

CTA200 Final

Isabelle Laing

May 2024

Project Overview

My project for CTA200H was to gain an introductory understanding on:

- How to plot data from the Dr. Ti Ling's server and access the DESI/APOGEE files
- How to find answers as to why my plots look the way that they do

To do this, I reproduced some of the figures created by 2023 SURP student Stephanie Lee and shown in her final SURP Presentation. Within this project, I was expected to do the following:

- Access the server where the data is stored
- Make DESI and APOGEE plots for metallicity and for $\log(g)/T_{\text{eff}}$
- Compare the plots between DESI and APOGEE
- Make inferences on the real stars being used based on the plots (involving research)
- Understand why the results have the structure that they have
- Look at a real DESI spectra for a specific star
- Find chemical abundances within that spectrum
- Make inferences about that star's properties
- Explain and write up my findings in a report

In fact, one of my main goals for this summer is to improve upon my scientific writing in English, as I completed my entire education in French and struggle to properly structure my writing in English.

Script Execution

My script begins by importing necessary packages and looking at the private symbolic link that the DESI spectra data falls under. By typing `ls...`, I was able to look at all the possible files and I found the one I was looking for. My next cell opens that `.fits` file and lists all the columns of data, which was necessary so I knew the keys I would need to input in order to access the numbers and plot them. To open the `fits` files, I used `astropy.io`. At every cell I rewrote the path to find the file and opened it anew so any cell would run immediately upon opening the folder to make my life easier. After that, it is just plot after plot making figures that I will discuss during the length of this report. I have left text and comments to explain my thought process while I coded. Now to the scientific report section of my project!

Final Project: the full report

Introduction

In an effort to build a timeline of the Milky Way's evolution and formation, the field of Galactic Paleontology has emerged. Mostly since the beginning of the 21st century, this field revolves around different stellar populations and treats them like living fossils, telling of a chronological history to be uncovered. A major contributor to these studies are the surveys that observe these stars, and two prominent surveys that help are DESI and APOGEE. The former, the Dark Energy Spectroscopic Instrument, is based in Arizona and started publishing very recently, as of only 2021. Meanwhile, the Apache Point Observatory Galactic Evolution Experiment is a part of the Sloan Digital Sky Survey (SDSS), an ongoing project that began in 2000. Both surveys play major roles in active research in galactic Palaeontology, however looking at the data they produce individually does result in major differences that are telling of how long they've been running, respectively. By looking at multiple stellar parameters, at the functionality of these surveys and at individual stars, the long history of our Milky Way and how it is studied begins to reveal itself. The goal of this paper is to report how DESI and APOGEE data can be used within a Jupyter Notebook to unveil details of data collection methods and of stars in the Milky Way as a first step towards understanding Galactic Paleontology.

All of the data used going forwards for figures and interpretation was pulled from the University of Toronto research group lead by Dr Ting Li, made available to the team on the eridanus server accessible on personal devices but restricted to the public. These figures are not reproducible without access to that proprietary data. For the most part, as an introductory project, the figures being created are meant to be reproductions of figures made in the summer of 2023 by SURP student Stephanie Lee. The main difference between these graphs and those is that the data has been added to and improved upon by the surveys

being pulled from. To reproduce these graphs, the same axis ranges were used, and will be used in the code as well. This explains any similarities to past SURP research and decisions made during the plotting process.

Stellar Parameters

As a starting place, stellar parameters of stars such as their metallicity, their surface gravity and their effective temperatures make room for lots of interpretation about the data given and what it gives away concerning their home galaxy. This is important as stars accumulate characteristics based on two main markers: their age and the environment they were born in. Take for example how metal rich a star is; the metallicity of a star is very dependent on whether that star formed naturally through a calm process, leaving it metal-rich, or whether it formed under turbulence and mergers, which would leave it metal-poor. Unfortunately, studying the quantitative abundance of a star is extremely difficult, and it is very rare for a survey to put out numbers about the “amount” of any element in a star. So, metallicity is instead defined by ratios. Generally, metallicity is defined by how abundant elements heavier than helium and hydrogen are in a star’s atmosphere. Here, it will be defined by the abundance of iron compared to hydrogen, denoted by Fe/H . That ratio is then taken over the ratio for the Sun, which is 0.012, and finally we take the log base 10 of that number so the range is less large and easier to plot, creating more consistent visuals. So the final equation to define metallicity, or $[\text{Fe}/\text{H}]$ of a star is:

$$[\text{Fe}/\text{H}] = \log_{10} \left(\frac{\text{Fe}_{\text{star}}/\text{H}_{\text{star}}}{\text{Fe}_{\odot}/\text{H}_{\odot}} \right)$$

