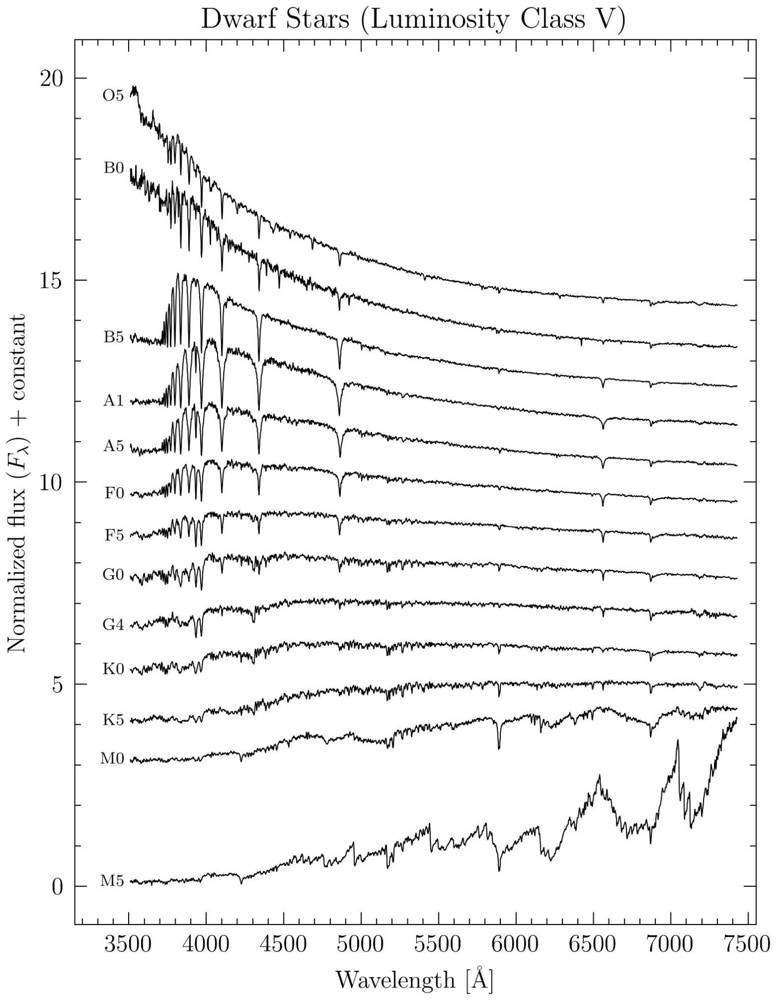
To understand the Sun, its formation and evolution and as part of the Galaxy

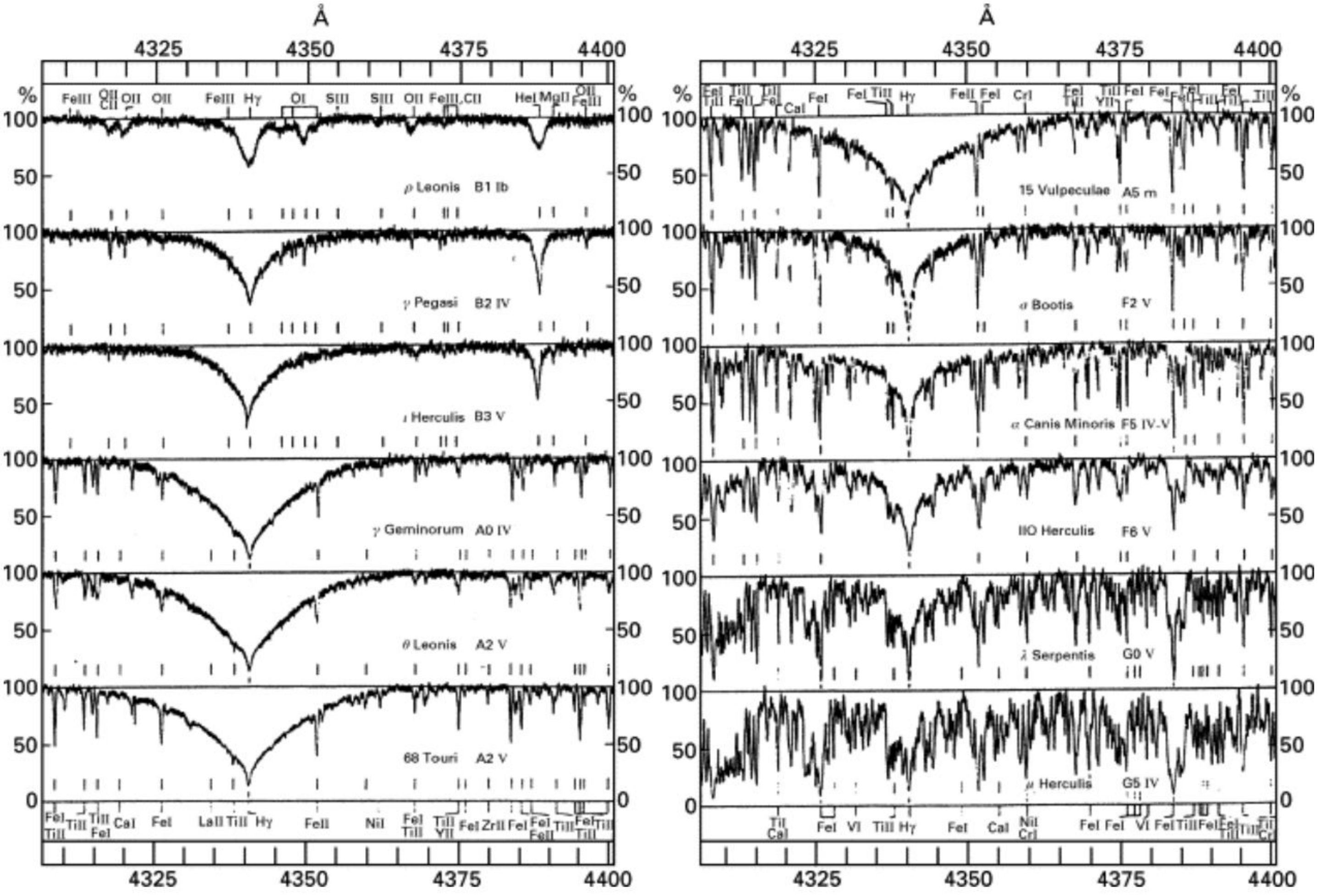
Use stars as to tool to study galaxies. ⇒ Stars contribute to the most of the visible wavelength emission of a galaxy

MK spectral classification -2D

Spectral class - OBAFGKMLT
Luminosity class - I,II,III,IV,V







Å
4325
4350
4375
4400
%
100
50
0
ρ Leonis B1 Ib
γ Pegasi B2 IV
ι Herculis B3 V
γ Geminorum A0 IV
θ Leonis A2 V
68 Tauri A2 V
15 Vulpeculae A5 m
α Bootis F2 V
α Canis Minoris F5 IV-V
110 Herculis F6 V
λ Serpentis G0 V
μ Herculis G5 IV
Hγ

Spectral classification

Autoencoders, CNN, Sharma et al. 2020

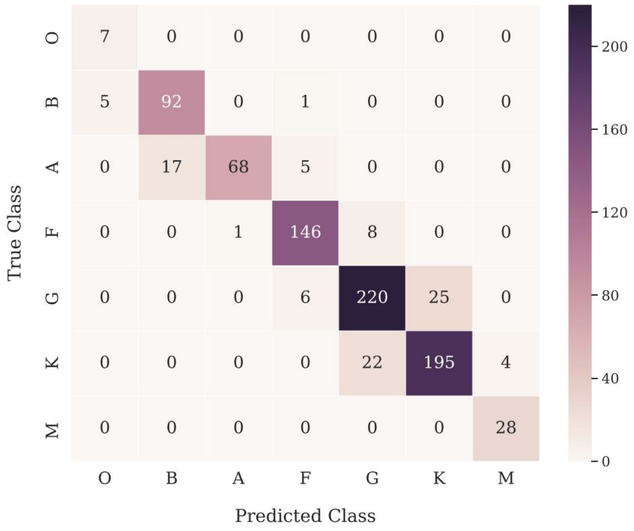


Figure 12. Confusion matrix for the main spectral classes computed for the CFLIB library using 1D CNN classification model.

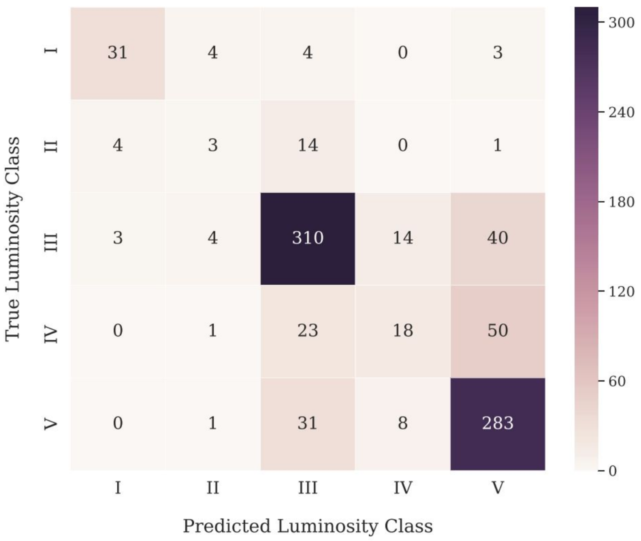
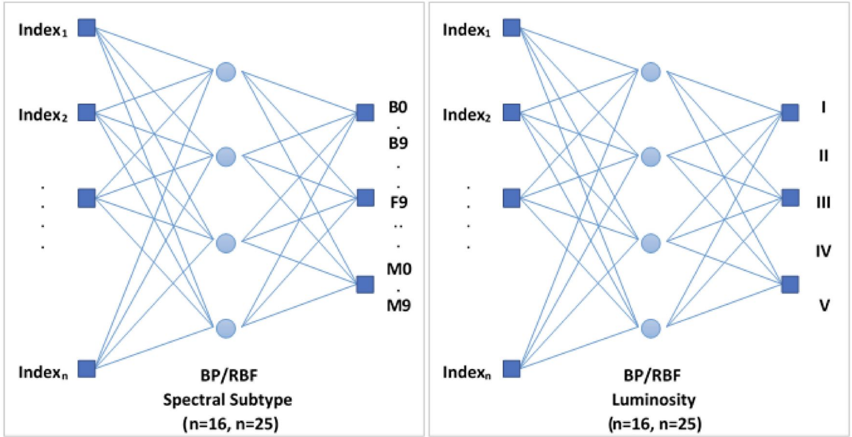
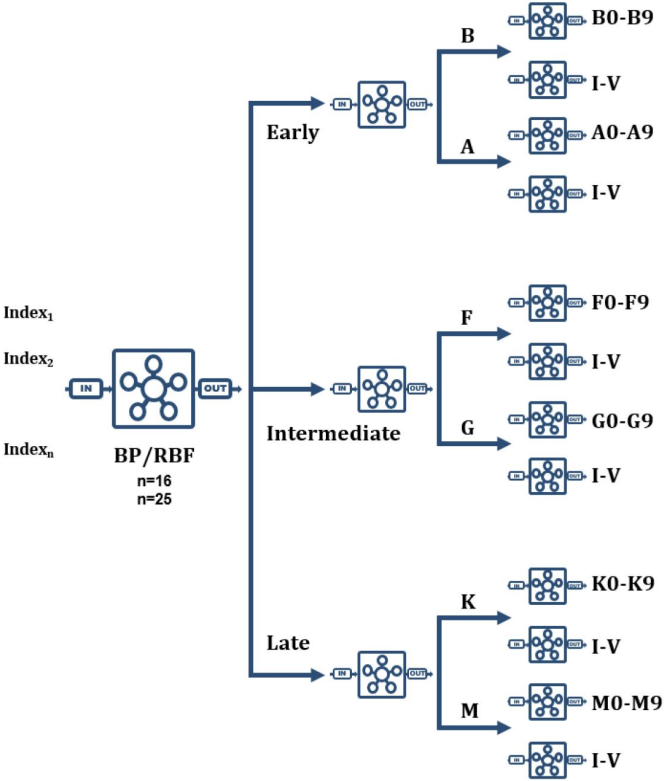


