

Term Project # 3

Galactic Astronomy (Astr 511); Winter Quarter 2015

prof. Željko Ivezić, University of Washington

This document can be found in GitHub repo 2015ASTR511/Project3

1. Introduction

We will boldly attempt heroic collaborative work: two awesome teams working on cutting-edge problems using recently released APOGEE-ASPCAP data¹:

- **EXD:** using abundances for 15 elements (from summary allStar file), we will study the variation of their 15-D abundance distribution in the 3-dimensional XYZ space, as a function of stellar population (halo, thin disk, thick disk); we will use Extreme Deconvolution (EXD) method to quantify the 15-D distributions, see astroML Book Figures 6.11 and 6.12; nevertheless, it may be prudent to run PCA on the set of these 15-D vectors before attempting (more complicated) EXD (see below).
- **DimRed:** Using spectra (apStar files, one per star), we will run dimensionality-reduction methods PCA, ICA and NMF and analysis results; see astroML Book Figures 7.4, 7.6 and 7.7.

The basic “building blocks” are same for both projects:

1. define the teams, and within each team assign responsibilities for the tasks below;
2. learning about the data properties and starting a paper/notes with a brief Introduction about APOGEE and data properties; paper templates can be copied from the class repo² to appropriate directory within the new 2015ASTR511 repo.
3. learning about how to download and access data (including writing python code);
4. learning about how to (modify and) run appropriate astroML tools;
5. writing analysis summary (including plots) for the paper

¹See <http://www.sdss.org/dr12/irspec/>

²[Astr511/tree/master/HW2015/Project3](https://github.com/2015ASTR511/tree/master/HW2015/Project3)

2. Project EXD

Copying verbatim from the SDSS Data Release webpage³, “The APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP) employs a two-step process to extract abundances: first, determination of the atmospheric parameters by fitting the entire APOGEE spectrum, and second, use of these parameters to fit various small regions of the spectrum dominated by spectral features associated with a particular element in order to derive the individual element abundance.” The first step of ASPCAP estimates seven parameters: effective temperature, surface gravity, microturbulence, overall metal abundance $[M/H]$, relative α -element abundance $[\alpha/M]$ (defined as O, Mg, Si, S, Ca, and Ti changing with solar proportions in lockstep), carbon $[C/M]$, and nitrogen $[N/M]$ abundances. The second step estimates abundances of the following 15 elements: Al, Ca, C, Fe, K, Mg, Mn, Na, Ni, N, O, Si, S, Ti, and V.

These stellar parameters and abundances (for 163,278 stars) are conveniently stored in a single fits file; the current version (Data Release 12) is `allStar-v603.fits`. I played a bit with this file and added my sample analysis code (plain `.py` files – not ipython notebooks!), in files `allStarAnalysis1.py` and `allStarAnalysis2.py`, to repo `2015ASTR511/Project3`. If you download these files and `allStar-v603.fits` data file, you will be able to reproduce figures 1 and 2 here.

This initial analysis is not very deep, and is missing a proper flag selection (I did some ad hoc shortcuts). The plots don’t show it, but I was happy when I managed to reproduce a good agreement between the $[\alpha/Fe]$ value estimated during the first ASPCAP stage, and a corresponding value based on the measured abundances of individual α elements. If the data are read as

```
data = pyfits.open('allStar-v603.fits')[1].data[:,1]
```

then $[\alpha/Fe]$ from the first stage can be accessed as

```
alphaFe = data['PARAM_ALPHA_M']
```

while individual abundances can be accessed as, e.g. for Mg and Ti,

```
MgH = data['MG_H']  
TiH = data['TI_H']
```

³<http://www.sdss.org/dr12/irspec/aspcap/>

The $[\alpha/Fe]$ based on individual abundances can be evaluated as (note that α elements Ne and Ar are missing):

$$[\alpha/Fe]^* = \frac{[O/H] + [Mg/H] + [Si/H] + [S/H] + [Ca/H] + [Ti/H]}{6} - [Fe/H]. \quad (1)$$

In other words, $[\alpha/Fe]$ corresponds to the *mean* value of the abundances of the above six elements, and the last term transforms the scale from “per H ” to “per Fe ”.

