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 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, offering the potential for chemical tagging. 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_{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. With sufficiently high accuracy in abundance measurements, it may become possible to perform chemical tagging, identifying clusters in abundance data as groups of stars that formed from the same gas cloud but are now dispersed throughout the Galaxy (Freeman & Bland-Hawthorn 2002). However,

Recent large-volume spectroscopic survey data from the Apache Point Observatory Galactic Evolution Experiment (APOGEE, Majewski et al. 2015), upcoming data from Gaia (Jordan 2008) and Galactic Archaeology with HERMES (De Silva et al. 2015), and other large spectroscopic surveys will offer rich spectral information about hundreds of thousands of stars. The quantity of stars provides a well populated abundance space. To take advantage of this, I develop an efficient method for determining important dimensions of abundance space based on the work of 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 model fitting methods. Thus I perform a more accurate investigation the dimensionality of abundance space for a sample of stars, including probing the relative significance of particular elements. I apply this method to red clump star spectra from APOGEE, while using open cluster stars with low abundance space scatter as a calibration sample . An understanding of trends in this space may place constraints on what chemical processes contributed to the formation of populations of stars, and clusters of stars in abundance space may correspond to co-natal groups. Identifying these groups through chemical tagging and understand formation processes may reveal details of the Milky Way's star formation history and statistically 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 ($T_{\rm eff}$), surface gravity ($\log(g)$), overall metallicity (Z), and alpha-element enhancements ($[\alpha/\rm Fe]$)(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 pixels in the spectra using the bitmask provided by the pipeline. Bits set in the bitmask indicate particular problems with individual pixels in that spectrum, including flags for the effects of cosmic rays, pixel saturation, atmospheric absorption, poor quality calibration files or inaccurate flux uncertainty estimations. I chose to mask on all of the effects listed as well as on pixels with a signal to noise ratio (SNR) below 50. I then perform a correction on pixels where uncertainties seem underestimated, where the SNR > 200, increasing the uncertainties calculated by ASPCAP at those pixels such that the new SNR does not exceed 200 at any pixel. The result is a masked set of spectra, and a set of corrected uncertainties to which the mask on the spectra is applied.

After masking the spectra and making corrections to the uncertainties, the unmasked flux values $(F(p))$ at a given pixel p for each star are weighted by their uncertainties and fit with a second order polynomial in T_{eff}, $\log(g)$ and Z. Some function for flux is found: $f(p, T_{\text{eff}}, \log(g), Z)$ and residuals are calculated from this fit function: $\delta = F(p) - f(p, T_{\text{eff}}, \log(g), Z)$. The mask applied to the input spectra is then applied to these residuals.

The residuals from the fit comprise the abundance space I wish to investigate. Trends relating to the bulk properties of the star have been removed by the polynomial fit in T_{eff}, $\log(g)$ and Z. Any deviation from the fit that remains must be attributed to differences in stellar abundances. In its raw form, the data δ populates a space in which each pixel is an apparently independent dimension. We can analyze these residuals directly, using expectation maximized principal component analysis (EMPCA - Bailey 2012). This algorithm assesses which dimensions of a data set contain the most variance, and produces eigenvectors ranked by how much of the data's variance they can explain. Applying EMPCA to the residuals returns eigenvectors with 7214 elements (the number of pixels in the input spectra). Pixels at which these eigenvectors have a large magnitude highlight wavelengths at which stellar flux (i.e. the depth of absorption features) varies the most across the sample. Determining which absorption features are most important for describing variance tells me which elements are best at distinguishing individual stars.

However, individual pixels in a stellar spectra are not uncorrelated and do not represent independent dimensions in abundance space. In particular, pixels at which different absorption features for the same element occur will be correlated with each other. To account for this, and with additional effect of making the data set more easily tractable on a human level, a parallel set of approaches was used to reduce the dimension of abundance space. In one stream, EMPCA was applied directly to the residuals δ. The resulting eigenvectors were weighted according window functions describing absorption features for the 15 elements identified in §2. In an alternate stream, the residuals were first weighted by these window functions, then EMPCA was applied. The results are in A FIGURE.

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

De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, MNRAS, 449, 2604

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

Freeman, K., & Bland-Hawthorn, J. 2002, ARA&A, 40, 487

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