# Machine learning in APOGEE: Unsupervised spectral classification with $K$-means

**Rafael A. Garcia-Dias**

**Instituto de Astrofísica de Canarias - IAC**

Supervisors:

**Carlos Allende Prieto**

**Jorge Sánchez Almeida**

May 23$^{th}$, 2017

# Machine learning vs. AI

## Machine learning

"such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions."

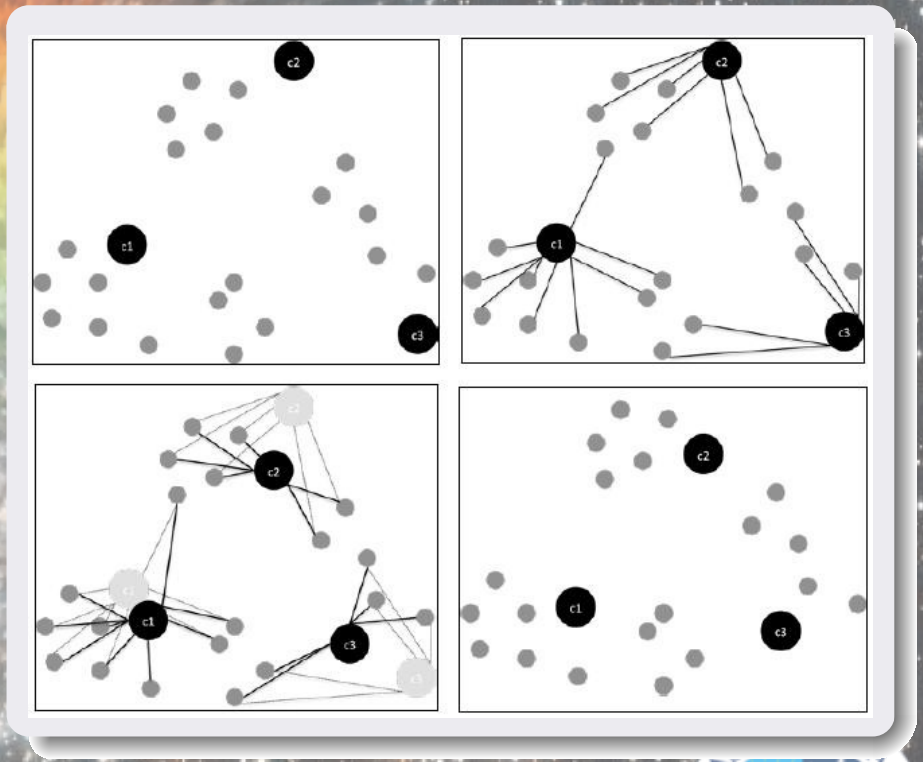
## Artificial Intelligence

"the term *artificial intelligence* is applied when a machine mimics *cognitive* functions that humans associate with other human minds."

