The dimensionality of stellar abundance space using red clump stars from APOGEE

Natalie Price-Jones, supervised by Professor Jo Bovy

ABSTRACT

Recent spectroscopic surveys have provided abundant data on stars in the Milky Way. I will analyze a subset of data from the Apache Point Observatory Galactic Evolution Experiment (APOGEE) survey, in the form of red clump stars. These metal-enriched stars offer a complex abundance space to explore, which I will do by studying the spectra themselves, rather than derived abundances. This direct use of the spectra facilitates the characterization of errors and systematic effects in interpretation of absorption features. An understanding of these uncertainties allows accurate investigation of the multi-dimensional abundance space, which may reveal trends in abundances with stellar age and position throughout the Milky Way. Tracing out a star's abundances and these possible trends offers a way to investigate the dynamical and formation history of the Milky Way.

1. Introduction

A star's history, from its formation environment to its migratory motions, can be traced through its composition. Measuring elemental abundances from stellar spectra establishes this composition. However, current methods for calculating these abundances require robust theoretical models and complete understanding of measurement uncertainties. At present, these methods still exhibit systematic trends with other stellar properties, such as effective temperature $T_{\rm eff}$ or surface gravity $\log(g)$ (Holtzman et al. 2015). Removing these contaminating effects from abundance calculations entirely may reveal statistically significant correlations between abundances and stellar age or location within a galaxy.

Recent large-volume spectroscopic survey data from the Apache Point Observatory Galactic Evolution Experiment (APOGEE, Majewski et al. 2015), upcoming data from the Gaia mission (Jordan 2008) and other large spectroscopic surveys will offer rich spectral information about hundreds of thousands of stars. This data is a prime opportunity for abundance analysis. To take advantage of this, I will develop an efficient method for investigating abundance space (Bovy 2015). The method, described in §3, uses spectra directly, rather than quantities derived by fitting multi-parameter model spectra. This reduces complexity in calculations of uncertainty, and is less sensitive to the systematic effects of other stellar properties than traditional methods. This will allow for a more accurate investigation the dimensionality of abundance space for a sample of stars, including probing the relative significance of particular elements. I will apply this method to

spectra from APOGEE and investigate the dimensionality of the sample in abundance space. An understanding of trends in this space may place constraints on what chemical processes contributed to the formation of populations of stars. This in turn may reveal elements of the Milky Way's star formation history or trace how stars moved through the Milky Way after their formation.

2. Data Set

This work will analyze the spectra of a sample of red clump stars from APOGEE's Data Release 12. These stars are selected as belonging to the red clump according to cuts in gravity, effective temperature, and metallicity, and leave us with an initial sample of 19936 stars (Bovy et al. 2014). These stars trace, to some extent, the stellar population of the Milky Way through the volume of the APOGEE survey, which covers a significant fraction of the Milky Ways disk. Data on these stars comes in the form of infrared (H-band) spectra covering a wavelength range from 1.514 μm to 1.696 μm (with some small nm gaps between detectors, Majewski et al. 2015). APOGEE's DR12 pipeline provides data on abundances of 15 elements (C, N, O, Na, Mg, Al, Si, S, K, Ca, Ti, V, Mn, Fe, Ni) as well as effective temperature (T_{eff}), surface gravity ($\log(g)$), overall metallicity (Z), and alpha-element enhancements ($[\alpha/\text{Fe}]$)(Holtzman et al. 2015). I will analyze all 15 of these elements.

3. Methods

Starting with our sample of APOGEE spectra, I will perform empirical fits to remove non-abundance stellar properties. I use spectra processed by the pipeline used to generate stellar parameters and chemical abundances, (ASPCAP - Elia Garcia Perez et al. 2014). These spectra are the combined results of all APOGEE spectra taken for a particular star. I begin by masking the spectra using the bitmask provided by the pipeline

Flux values from each spectrum at a given pixel p ($F(p)$) are fit with a low order polynomial in T_{eff}, $\log(g)$ and Z, so some function for flux is found: $f(p, T_{\text{eff}}, \log(g), Z)$. At each pixel residuals are calculated from the fit: $\delta = F(p) - f(p, T_{\text{eff}}, \log(g), Z)$. We can analyze these residuals directly, using principal component analysis (PCA). This assesses the amount of variance in each dimension of that data. In the case of the residual data, each pixel represents a dimension, with individual stars populating that dimension. Before running PCA with each pixel as an independent dimension we can get a broad sense of the influence of a particular element X by constructing a set of weighted residuals for that element. Residuals for element X pixels are weighted by normalized window functions which trace each X absorption feature and are then summed for each stellar spectrum. Performing PCA on these results reduces the problem from having dimension equal to the number of pixels to having dimensions equal to the number of elements (15). The resulting PCA eigenvectors will show which element (if any) is dominate in determining the residuals after

other stellar parameters have been removed.