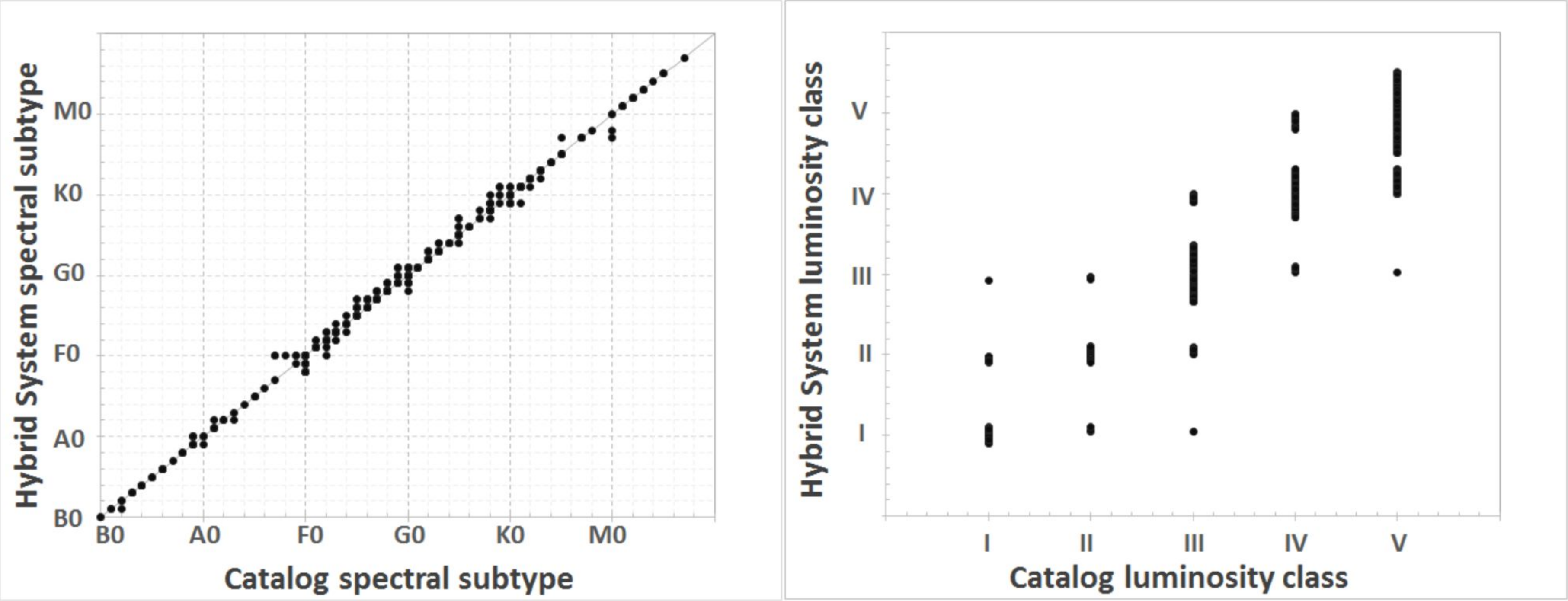
Figure 14. Confusion matrix for luminosity classification on CFLIB using 1D CNN.

Spectral classification - Hybrid approach - Dafonte et al.





Spectral classification - Hybrid approach - Dafonte et al.

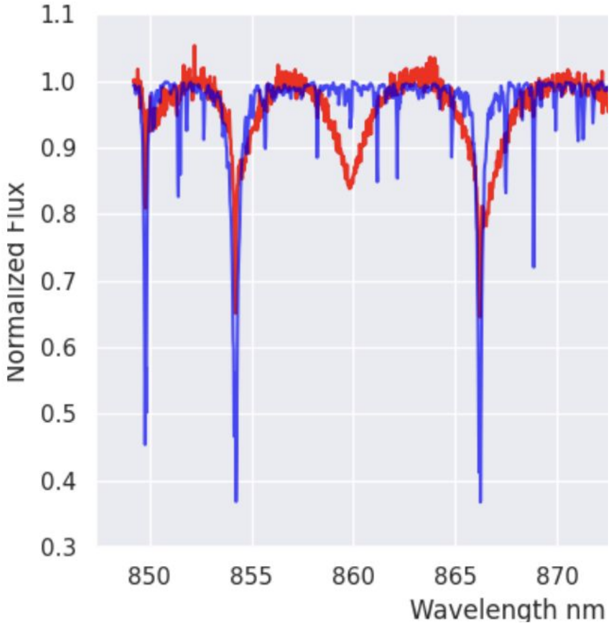


(a) (b)

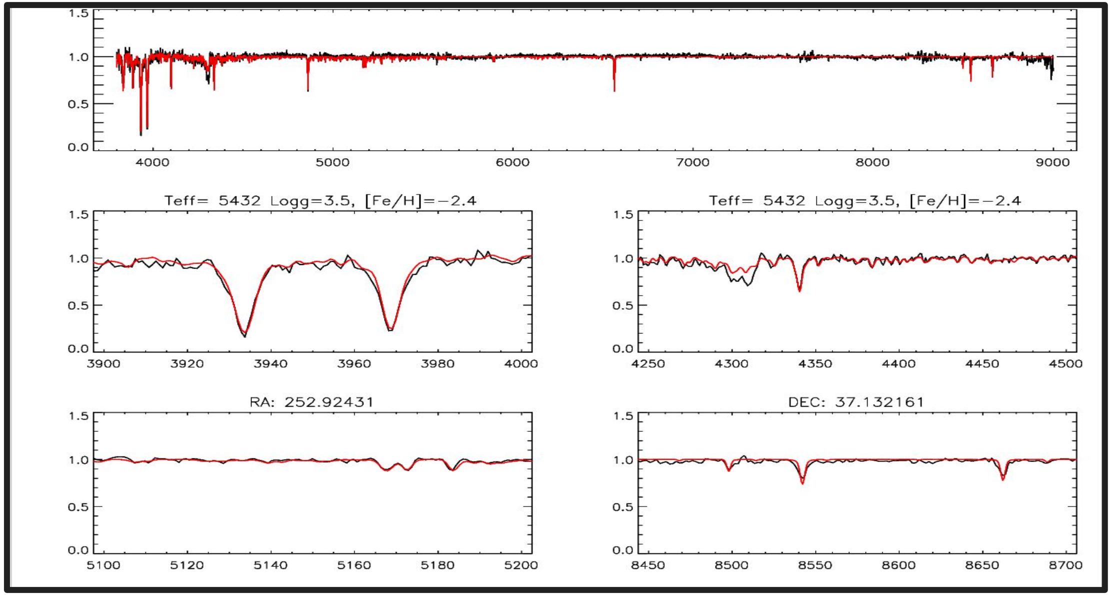
Spectral classification

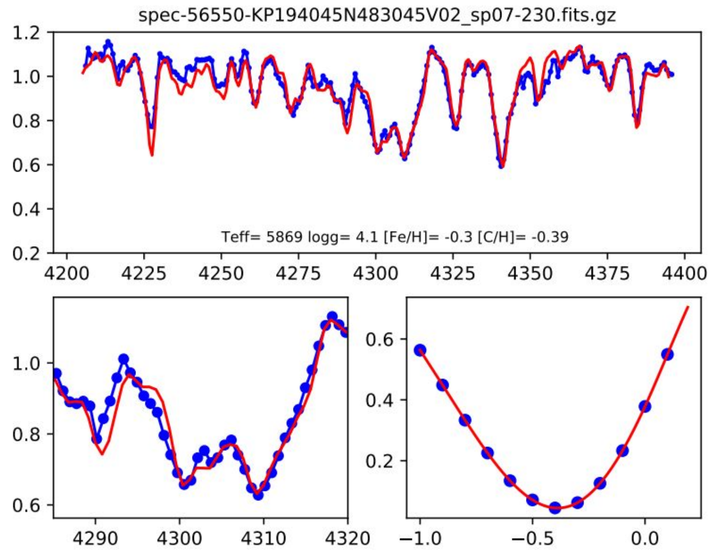
Matching with the set of benchmark training stars to get luminosity and distances

Finding rare objects - similarity index and twins - identifying Metal poor stars from Gaia-XP spectra – APOGEE, LAMOST is used to train Gaia XP spectra



Stellar parameterisation

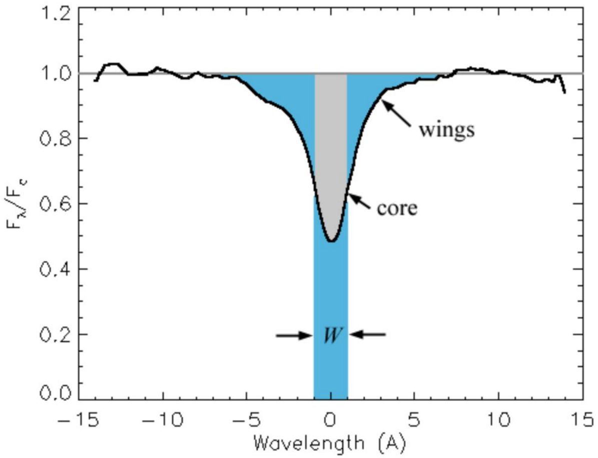


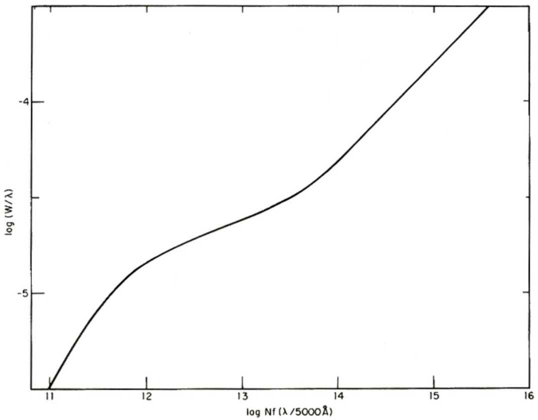


Estimation of physical parameters, Teff, logg, [Fe/H], matching with synthetic spectra using global optimizer Simulated Annealing. Issue: Degeneracy between Teff and logg.