# $K$-means

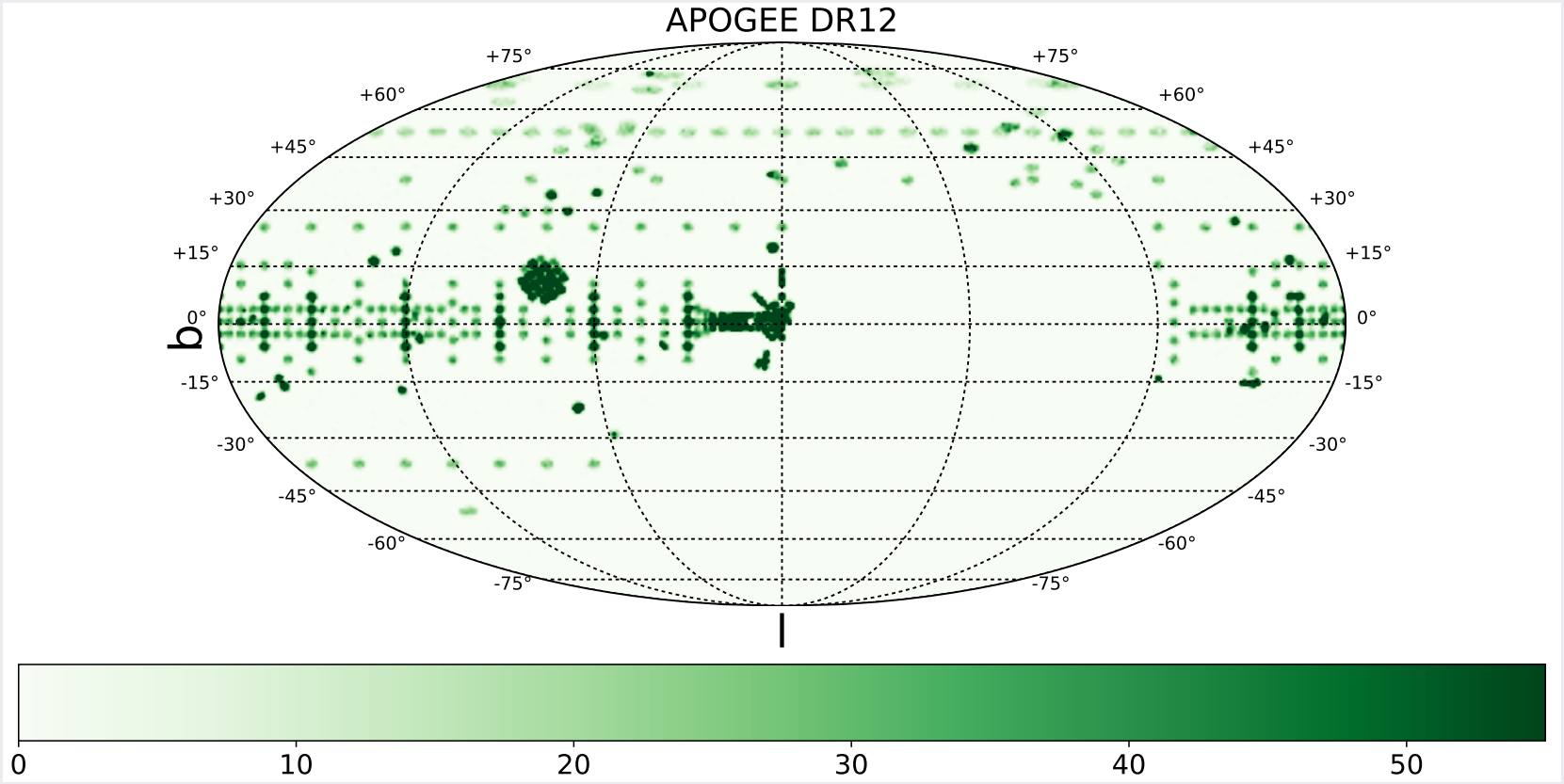
## Algorithm

- Choose initial centroids;
- While != Stability condition:
  - Assign objects to the nearest centroid;
  - Recalculate centroids as the mean of the objects in each class;
- Finish.
- Repeat as many times you need to be sure the random initialization doesn't affect the results.