I can improve upon the results of classic PCA method (applied to either the pixel or weighted residuals) by using EMPCA (Expectation Maximization PCA). This weights the input data according to its uncertainty (Bailey 2012). In order to use this, I need to find a good set of weights to use. A good place to start is to use the flux measurement uncertainties to serve as weights. However, it seems that for many pixels these uncertainties may be underestimated. To correct for this, I use a sample of stars from open clusters, which should be internally homogeneous in terms of elemental composition. I calculate the residuals in the same way as I described above and take their covaraince. Assuming that the open clusters really are perfectly internally homogeneous (and thus any residual is due to some other source of noise) allows me to use the square root of the terms on the diagonal as a multiplicative factor to correct the uncertainties in the red clump sample. This assumes the pixels are not correlated with each other. These corrected uncertainties can then be used as weights for EMPCA.

4. Timeline

- Nov 2015, Q: Can I see scatter in residuals of individual elements above our level of intrinsic and measurement uncertainty?

    – Characterize measurement scatter by drawing from a Gaussian determined by the uncertainty in flux measurements. [DONE]

    – Fit open clusters (from Mészáros et al. 2015) in T_{eff} according to the outline in §3 and take residuals from this fit as representative of intrinsic abundance scatter. [DONE]

- Dec 2015, Q: What is the dimensionality of abundance space for all elements? Which elements are most significant for distinguishing stars?

    – Use principal component analysis to identify elements of primary importance and the number of dimensions that are not due to noise. [MODIFIED - NOW DOING THIS ON ALL PIXEL SPACE AS WELL - CODE WORKS WITH PRELIMINARY WEIGHT ESTIMATES]

    – Write a code to test existing fitting framework [NEW - DONE]

    – Assess uncertainties from pixel fit. [NEW - CODE EXISTS, NOT YET RUN ON RED CLUMP DATA]

    – Decide on appropriate weights for red clump sample based on results from open clusters. Use open clusters to correct given uncertainties at each pixel [NEW]

- Jan 2016 Q: How can we most accurately find weights for PCA?

    – Trace sources of uncertainty through data (i.e. uncertainty from fitting - how does this translate to uncertainty in open cluster pixel corrections? how in turn does this translate

to uncertainty in red clump residuals, which also have uncertainty from their own fitting? this is necessary for accurate weights). [NEW]

– Use newly more accurate weights in EMPCA in both pixel residual and weighted residual spaces. Use empca.R2(n) where n in the number of the eigenvector to determine where the explained variance is equivalent to the noise variance. How many eigenvectors are needed to explain the variance in each space (pixel vs weighted)? [NEW]

• Feb 2016, Q: Is there any positional dependence of populations in small bins in the remaining dimensions of abundance space?

– Collect stars from small bins in the multidimensional residual space and plot their radial and vertical positions.

– Determine if stars belonging to different bins inhabit different sections of the galaxy, as well as positional variations for stars within a bin.

• Mar/Apr 2016, Q: Can I understand the reasons certain elements may be more significant after PCA by modeling spectra?

– Create model spectra by varying abundances of the 15 elements.

– Perform fitting described in §3 and compare with fitting done on APOGEE data. Ascertain whether model spectra created with derived abundances for the red clump stars can reproduce residual results from the data set.

REFERENCES

Bailey, S. 2012, PASP, 124, 1015

Bovy, J. 2015, ApJ submitted, arXiv:1510.06745

Bovy, J., Nidever, D. L., Rix, H.-W., et al. 2014, ApJ, 790, 127

Elia Garcia Perez, A., Allende-Prieto, C., Cunha, K. M., et al. 2014, in American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts #223, 440.07

Holtzman, J. A., Shetrone, M., Johnson, J. A., et al. 2015, AJ, 150, 148

Jordan, S. 2008, Astronomische Nachrichten, 329, 875

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2015, AJ, submitted, arXiv:1509.05420

Mészáros, S., Martell, S. L., Shetrone, M., et al. 2015, AJ, 149, 153