Stellar parameterization suffers from degeneracies
Most surveys use empirical calibration to match high resolution spectroscopy or Trained on high resolution

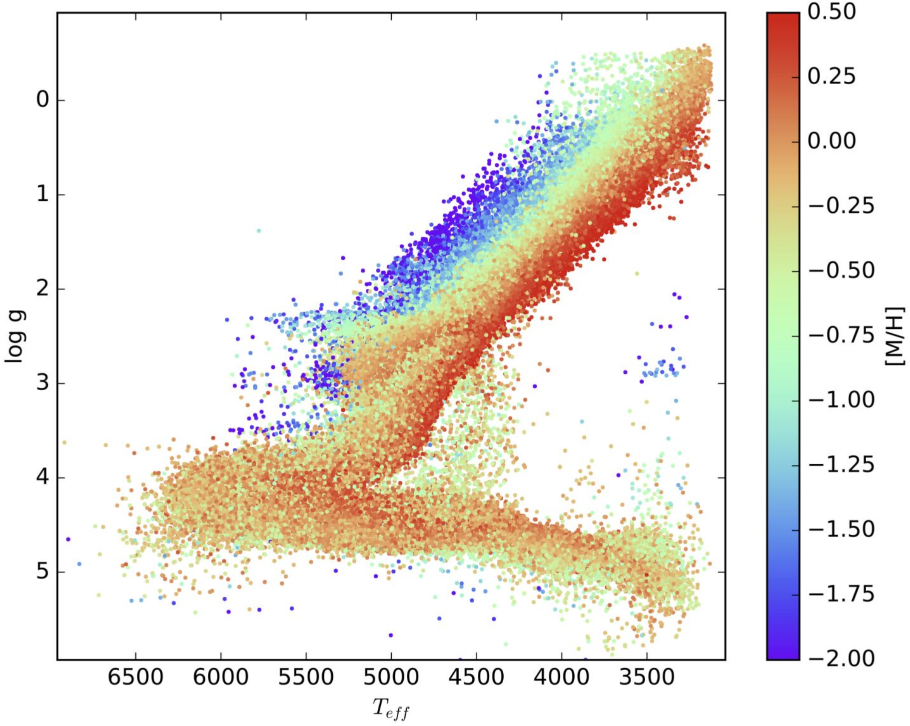
High resolution uses weak lines, low resolution uses strong lines, affected by microturbulence and line broadening, nonlinear dependency of optical depth. Carefully using wings of the lines would work better.



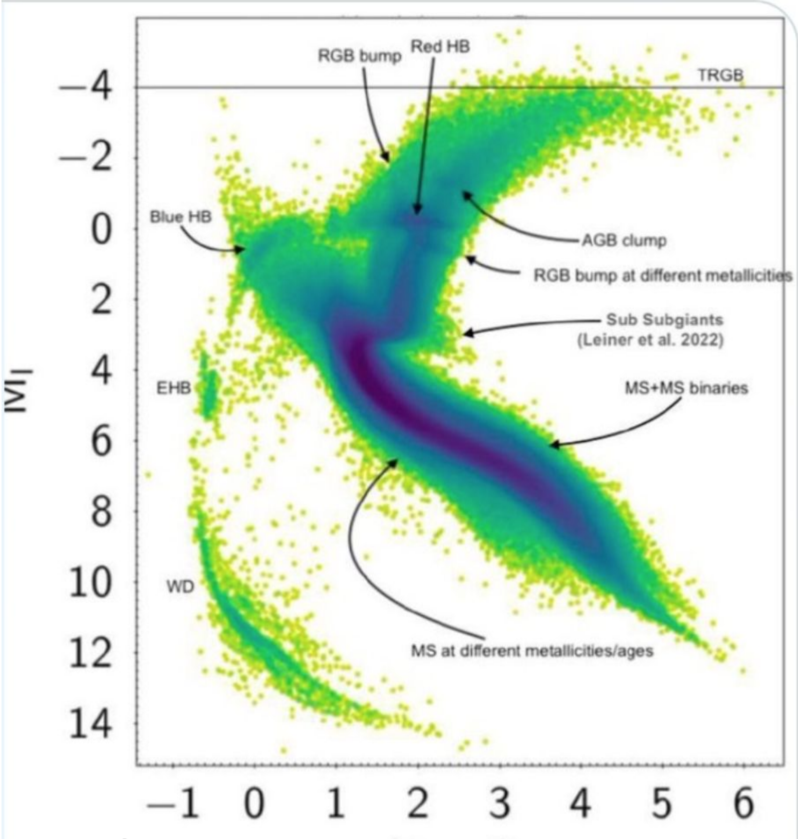


A general curve of growth for the Sun, where N is in atoms/cm^2. (Figure from Aller, *Atoms, Stars, and Nebulae*

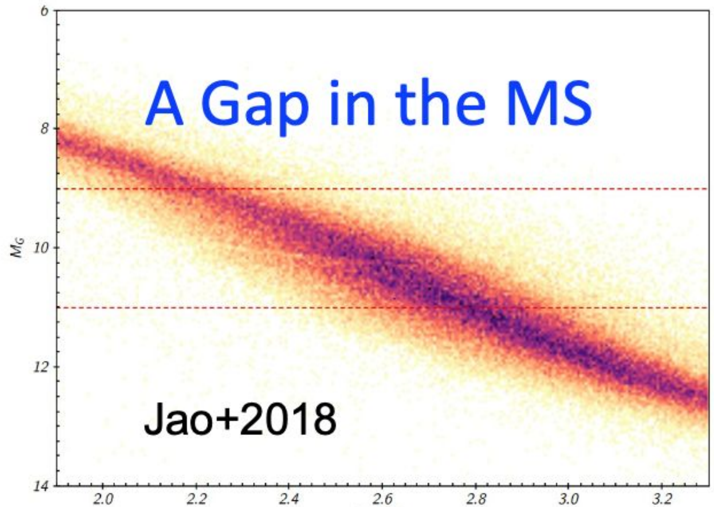
Comparison of observed and derived HR diagram