The difference of the two estimates of $[\alpha/Fe]$ has a median of -0.008 dex and rms of 0.042 dex, which represents excellent agreement! I also compared the values of $[M/H]$,

```
MH = data['PARAM_M_H']
```

and $[Fe/H]$,

```
FeH = data['Fe_H']
```

and their difference has a median of 0.007 dex and rms of 0.077 dex, which is also excellent.

2.1. Milestones

Based on the above analysis, here are specific few milestones:

1. Reproduce Figures 1 and 2 here, but use proper flag checking.
2. Run the PCA on the vector of 15 abundances for the cleaned sample (about 100,000 stars) and plot the first few eigen components. What do they mean? How much variance is captured by these components? Plot diagrams constructed with eigen coefficients. Do the same separately for high Galactic latitudes ($|b| > 20^\circ$) and low latitudes ($|b| < 5^\circ$). Do the same separately for halo, thick and thin disk populations defined by the cuts based on $[Fe/H]$ and $[\alpha/Fe]$ listed below. Do we see any differences? Can we recognize halo/thick disk/thin disk stars using diagrams constructed with eigen coefficients for the full sample?
3. Based on the PCA analysis, decide whether to run the full EXD.

Halo, thick and thin disk populations can be approximately defined (see the bottom right panel in Figure 1) as:

```
# define thin/thick/halo subsamples
condHalo = (data['PARAM_M_H'] > -2.5) & (data['PARAM_M_H'] < -1.1)
condDisk = (data['PARAM_M_H'] > -0.9) & (data['PARAM_M_H'] < 0.6)
condThin = (condDisk & (data['PARAM_ALPHA_M'] > 0.0) & (data['PARAM_ALPHA_M'] < 0.15))
condThick = (condDisk & (data['PARAM_ALPHA_M'] > 0.15) & (data['PARAM_ALPHA_M'] < 0.35))
```

3. Project DimRed

APOGEE spectra are stored one per star in apStar files⁴. The combined (from several independent observations, or visits) spectra are generated by resampling individual visits onto a common, logarithmically-spaced wavelength scale. The spectra are corrected for each visits derived radial velocity: *the resulting spectra are in rest, vacuum wavelengths!*. Also, these spectra include the barycentric correction (the Earth’s motion is corrected for).

The primary HDU of each file contains an image which gives the two versions of the combined spectrum for each object, as well as the individual visit spectra that went into these combinations. The two versions of the combined spectrum correspond to different weighting schemes (using the square of the signal-to-noise ratio): global weighting and pixel-by-pixel weighting. We will want to work with the version based on pixel-by-pixel weighting. The combined spectra include the continuum (determined using a 500-pixel median filter, with bad pixels masked).

I have not yet played with apStar files. I would suggest to first download and plot a dozen or so spectra to learn technical details before we attempt massive processing. In particular, we will need to understand how to treat pixel masks (bitmasks).

Generally speaking, we will want to process various subsets of the full set. Hence, when writing code to feed tools from astroML with these spectra, we need to develop a “pre-processor” that will collect the required spectra and send the selected subset downstream.

3.1. Milestones

Here are a few milestones to start with

1. Ask the other team for a few examples of stars (say, a dozen each) from halo, thick

⁴See <http://www.sdss.org/dr12/irspec/spectra/>

and thin disk populations. Plot their spectra.

2. Run the PCA method on this small set.
3. If everything went fine, download the entire dataset (about 163,278 apStar files) and run the PCA on all of them. Beware of flags, missing/masked pixels, etc.
4. Plot the first few eigen components. What do they mean? How much variance is captured by these components?
5. Plot diagrams constructed with eigen coefficients. Can we recognize halo/thick disk/thin disk stars (classified using cuts based on $[Fe/H]$ and $[\alpha/Fe]$, see other project) using diagrams constructed with eigen coefficients? That is, color-code points according to population (or use contours of different colors).
6. Based on the PCA analysis, decide whether to run the ICA and NMF analysis.