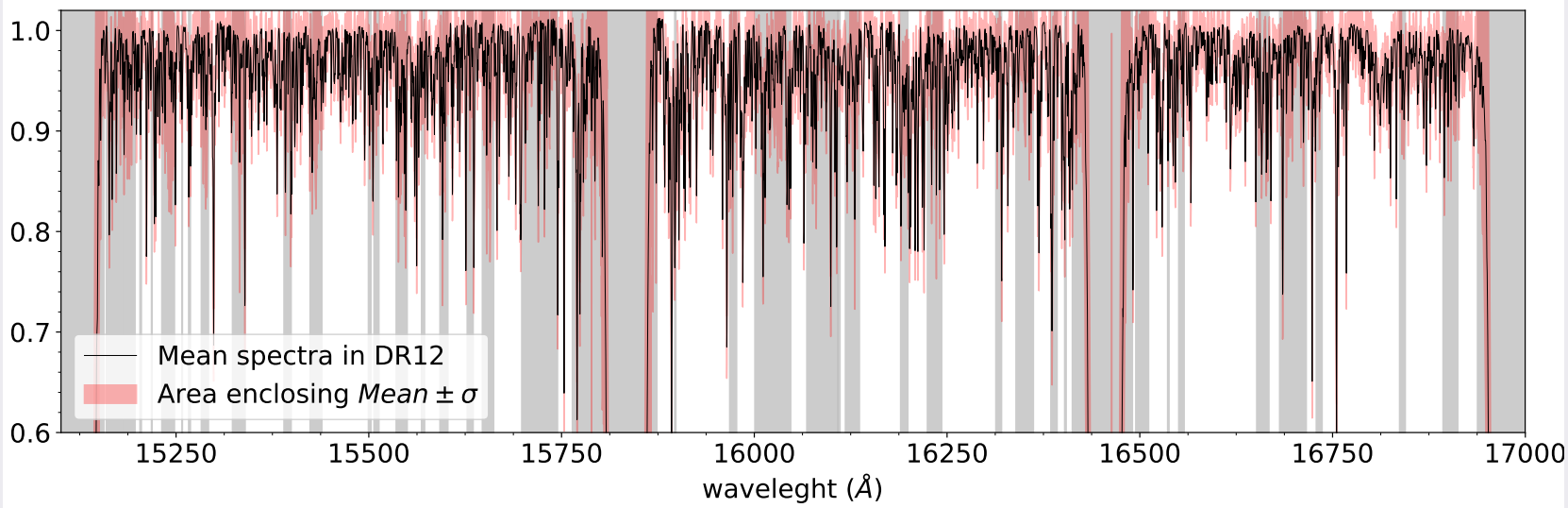


# Data

## 153,847 stars



# Data

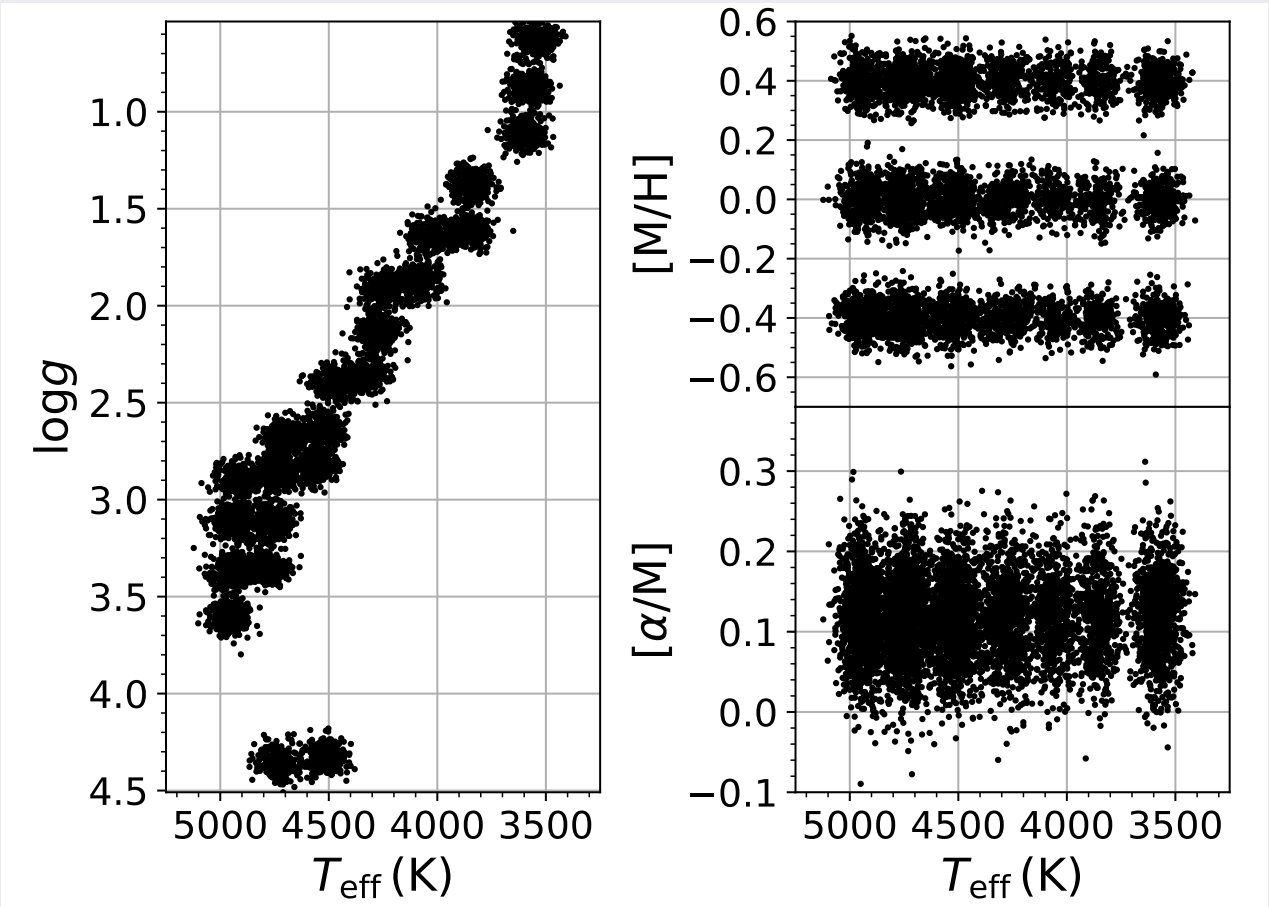



## Mask

- Mask were based in telluric absorption and sky emittions.
- We use 4838 pixels from the 8575 pixels in APOGEE.