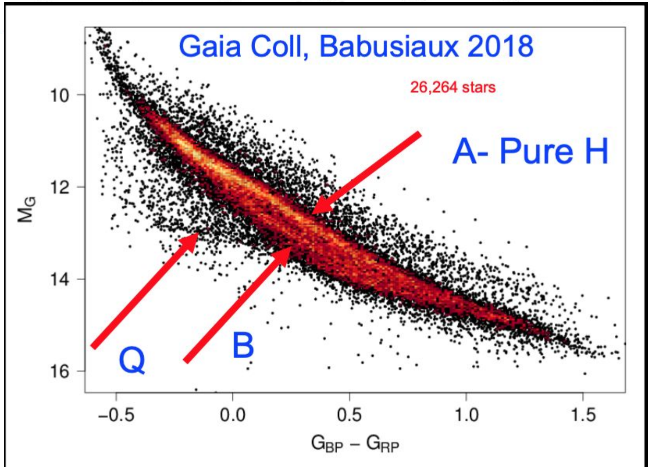


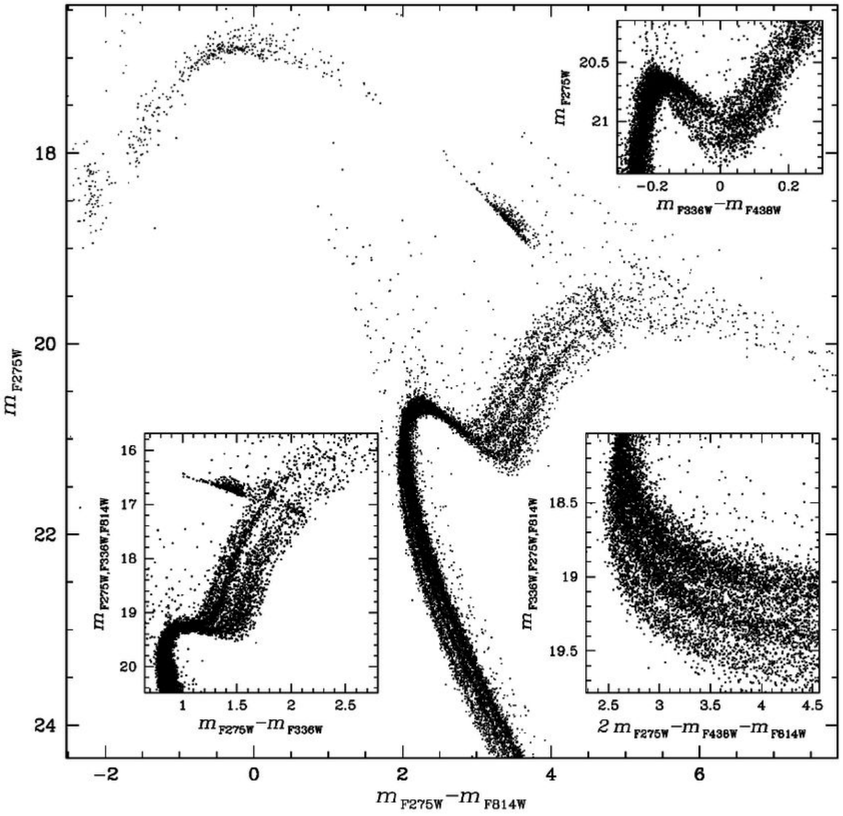
APOGEE



Gaia


A Gap in the MS
Jao+2018
M_G
6
8
10
12
14
2.0
2.2
2.4
2.6
2.8
3.0
3.2

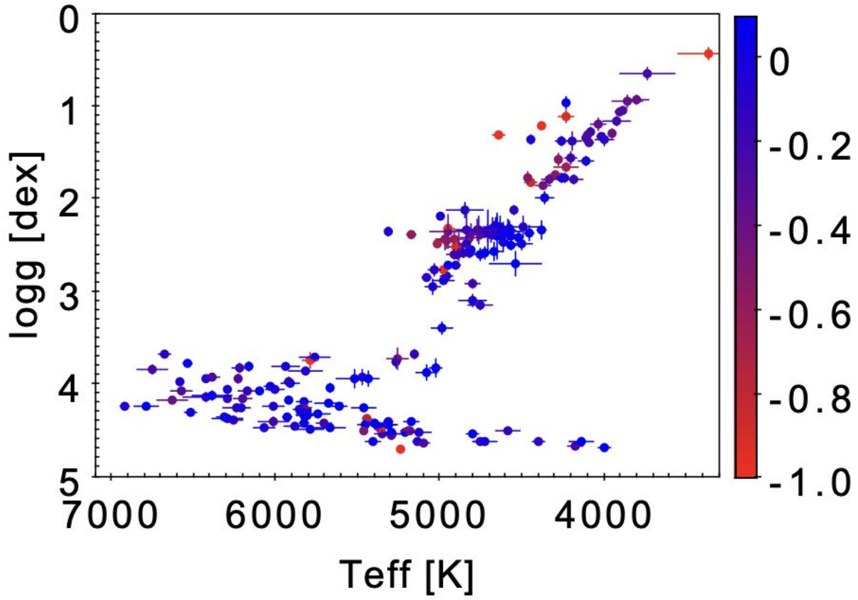

Gaia Coll, Babusiaux 2018
26,264 stars
A- Pure H
Q
B
M_G
10
12
14
16
−0.5
0.0
0.5
1.0
1.5
$G_{BP} - G_{RP}$


m_{F275W}
18
20
22
24
−2
0
2
4
6
$m_{F275W} - m_{F814W}$
m_{F275W}
20.5
21
−0.2
0
0.2
$m_{F336W} - m_{F438W}$
$m_{F275W,F336W,F814W}$
16
17
18
19
20
1
1.5
2
2.5
$m_{F275W} - m_{F336W}$
$m_{F336W,F275W,F814W}$
18.5
19
19.5
2.5
3
3.5
4
4.5
$2\,m_{F275W} - m_{F438W} - m_{F814W}$

Towards precision spectroscopy - Bridge the gap between models and observations

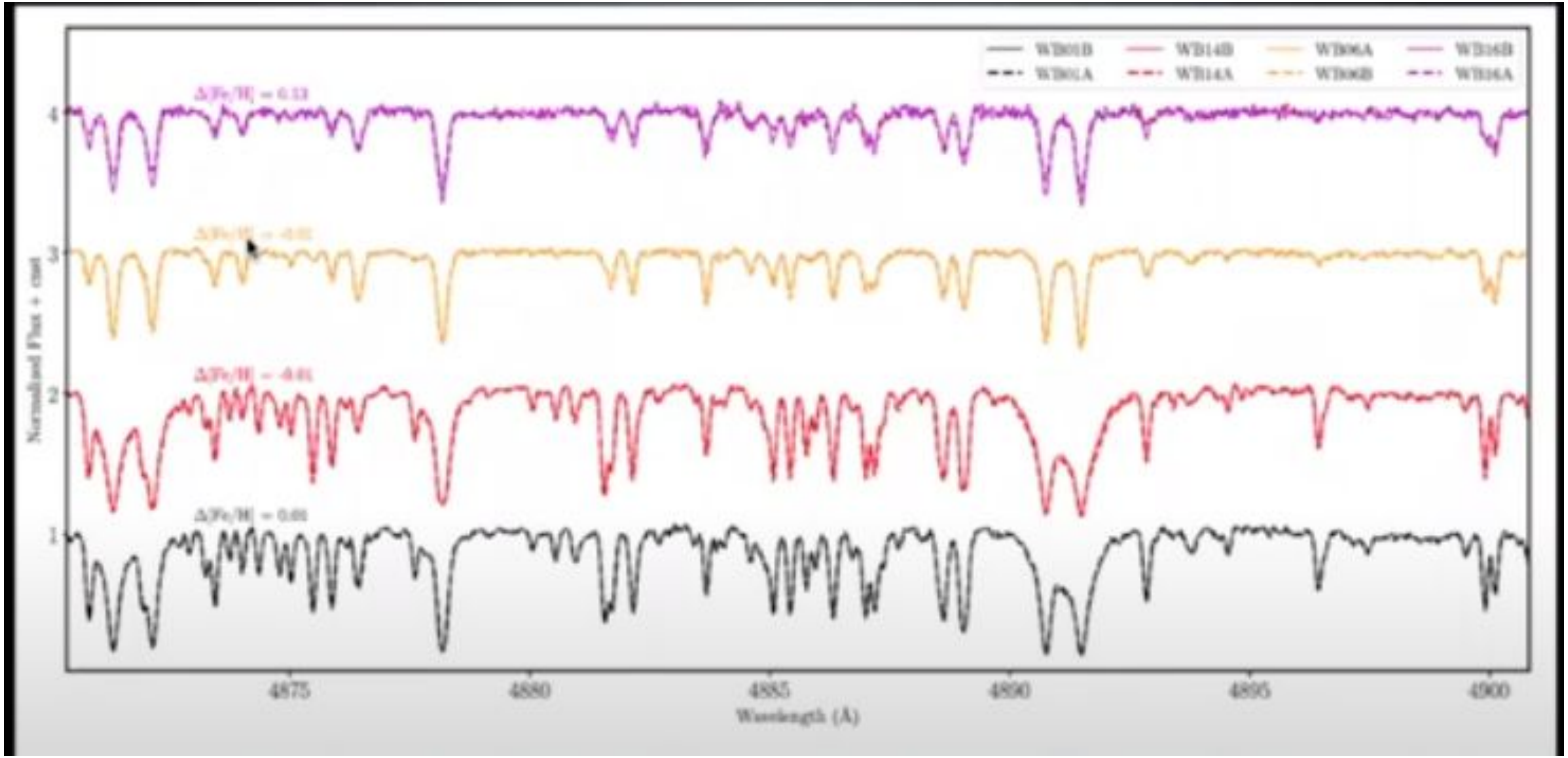
Teff and log g through the fundamental relations with a minimum of assumptions and theoretical modelling

Gaia benchmark stars - 200 stars, stellar diameters and parallax

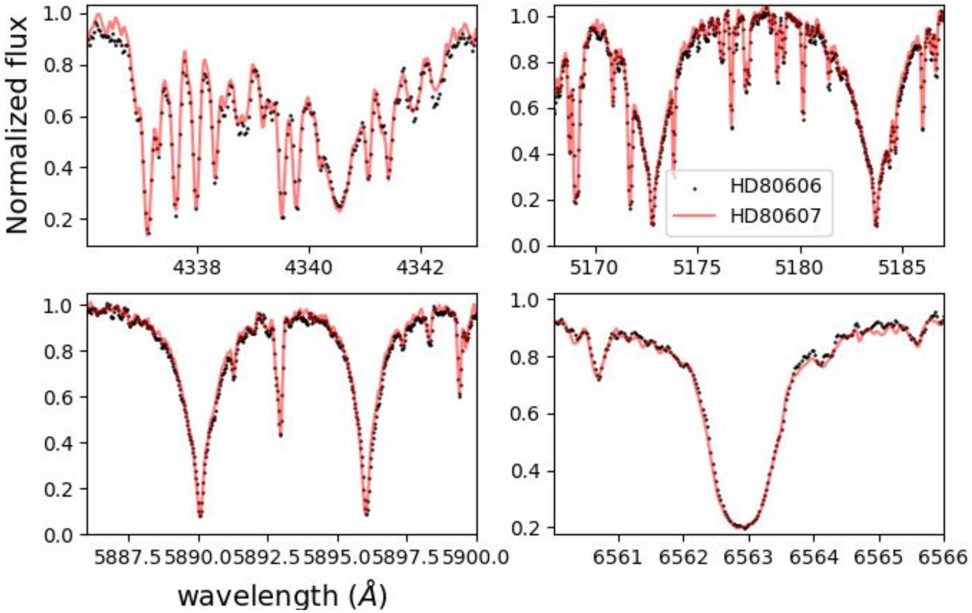


Soubiran et al (2024)

Differential Abundance Stellar twins



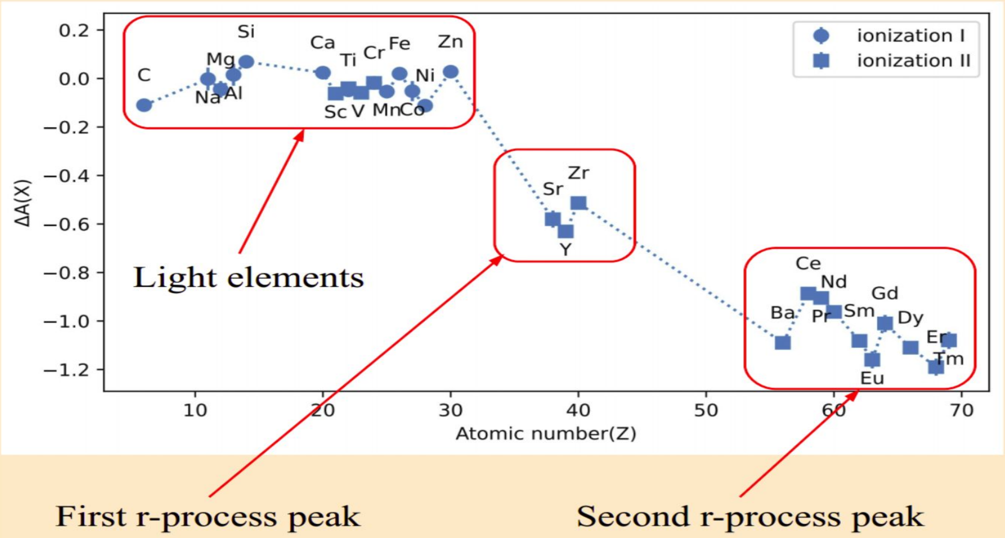
Differential approach for solar twins



Parameter estimated	HD 80606	HD 80607
Teff (k)	Δ 37 K	5650 K
logg	Δ 0.03	4.4
[Fe/H]	Δ 0.012 dex	0.32 dex
εt (km/s)	Δ 0.3	1.30

Unni et al. in preparation

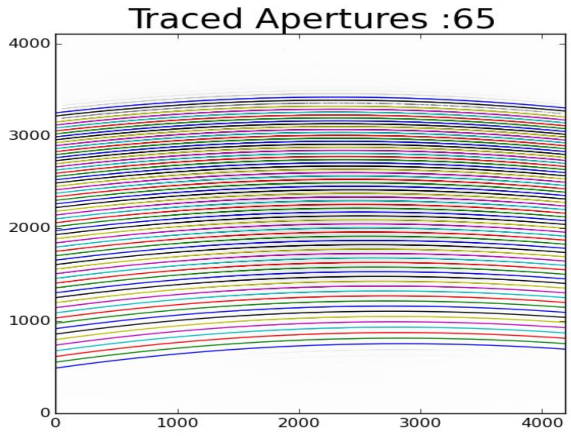
Is r-process pattern universal?