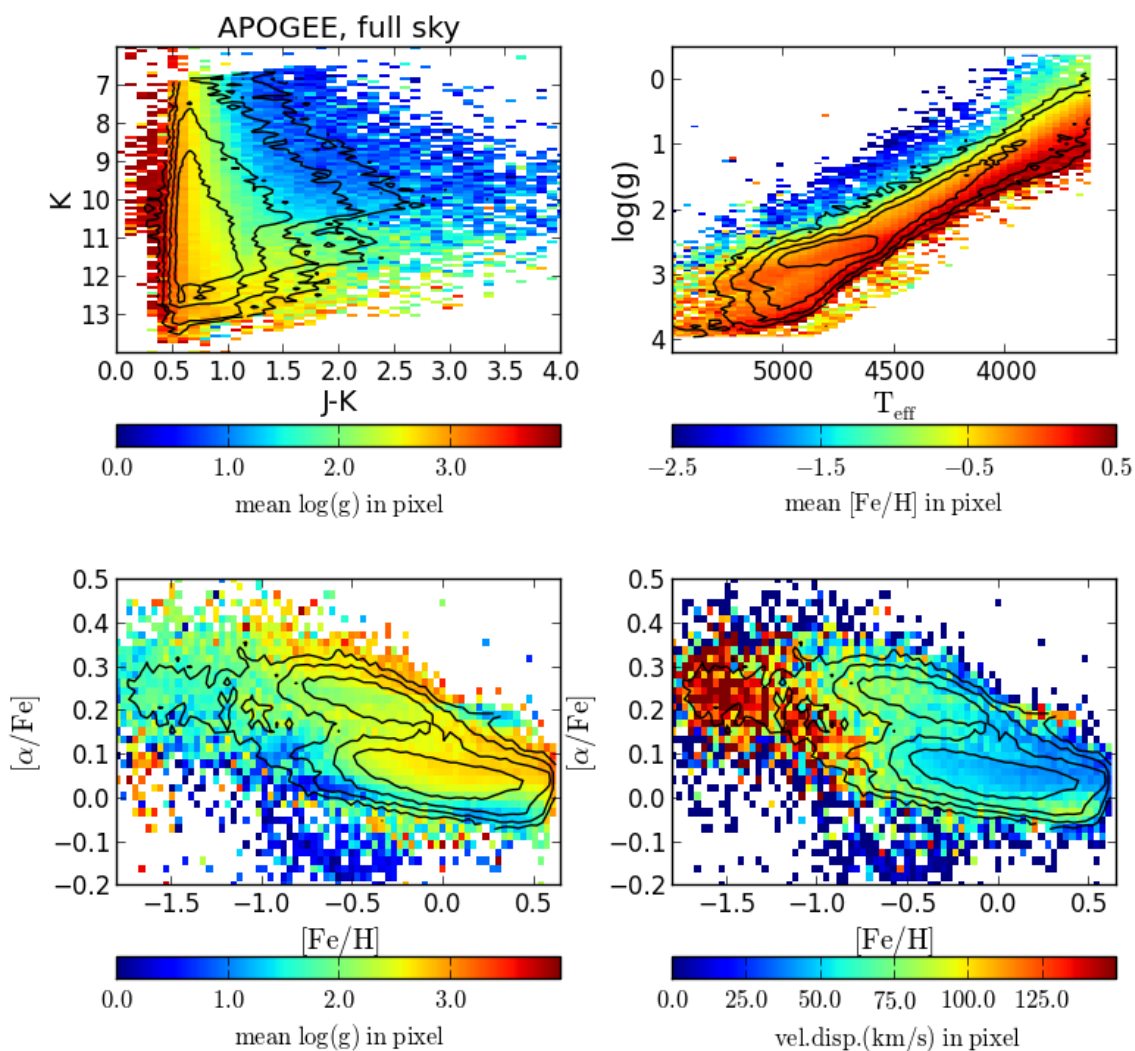


Fig. 1.— This figure was produced with `allStarAnalysis1.py`. It shows various two-dimensional diagrams for a subset of stars from `allStar-v603.fits` data file, color-coded using a third quantity marked below each panel.