# Mock data

## Ferre interpolated Kurucz spectral models



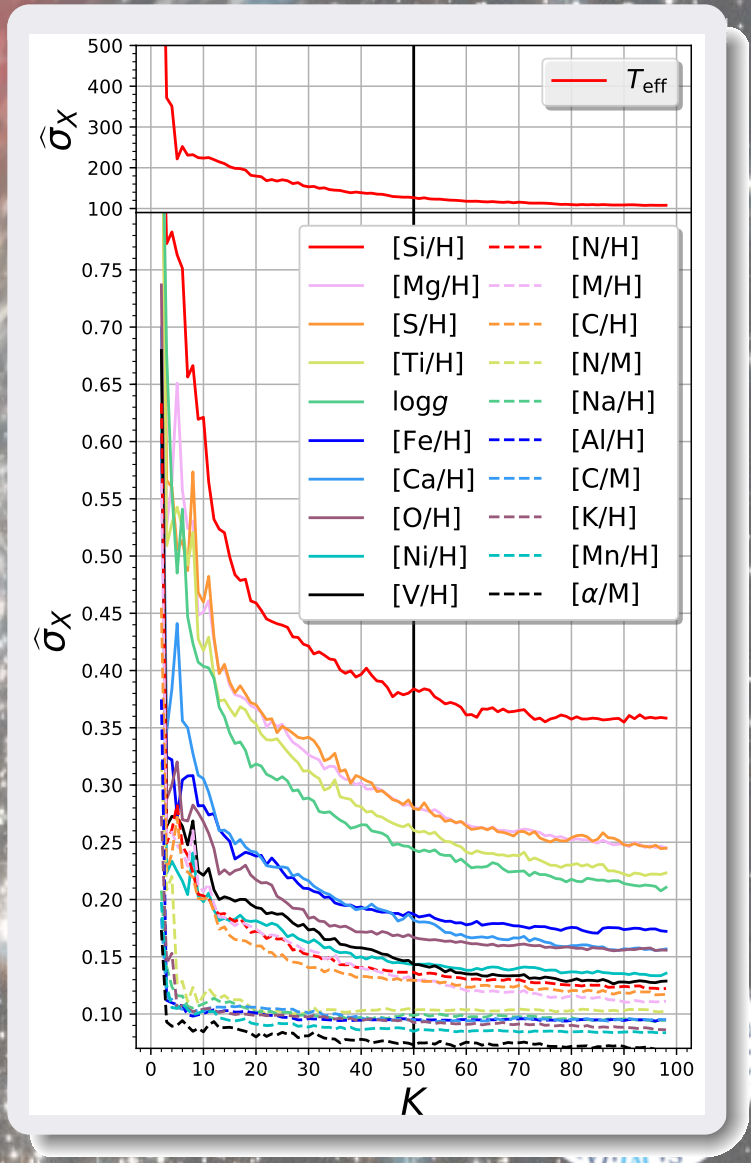

https://github.com/callendeprieto/ferre