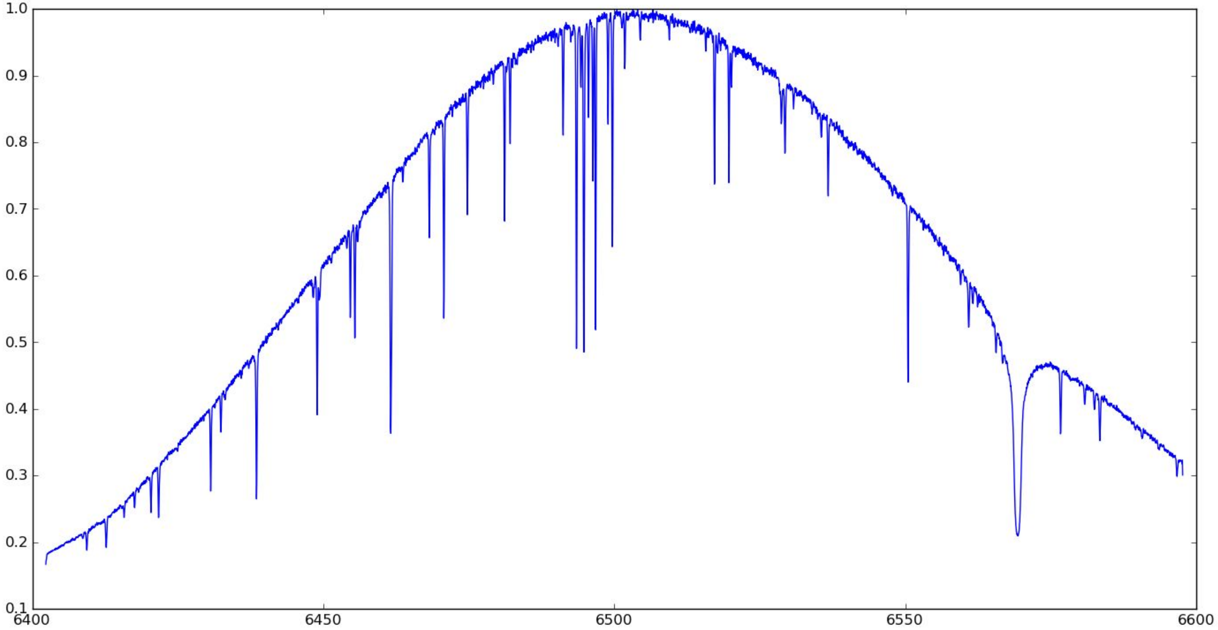


Saraf+2024a (with collaborators)

- ➢ Alpha and Fe-peak elements in r-I and r-II stars come from same site that does not produce n-capture elements.
- ➢ Lighter n-capture elements (Sr,Y,Zr) need additional production site compared to heavy n-capture elements (Ba, Eu, …)
- ➢ Even among the main r-process elements there is significant difference (0.1 − 0.2 dex) between r-I and r-II stars.

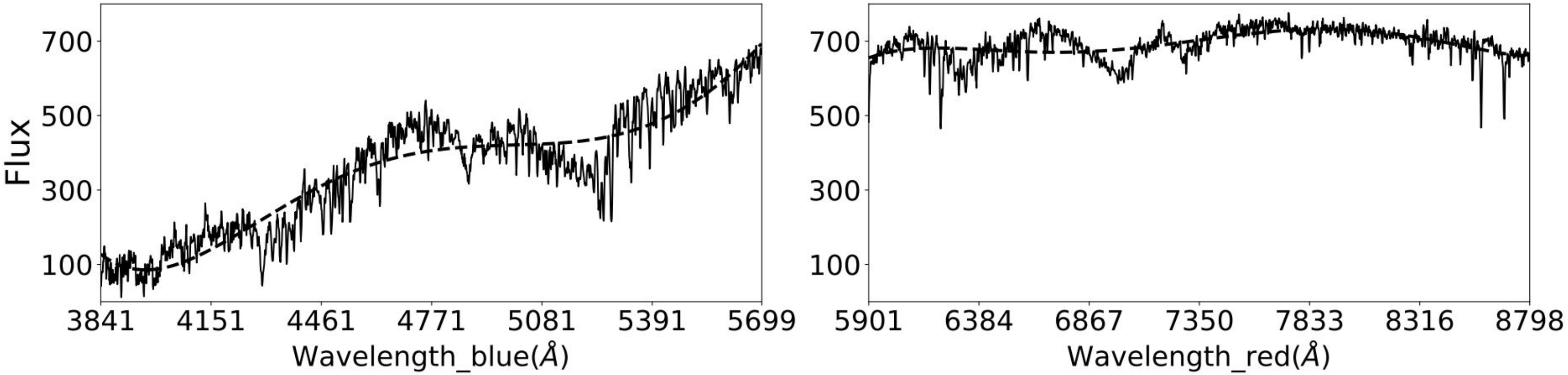
Going down a step into data reduction and applying data driven approach



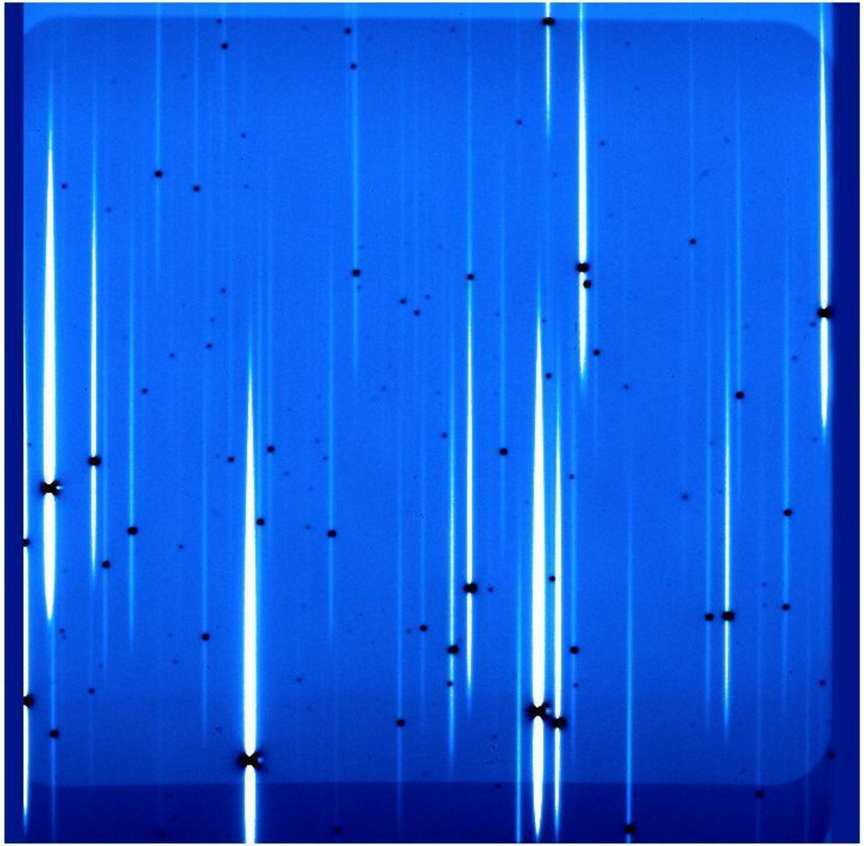


Instrument model based extraction and calibration – Stability places a crucial role
Chanumolu et al. 2015, 2016