As such, a number equal to zero would indicate that the ratio was equal to 1, or that the star has the exact same metallicity as the sun. The Sun is considered relatively metal-rich, and as such it is very common for the metallicity of a star to be negative, indicating it is more metal-poor than our sun. A number higher than zero would make the star more metal-rich than the Sun. The plot used to show this number is a histogram, which counts upwards for the frequency of a given star having a given metallicity number. By looking at this graph, we can see how many stars are equally as metal-rich as the Sun, more-so, or less-so. By plotting both DESI and APOGEE data, we can see how accurate their information is.

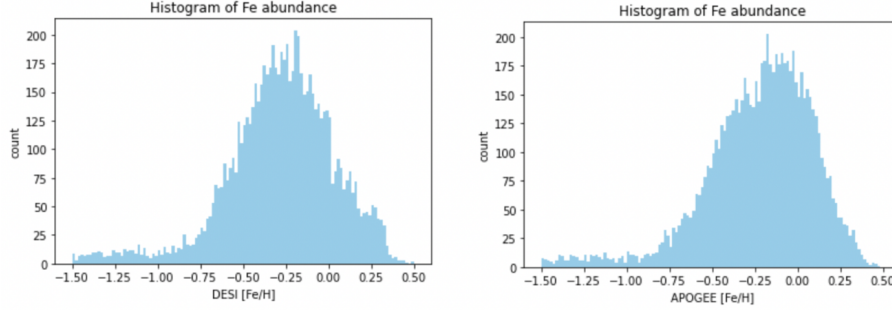


Figure 1: $[\text{Fe}/\text{H}]$ from DESI and APOGEE

In terms of information presented by both graphs, we can see the most common result finds itself at just below 0 until 0, so metal-rich stars that are just poorer than the Sun, which is the case for most main sequence stars. This tells us most stars in the Milky Way are metal-rich, obviously however the sample here is not representative of every star in the Milky Way. However there is no doubt that the large majority of stars in the Milky Way are red dwarfs, which are indeed metal-rich. So while this data is a generalisation of a diverse stellar population, it is nonetheless correct to identify the large frequency of metal-rich stars.

Looking into the differences of the two graphs however tells us a whole new story about which survey produces better data. To make this more clear, the other graph shows just the outline at the top of the bins on the histogram.

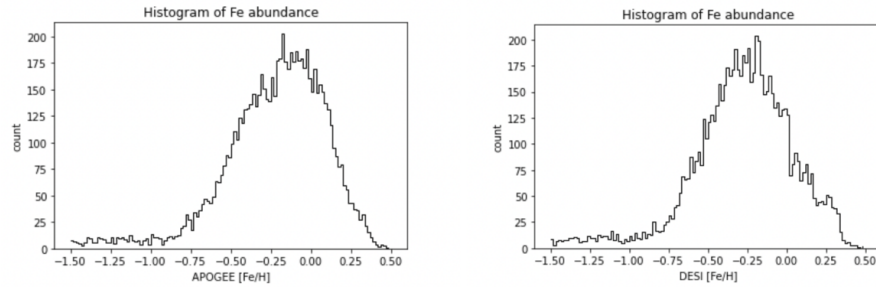


Figure 2: $[\text{Fe}/\text{H}]$ but in black and white and it's transparent

Here, it becomes incredibly obvious that the DESI data is choppy and has strange moments of inconsistency. The climb up to the highest count is far smoother in the APOGEE data, with fewer leaps up and down again as seen in DESI's data. Then, and most obviously, on the comedown, DESI has multiple instances of the data lowering in count dramatically then rising again until it gets down to near 0. The APOGEE data largely avoids this issue, with bars go-

ing down one after the other at almost every bin. The jarring appearance of the DESI data is the beginning of a trend this report will cover where the younger and less perfected DESI survey will not quite meet up to its elder counterpart. The faults observed are a mix of human error, systematic issues and perhaps machine learning mistakes. It is not possible to identify the root cause of every issue, so going forward assume the issues are of those types.

The next two parameters are closely connected, and form a very interesting and telling relationship, and for that reason are placed together as the axes of a scatter plot.