# Choosing $K$

## Median within cluster $\sigma$

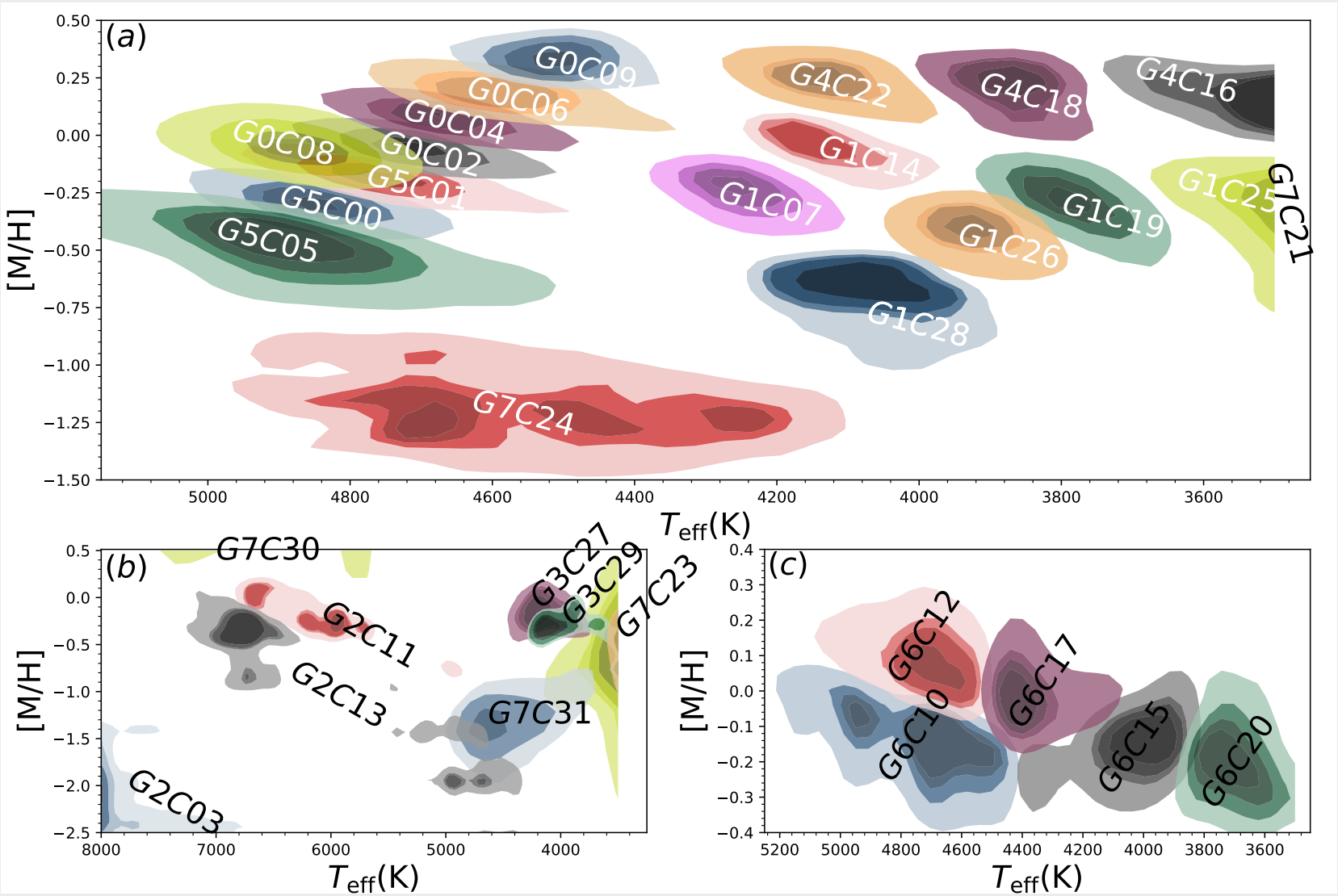
K was chosen based in the variation of the median within cluster standard deviation with K.