Continuum fitting in LAMOST



Data driven approach to Background subtraction slitless spectroscopy



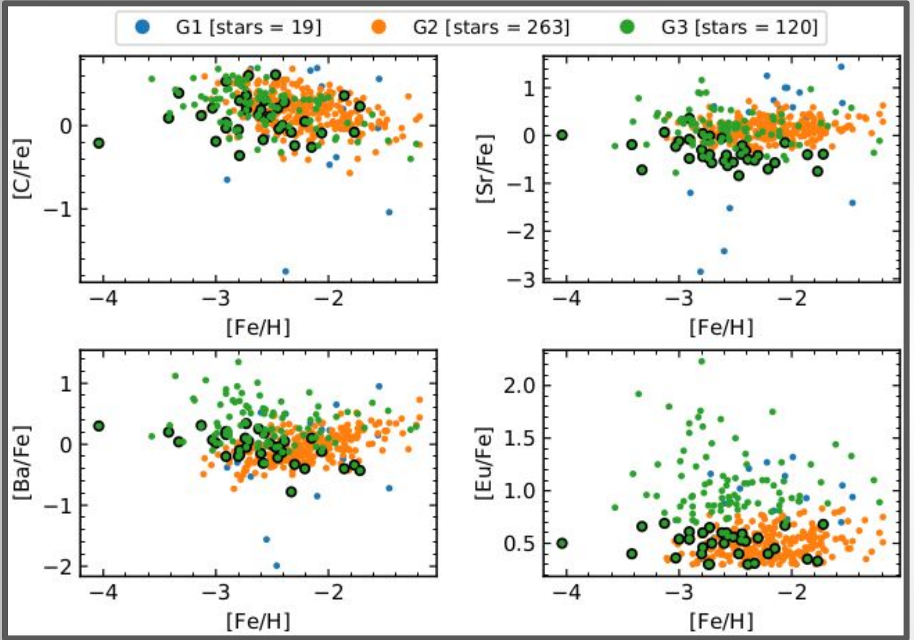
Divakar et al. in preparation

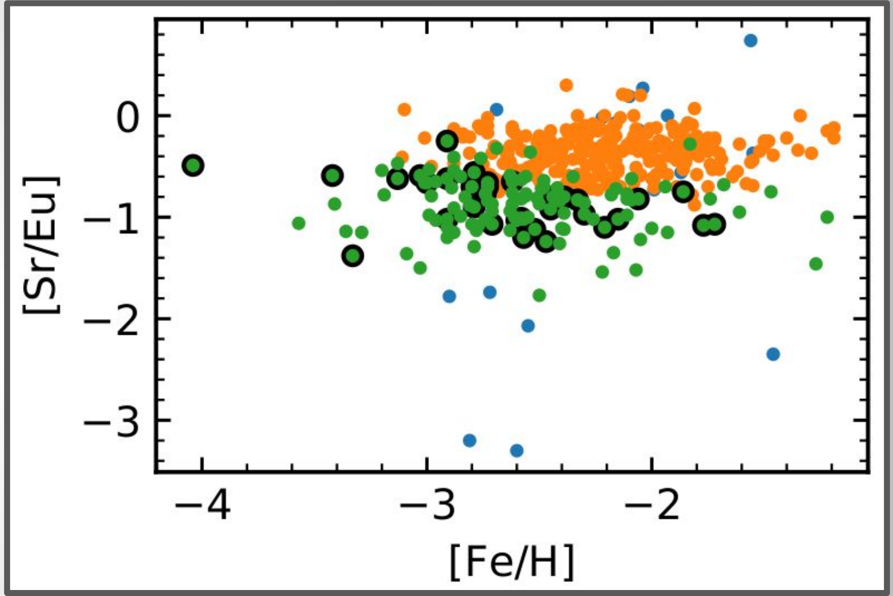
Application at the higher level

Description	Definition	Abbreviation
Population III stars	Postulated first stars, formed from metal-free gas	Pop III
Population II stars	Old (halo) stars formed from low-metallicity gas	Pop II
Population I stars	Young (disk) metal-rich stars	Pop I
Super-metal-rich	$[Fe/H] > 0.0$	MR
Solar	$[Fe/H] = 0.0$	None
Metal-poor	$[Fe/H] < -1.0$	MP
Very metal-poor	$[Fe/H] < -2.0$	VMP
Extremely metal-poor	$[Fe/H] < -3.0$	EMP
Ultra-metal-poor	$[Fe/H] < -4.0$	UMP
Hyper-metal-poor	$[Fe/H] < -5.0$	HMP
Mega-metal-poor	$[Fe/H] < -6.0$	MMP
Septa-metal-poor	$[Fe/H] < -7.0$	SMP
Octa-metal-poor	$[Fe/H] < -8.0$	OMP
Giga-metal-poor	$[Fe/H] < -9.0$	GMP
Ridiculously metal-poor	$[Fe/H] < -10.0$	RMP

Frebel 2018

Data driven classification





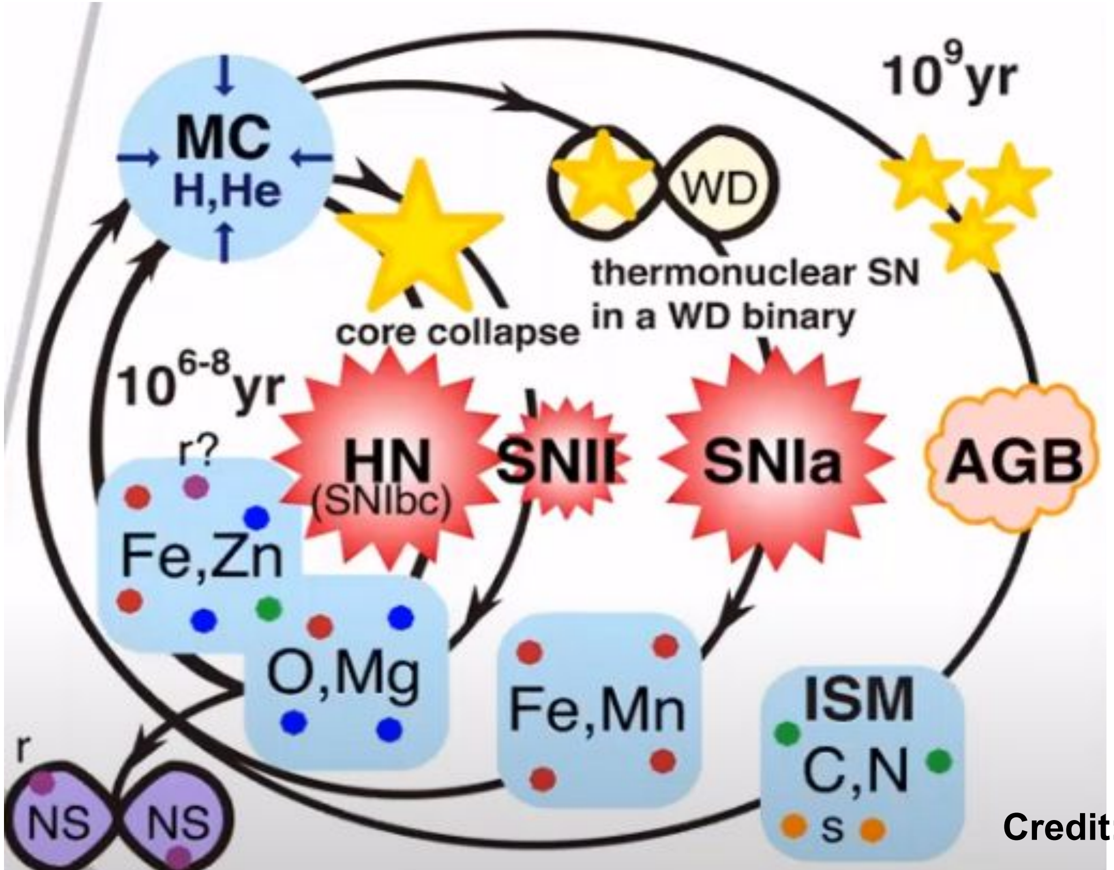
<u>group G1:</u> % of r-I = 31.58
% of r-II = 68.42

<u>group G2:</u> % of r-I = 92.02
% of r-II = 7.98

<u>group G3:</u> % of r-I = 29.17
% of r-II = 70.83

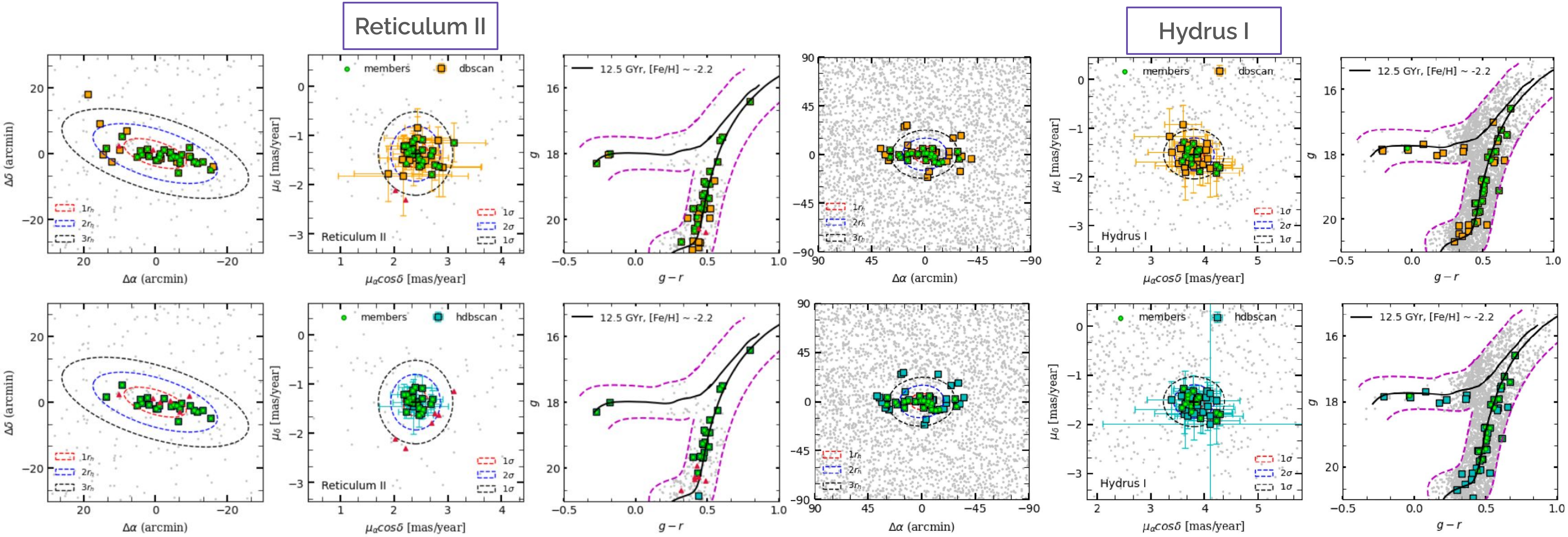
Saraf et al. 2024

Data driven approach to chemical tagging



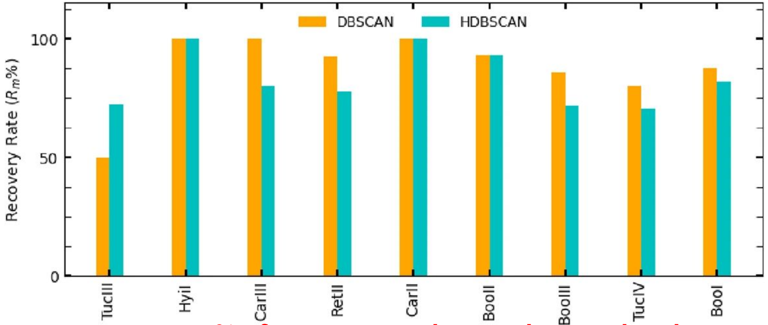
Credit: Kobhayashi

Machine Learning Techniques to find new members of UFDs - Results

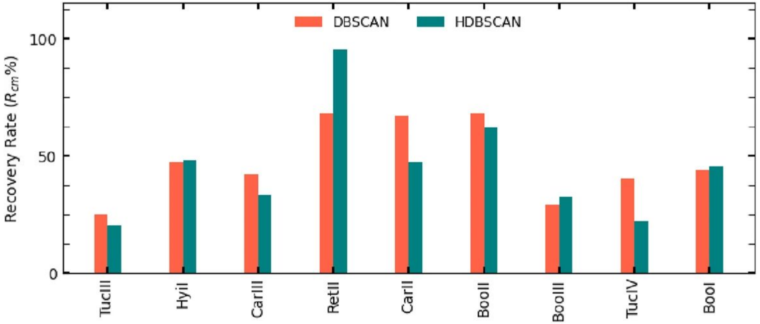


Green squares are known spectroscopic members, Yellow squares are identified new candidates using DBSCAN, Blue squares are identified new new candidates using HDBSCAN