The effective temperature of a star is defined as such; take a star and imagine it as an idealised blackbody object, in which case it absorbs all incident electromagnetic radiation, and then take its temperature in Kelvin units. When we take the ‘raw’ or ‘natural’ temperature of a star, it is inflated by that radiation and that inflation rate is never consistent with stars, being influenced by plenty of factors. By taking T_{eff} , there is an established validity to comparing these stars. The Sun, for example, has a $T_{\text{eff}} = 5778$ K. Moving on, the surface gravity of a star, or $\log(g)$ of that star, is the gravitational acceleration experienced at a star’s surface at its equator (including the effects of rotation). Think of it as the acceleration due to gravity on a star on its surface, in cgs units. However, we use the log function with base 10 as between a neutron star and a giant, the range of gravity is insanely high, spanning several magnitudes. By using \log_{10} , we can express that whole range without a condensed axis. If surface gravity is equal to zero, then $g = 1$, which is the gravity on Earth. As such, it will almost always be higher than zero. When we plot T_{eff} on the x axis and $\log(g)$ on the y axis, a very distinct structure forms, regardless of which survey we observe.

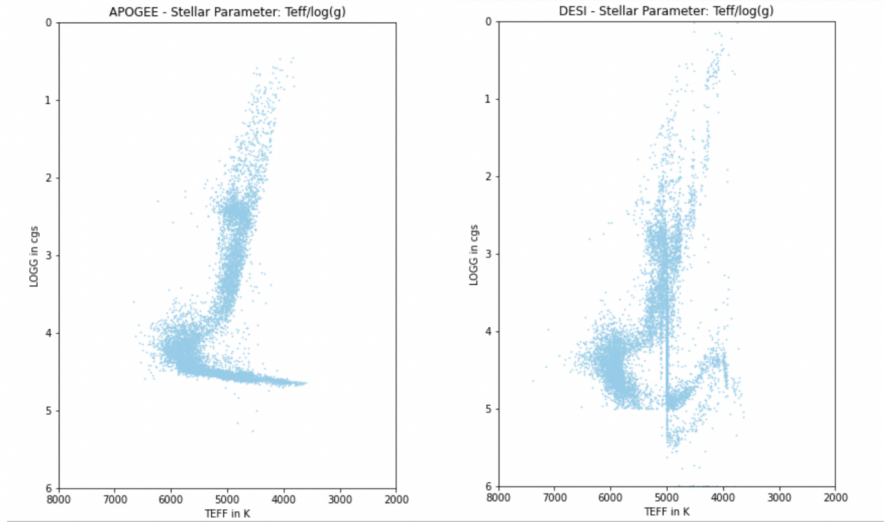


Figure 3: The $T_{\text{eff}}/\text{Log}(g)$ relationship

Immediately, the shape is reminiscent of the Hertzsprung-Russell Star Diagram, a scatter plot of stars showing the relationship between the stars' absolute magnitudes or luminosities and their stellar classifications or effective temperatures [1]. More specifically, the Hertzsprung gap seems to have formed about the curve that this 'hook' shaped plot gives. In fact, plots that showed the same relationship referred to their plots as a "spectroscopic H-R" diagram [2]. However, that plot was only using red stars, whilst ours used a selection of 7910 stars within the Milky Way. In no coincidence at all, the shape is still distinctly there. From this, it seems obvious that the majority of stars being plotted are main sequence stars as well. And by looking at an H-R diagram, and by observing the path of main sequence stars to giants, there is a familiar shape once again. In fact, by assuming this graph is plotting data from a majority of main sequence stars, it would be entirely possible to predict its shape. Main sequence stars have a specific relationship between their surface gravity and effective temperature: the surface gravity decreases when the surface gravity is increasing [3]. For the bottom part that creates a slightly diagonal line from 3000K to 6000K, there is a slow mount upwards explained by this fact. As for the stars above $\log(g) = 4$, look towards the common red star; red giants typically have a surface temperature around 5000K and weak surface gravity levels. So as a star transitions from main sequence to a dying red star, and notably the red giants in the Milky Way, their trajectory is mapped on H-R diagrams as well as the above. That explains the exact structure of these graphs.

As for the differences between them, there's no surprise that the DESI data produces strange results. There is an awkward line at exactly 5000K that seems completely solid, and not the results of actual data points. Furthermore, a square corner somehow appeared at the base of the hook. There are also stars at null $\log(g)$, which feels very unlikely. Again, this is not unexpected and is likely the result of some error within the processing or collection of the data, human or machine via a human. Either way, the most important part of this graph, its structure, is respected by the DESI data and these graphs have provided more of the same in terms of insight into the Milky Way, a galaxy filled with mains sequence stars and giants. However, considering the degree of which APOGEE provides better data, it would be of interest to look into how they collect their data.