$$\lim_{K \to \infty} \frac{d\widehat{\sigma}}{dK} \to 0$$



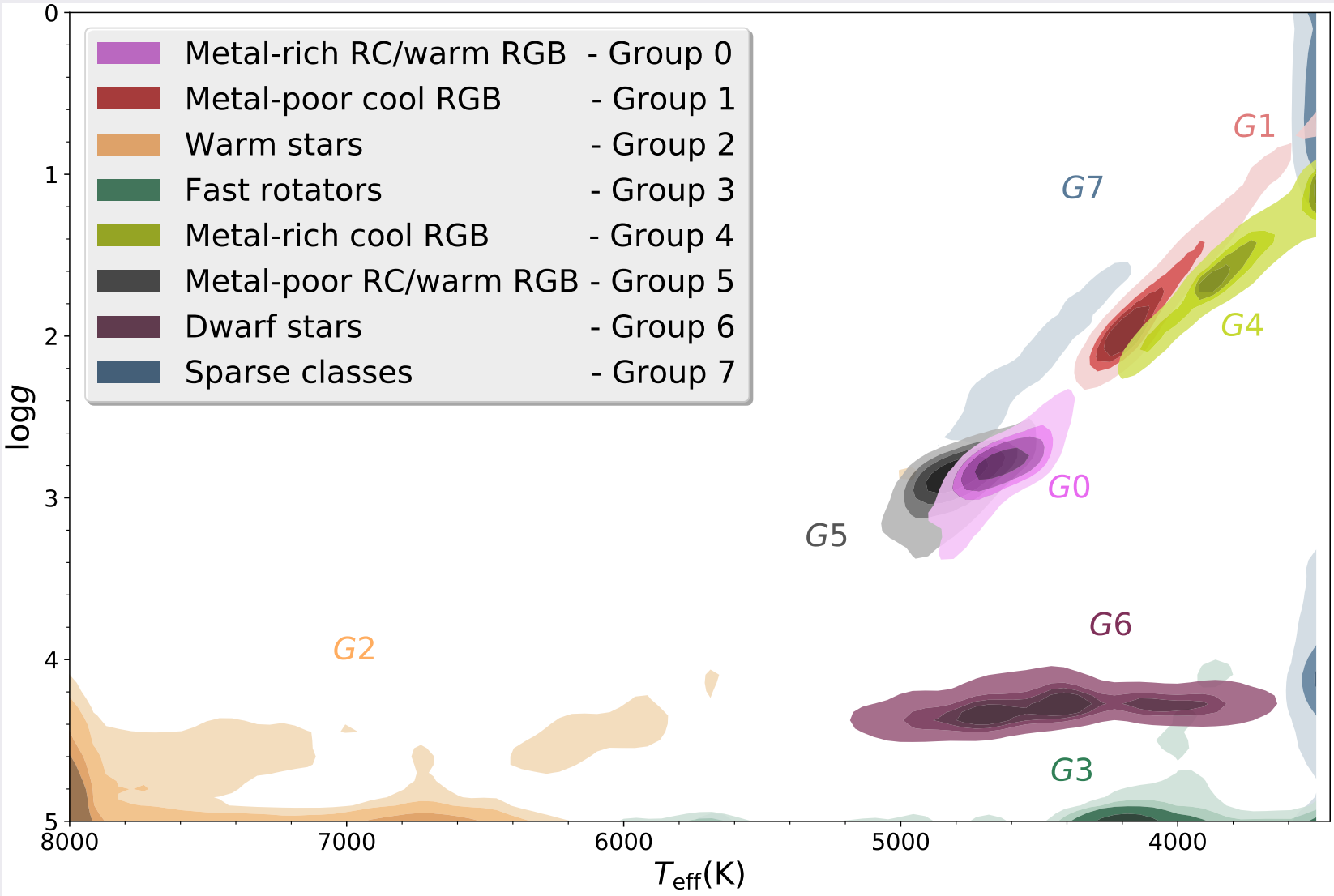
# Final classification

## 32 classes contain 99% of the stars.



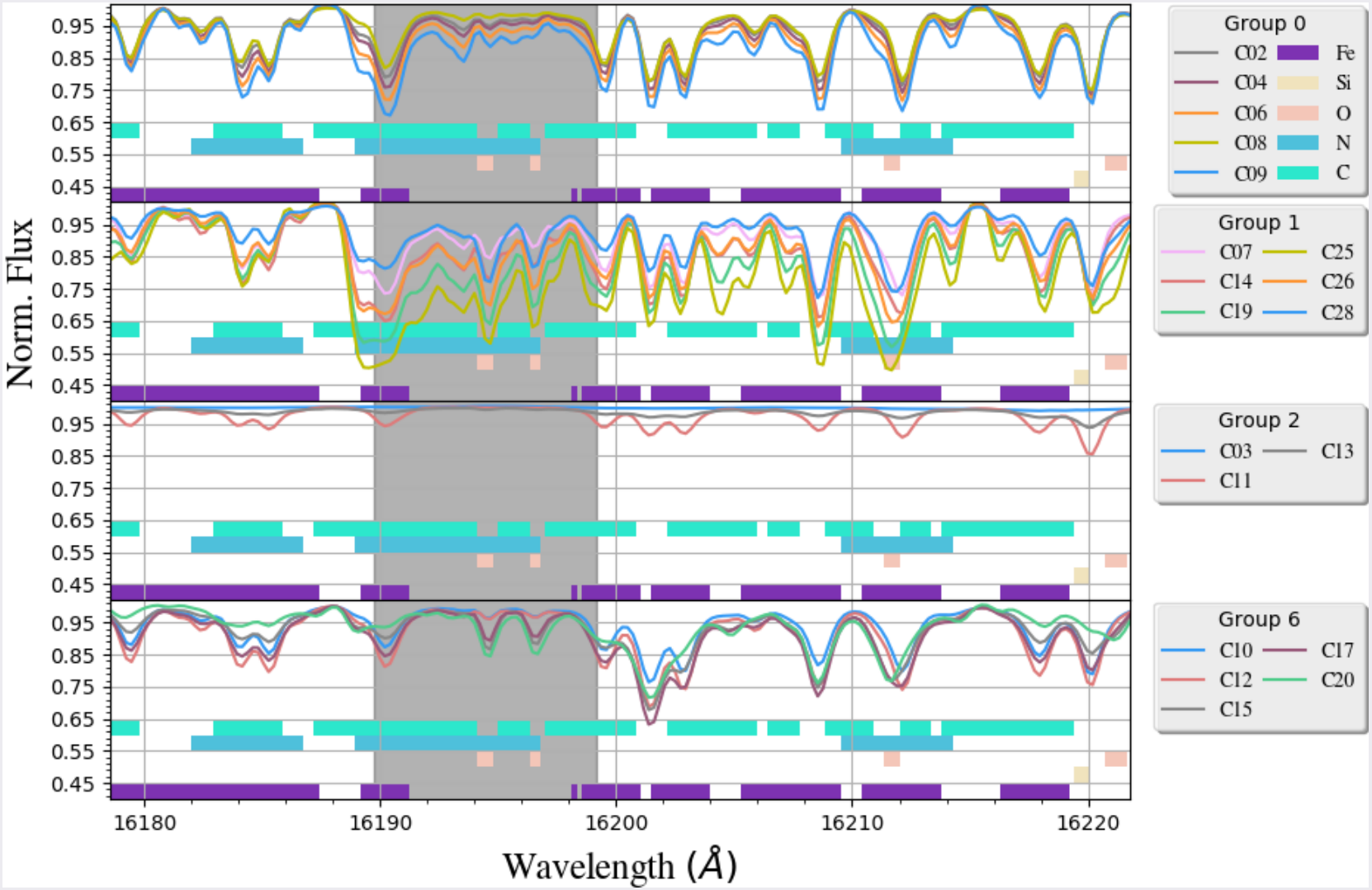
# Final classification

## Groups



# Final classification

## Mean spectra for classes



# Summary

**Table A1.** Summary of the classes and complementary material.

| Group | Class[a] | Stellar type[b] | Gal. component[c] | Comment |
|---|---|---|---|---|
| Metal-rich RC and RGB | | | | |
| 0 | Class 02 | K-Giants | Thin disk | Lowest [M/H] in the group, 31% RC. |
| 0 | Class 04 | K-Giants | Thin disk | 26% RC |
| 0 | Class 06 | K-Giants | Thin disk | 26% RC |
| 0 | Class 08 | Sub Giants, K-Giants | Thin disk | Warmest in the group, 1% RC. |
| 0 | Class 09 | K-Giants | Thin disk | [M/H] near to grid limits, 21% RC. |
| Metal-poor cool RGB | | | | |
| 1 | Class 07 | K-Giants | Disk | Thick disk. |
| 1 | Class 14 | K-Giants | Disk | — |