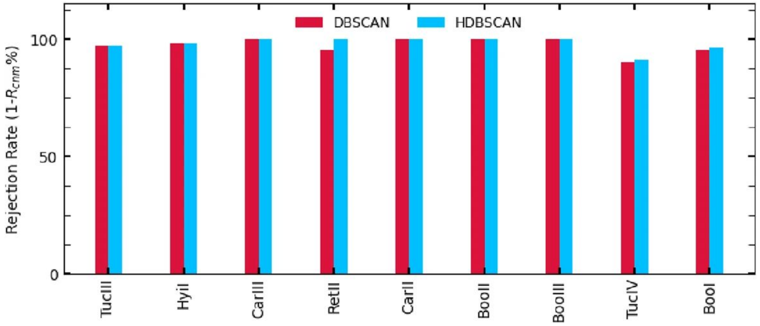
Performance of DBSCAN



> 75% of spectroscopic member retrieval

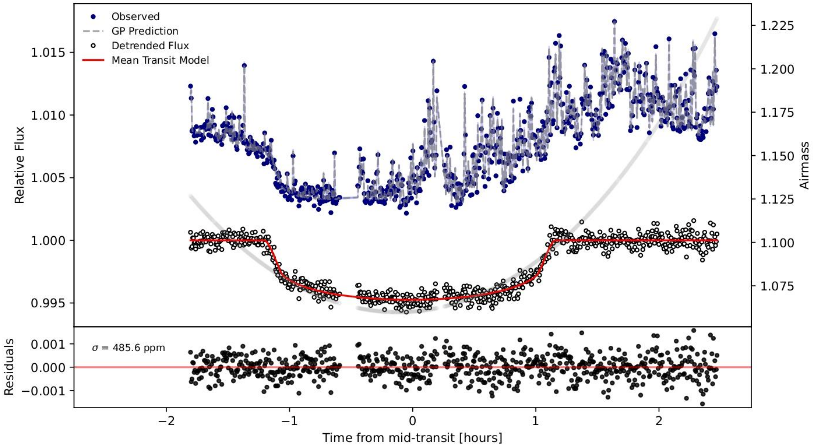


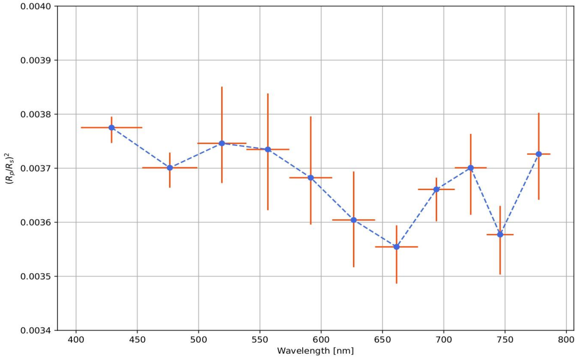
DBSCAN in terms of candidate member retrieval



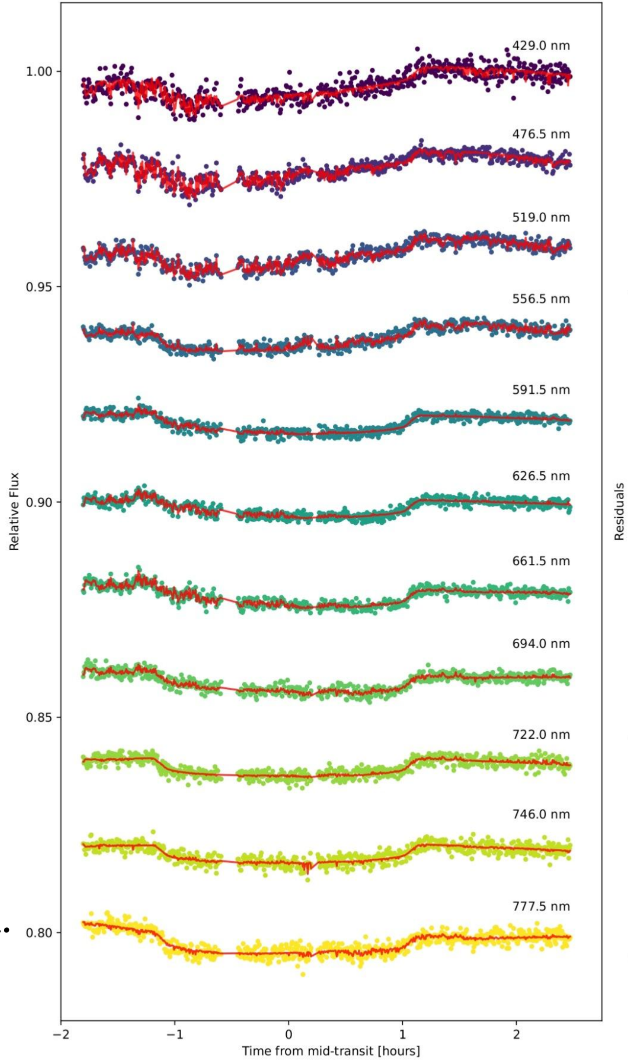
> 95% of non-member rejection

Exoplanet Transit spectroscopy
2D Gaussian process regression



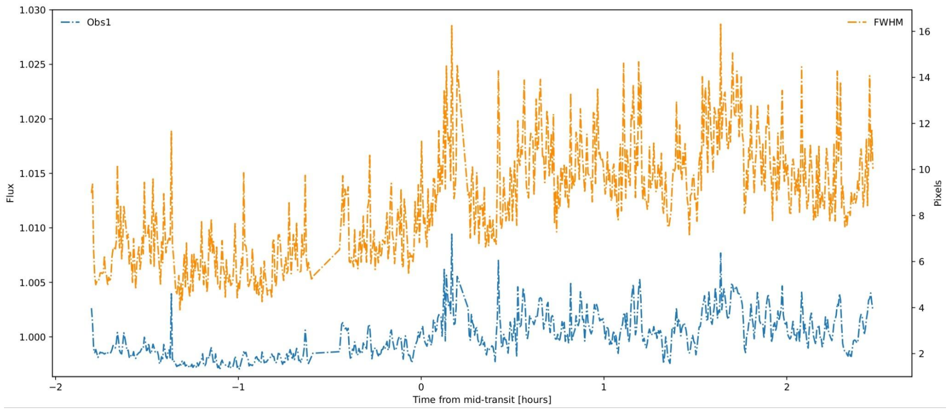


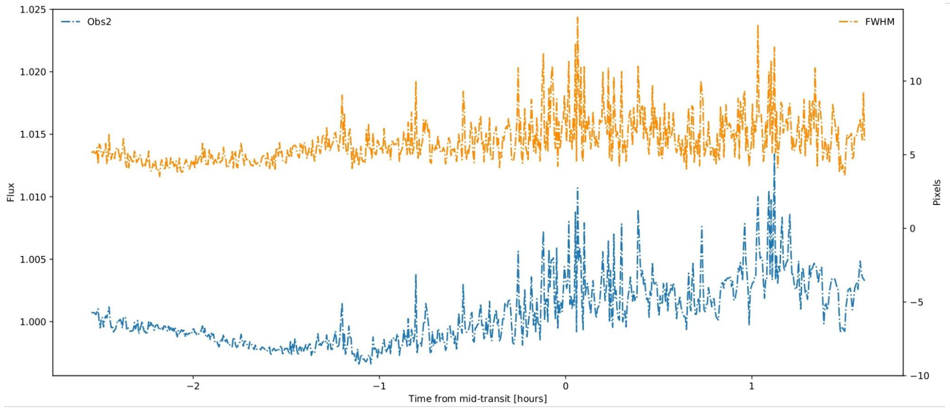
Manickavasaham et al.
in preparation



Comparison of WLC systematics and FWHM of spectral profile

August 30, 2016





September 25, 2017

Summary and future direction

Even with lot of data, high resolution and high SNR, we are yet to reach high precision stellar abundances

Information content in stellar spectra at different spectral resolution

Exoplanet host star studies has added more data, to obtain time resolved and disk resolved spectroscopy for stars, this adding more information and complexity

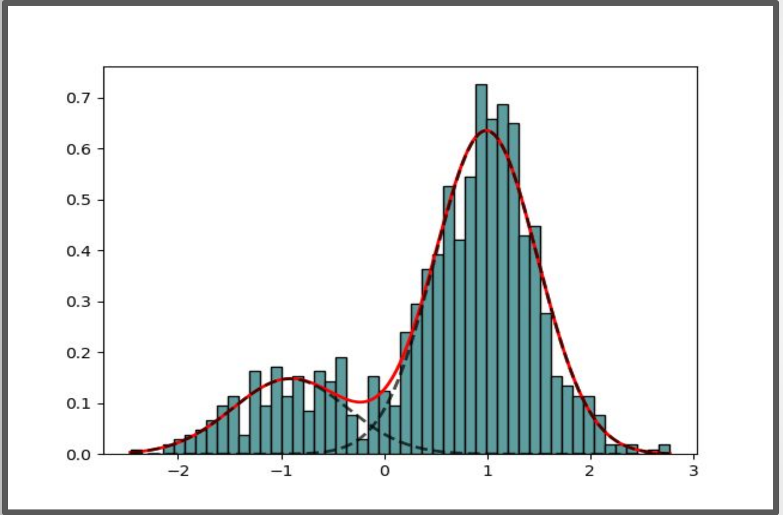
Astrophysical sites of elements production and chemical tagging of stars

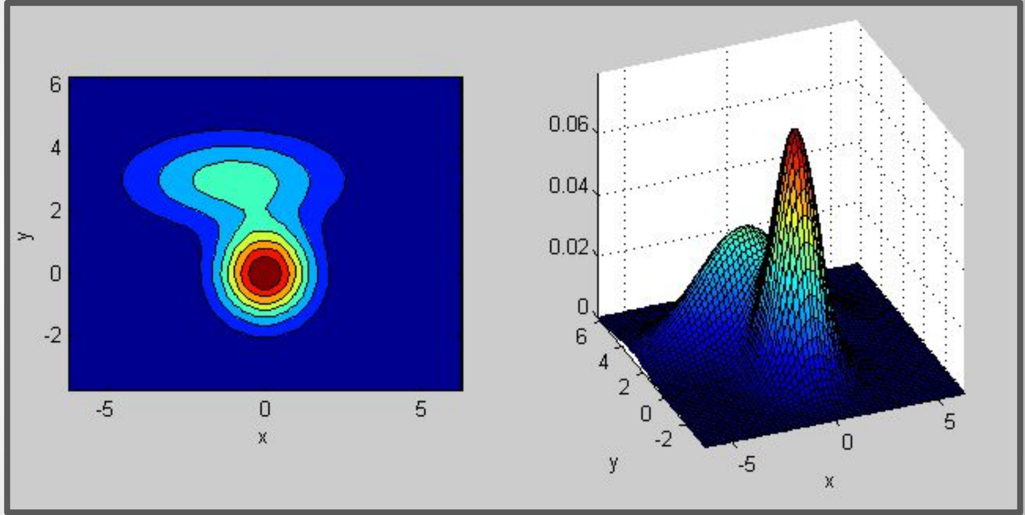
Chemical evolutionary models

Data driven approach to theoretical stellar models and improve the atomic and molecular data

Thank You!