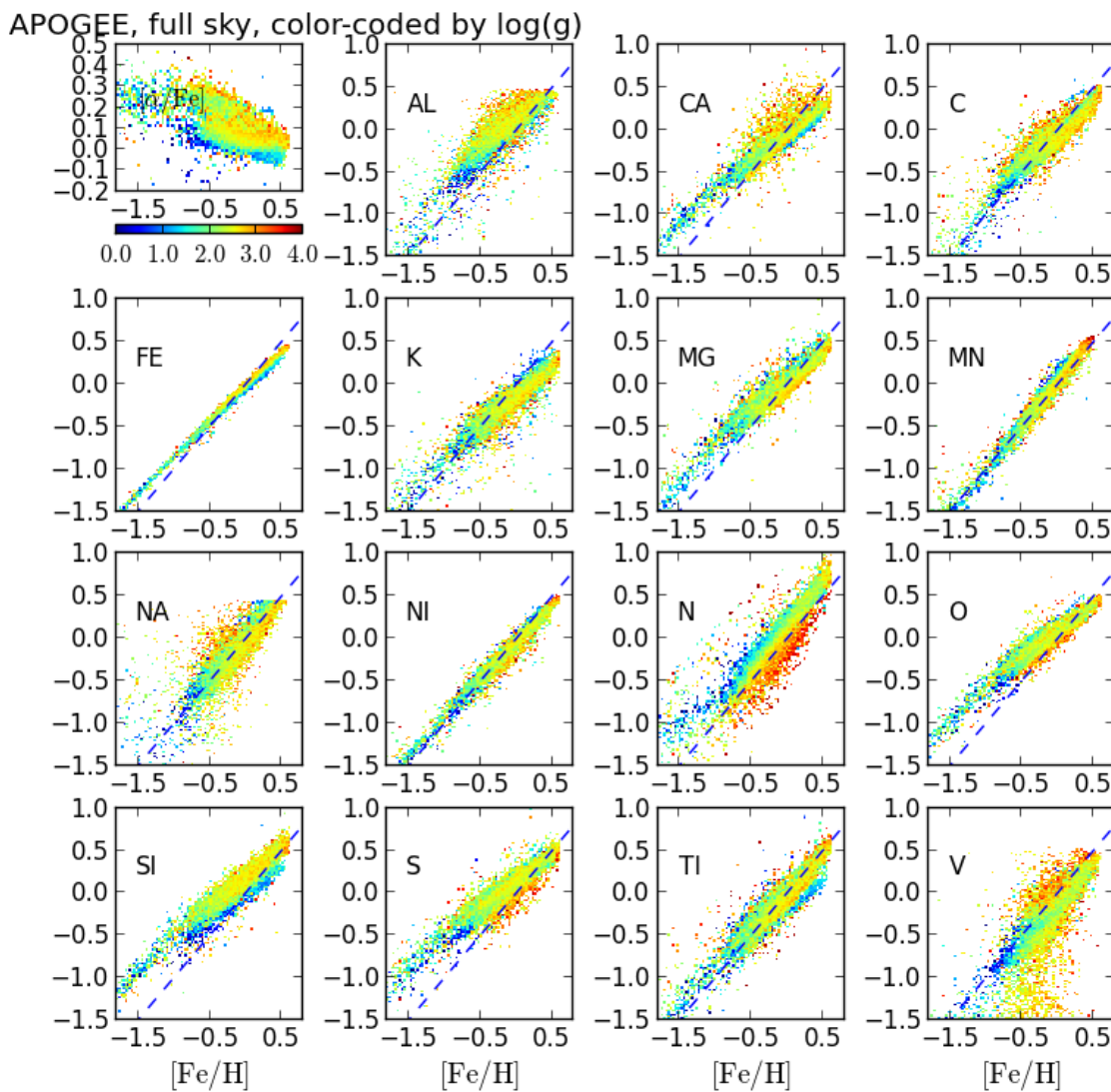


Fig. 2.— This figure was produced with `allStarAnalysis2.py`. The top left panel shows the $[\alpha/Fe]$ vs. $[Fe/H]$ diagram for a subset of stars from `allStar-v603.fits` data file, color-coded using $\log(g)$. The remaining 15 panels show the abundances of 15 individual elements, labeled in each panel, vs. $[Fe/H]$, using the same color-coding as in the top left panel.