| 1 | Class 19 | K/M-Dwarfs | Disk | $T_{\rm eff}$ near to the grid limits. |
| 1 | Class 25 | M-Giants | Disk | $T_{\rm eff}$ near to the grid limits. |
| 1 | Class 26 | K-Giants | Disk | High alpha blob. |
| 1 | Class 28 | K-Giants | Bulge/centre | Most metal-poor stars. |
| Warm stars | | | | |
| 2 | Class 03 | Blue stars | Disk | Warmest telluric standards |
| 2 | Class 11 | F/G-Dwarfs | High g. latitude | Warm, telluric standards. |
| 2 | Class 13 | Blue stars | — | Warm fast rotation stars. Telluric standards. |
| Fast rotators | | | | |
| 3 | Class 27 | K/M-Dwarfs | — | Fast rotators. |
| 3 | Class 29 | M-Dwarfs | — | Fast rotators. |
| Metal-rich cool RGB | | | | |
| 4 | Class 16 | K/M-Giants | Disk | $T_{\rm eff}$ near to the grid limits. |
| 4 | Class 22 | K-Giants | Thin disk | [M/H] near to the grid limits. |
| Metal-poor RC and RGB | | | | |
| 5 | Class 00 | K-Giants | Disk | Broad in atmospheric parameters. |
| 5 | Class 01 | K-Giants | Disk | Whole RGB |