What is GMM (Gaussian Mixture Model)?





Information content of stellar spectra at different resolution

Forward model - Retrieval

Models incomplete physics,

Incomplete molecular atomic data. Even wrong data

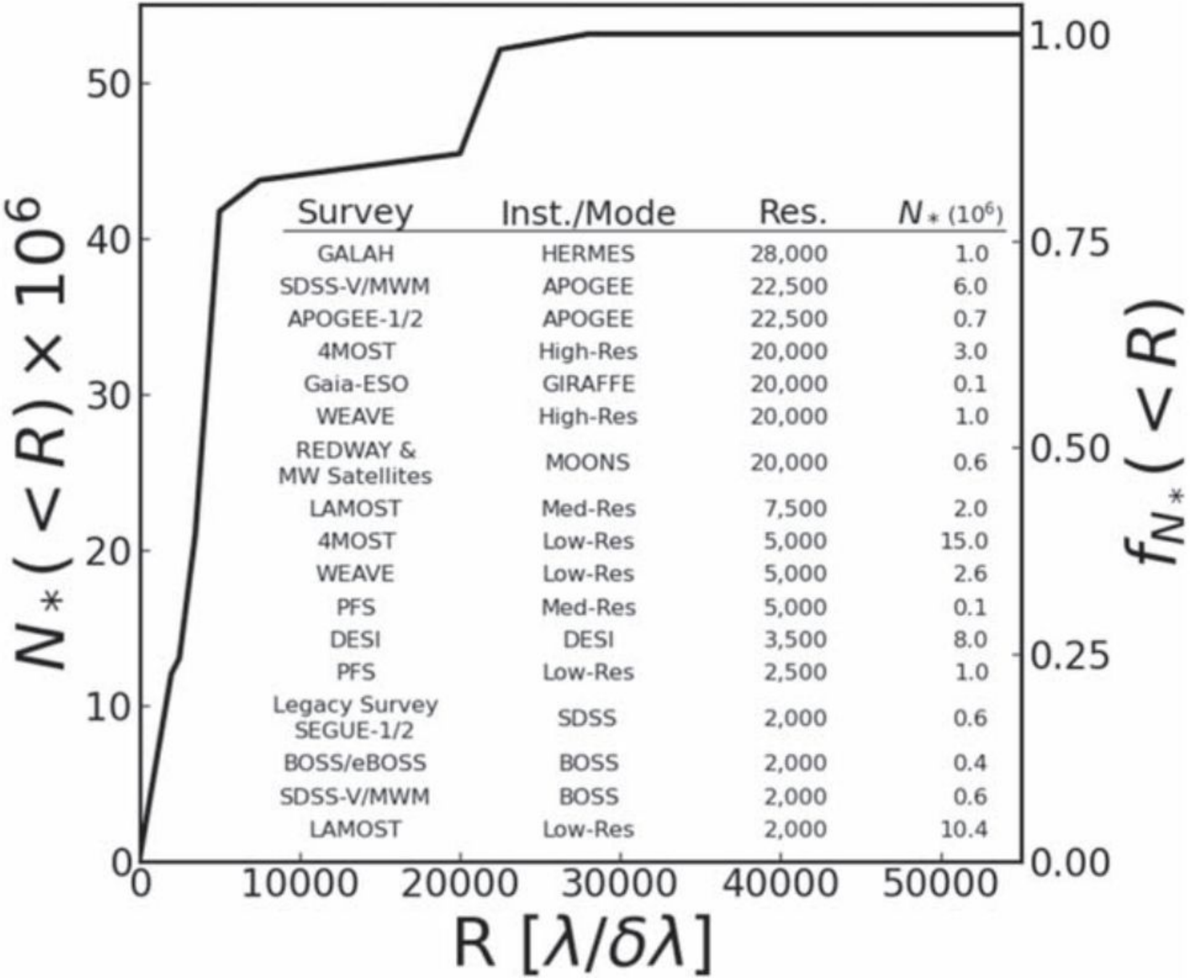
Telluric subtraction, Continuum fitting

Data processing

Normalisation – Continuum fitting or renormalisation (dividing by median) ⇒ Instrument stability

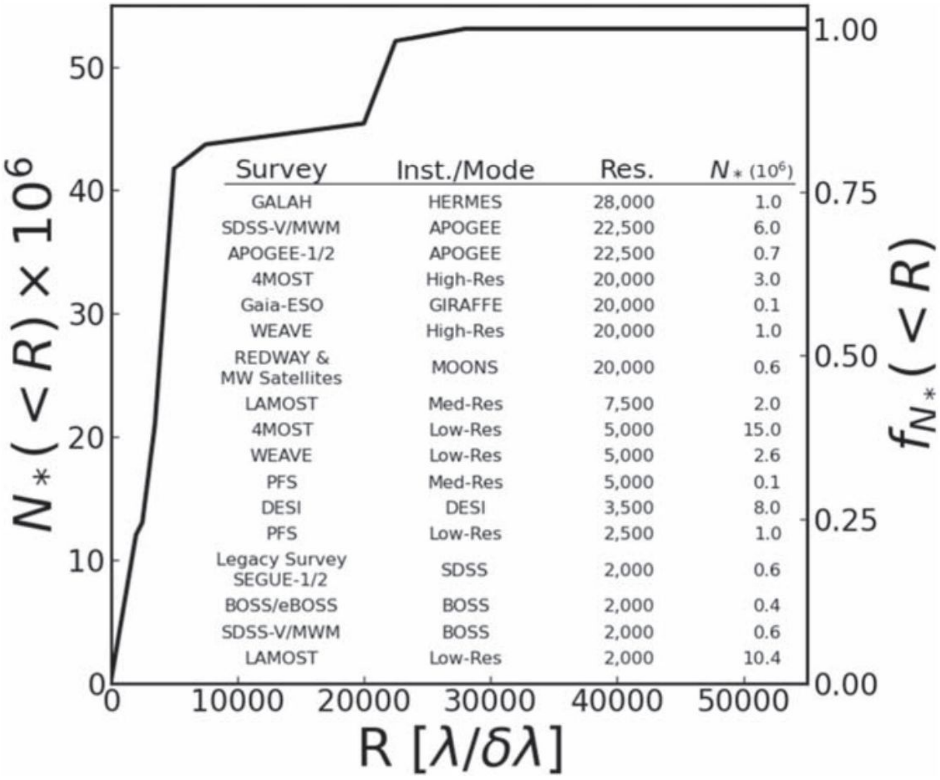
Stellar classification and Stellar Parametrisation

Apache Point Observatory Galactic Evolution Experiment (APOGEE) (Prieto et al. 2010), the Galactic Archaeology with HERMES Survey (GALAH) (De Silva et al. 2015), the Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST) Experiment for Galactic Understanding and Exploration (LEGUE) (Deng et al. 2012; Zhao et al. 2012), The Gaia-ESO Public Spectroscopic Survey (GaiaESO) (Gilmore et al. 2012), the Sloan Extension for Galactic Understanding and Exploration (SEGUE) (Yanny et al. 2009), the RAdial Velocity Experiment (RAVE) (Steinmetz et al. 2006)



Survey	Inst./Mode	Res.	N_* (10^6)
GALAH	HERMES	28,000	1.0
SDSS-V/MWM	APOGEE	22,500	6.0
APOGEE-1/2	APOGEE	22,500	0.7
4MOST	High-Res	20,000	3.0
Gaia-ESO	GIRAFFE	20,000	0.1
WEAVE	High-Res	20,000	1.0
REDWAY & MW Satellites	MOONS	20,000	0.6
LAMOST	Med-Res	7,500	2.0
4MOST	Low-Res	5,000	15.0
WEAVE	Low-Res	5,000	2.6
PFS	Med-Res	5,000	0.1
DESI	DESI	3,500	8.0
PFS	Low-Res	2,500	1.0
Legacy Survey SEGUE-1/2	SDSS	2,000	0.6
BOSS/eBOSS	BOSS	2,000	0.4
SDSS-V/MWM	BOSS	2,000	0.6
LAMOST	Low-Res	2,000	10.4

Stanford et al.
2023, ApJS, 267



Survey	Inst./Mode	Res.	N_* (10^6)
GALAH	HERMES	28,000	1.0
SDSS-V/MWM	APOGEE	22,500	6.0
APOGEE-1/2	APOGEE	22,500	0.7
4MOST	High-Res	20,000	3.0
Gaia-ESO	GIRAFFE	20,000	0.1
WEAVE	High-Res	20,000	1.0
REDWAY & MW Satellites	MOONS	20,000	0.6
LAMOST	Med-Res	7,500	2.0
4MOST	Low-Res	5,000	15.0
WEAVE	Low-Res	5,000	2.6
PFS	Med-Res	5,000	0.1
DESI	DESI	3,500	8.0
PFS	Low-Res	2,500	1.0
Legacy Survey SEGUE-1/2	SDSS	2,000	0.6
BOSS/eBOSS	BOSS	2,000	0.4
SDSS-V/MWM	BOSS	2,000	0.6
LAMOST	Low-Res	2,000	10.4