APOGEE Insights: a Mollweide Projection

Within the provided data, there are two columns that give information on the location of the stars being observed, and more specifically, their coordinates. Stellar coordinates were a tricky task to complete as it required picking a way to locate stars despite the Earth's tilt, spin, orbit and the movement of our Solar System throughout its galaxy. In short, the perspective of these telescopes is constantly moving and rotating through space. Furthermore, telescopes are all over Earth, at different heights and angles on the planet. As such, the RA/DEC system [4] was developed; RA, or right-ascension, is how far 'right' an object is

from one specific spot in terms of degrees (or as a common alternative, hours). The minimum is 0 degrees, or at the point directly, to a maximum of 360 degrees wherein the star can be so far to the ‘right’ it is a full turn just to the left of spot 0. That null spot is the vernal equinox. Then the DEC, or declination, is how far up or down from that equinox that the star is located, with 180 degrees of tilt up and down (maximum 90 degrees upwards and minimum -90 degrees downwards). All stars find themselves a certain degree to the right then up and down, giving them their RA/DEC coordinates. In theory, by plotting the coordinates of every one of the 7910 stars in the APOGEE data, we’ll be able to gain information on how the SDSS collects data for this file.

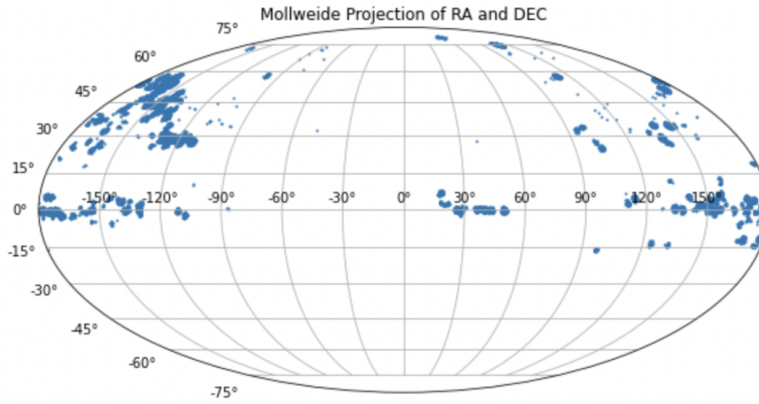


Figure 4: Visual Projection of APOGEE observations in the night sky (RA/DEC coords system)

Once plotted, this figure is very telling; all the data is collected in small circles that can easily be singled out, looking like freckles on a map. Given how much data is in this file, it goes without saying that the spots are not each a star (there would be significantly more spots if that were the case). Rather, assume the spots are each a collection of stars in proximity to one and other. A quick research into the SDSS press releases confirms as much. [5] According to one press release, to get their observations, APOGEE technicians and telescope operators or engineers drill holes into a large aluminium sheet placed at the back of the telescope. Then, optical fibres are carefully plugged by hand into each hole and those carry the light from the stars observed into a spectrograph. Through this method, observers can single out stars or regions of stars that they wish to observe as individual objects, leading to thousands of holes in each individual sheet. As such, all the data will come out in small circles, spread across the entire night sky. This ability to single out sectors leads to immense data and accuracy in the observations made by SDSS and as such, the APOGEE files are

brilliant archives of information.

DESI: Spectrum of a Single Star

After having looked at so many general cases and overall plots for stars, it would be interesting to look into the spectrum of one single star, and see what it can tell us. Within the DESI file, there are three different wavelengths, with a range of wavelengths that are all connected: blue, red and near-infrared. These cover from about 3500 Å to 10000 Å, providing lots of information about any specific star. The spectrum of a star is achieved by DESI by each of DESI's 10 identical spectrographs. Similar to the optic fibres discussed with APOGEE, here the spectrographs accept one fibre-optic cable with 500 fibres. The light from that cable enters the spectrograph through a slit and the light is split using dichroics into three cameras, each sensitive to different wavelength ranges, which explains the three columns of data [6]. To plot these spectra, the data in angstroms will make up the x-axis and the y-axis will represent the flux. Here, the script provides four plots, one for each colour and then one connecting them all, giving a full spectrum. Boringly, this plot is observing the first star in the data file.