| 5 | Class 05 | Sub Giants, G/K-Giants | Disk | Broad in atmospheric parameters. |
| Dwarf stars | | | | |
| 6 | Class 10 | G/K-Dwarfs | Thin disk | — |
| 6 | Class 12 | K-Dwarfs | Thin disk | — |
| 6 | Class 15 | K-Dwarfs | High g. latitude | — |
| 6 | Class 17 | K-Dwarfs | Thin disk | — |
| 6 | Class 20 | M-Dwarfs | High g. latitude | Atmospheric parameter near to the grid limits. |
| Sparse classes | | | | |
| 7 | Class 21 | M-Giants | Bulge/Centre/Disk | Atmospheric parameter near to the grid limits. |
| 7 | Class 23 | M-Dwarfs | — | Atmospheric parameter near to the grid limits. |
| 7 | Class 24 | Giants | Halo | High alpha metal-poor stars. |
| 7 | Class 30 | — | — | Poor fit, M31 clusters, high g. latitude. |
| 7 | Class 31 | Giants | High g. latitude | metal-poor high [$\alpha$/M]. |

# Thank you

## Paper

- https://garciadias.github.io/cv/k-means-apogee.pdf

## Contact

- rafaelagd@gmail.com

## Soon the code will be available in

- https://garciadias.github.io/StarClustering/