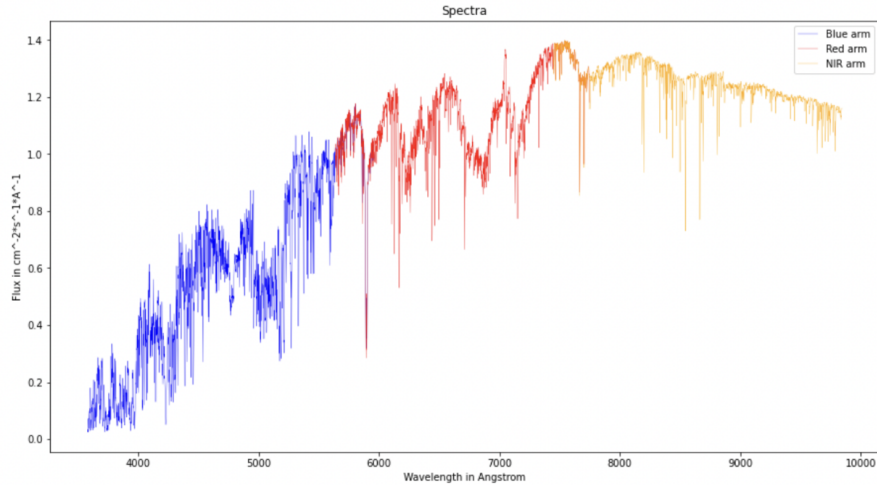


Figure 5: Spectra of non-identified star using DESI data

The best way to understand what a spectrum is telling is by looking at specific abundances noticeable in sudden spikes downwards in the graph, of which there are plenty here. Using a guide by SDSS, the following is unveiled [7]:

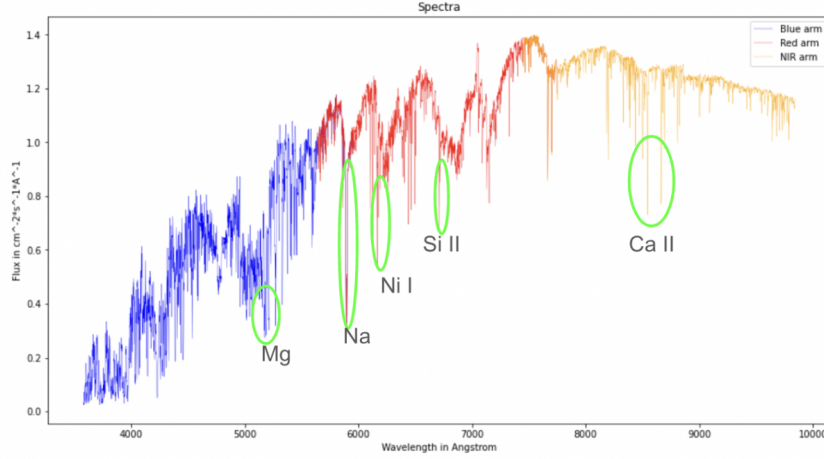


Figure 6: Spectrum of star 0 with identified and labelled a

Clearly, this is a metal-rich star, showing an abundance in magnesium, sodium, nitrate I, silicon II and calcium II. That continues the never ending train of metal-rich, younger stars in the Milky Way. Spectral imaging of stars provides information about metallicity, temperature, and information on the magnetic field of a star. The only issue with a figure like the one above is that the abundances noted are all based on the deduction of an undergraduate student at the University of Toronto who wasn't wearing her contact lenses and was tired from report writing. In sum, there is likely some mistake there caused by human error. That is why the University of Toronto astrophysics department has put a large emphasis on training neural networks to essentially pick out dips that indicate presence of specific elements. This technique is still being developed this summer by SURP students, and will likely continue to improve. Nonetheless, we can see there are numerous abundances, backing up the main results discussed so far.

Conclusion for this Project

The main goal of this project was for me to gain a better understanding of the data I'm working with. Combined with a paper I read going into my first meeting, I now hold a much more complete understanding of the main vocabulary surrounding this project and main concepts such as stellar populations, metallicity and other stellar parameters, what makes different stars unique, and what our Milky Way is composed of. Furthermore, I was able to better understand the sources of the data I will be using throughout the summer, DESI, and how it compares to APOGEE, concluding that the newer DESI data is imperfect, which will lead to many sanity checks with the results I end up working with.

Another huge aspect of this project was making me more comfortable with plotting using python and navigating the data within my terminal and notebooks. This comes across less in my script as it is commented out, but accessing the data took a solid two days before I even started plotting. Overall, this project was an amazing introduction to the material I'll need to understand this summer and it was a delight to try my hand at this project.

References

- [1] Australia Telescope National Facility, “Introduction to the Hertzsprung-Russell Diagram,” Csiro.au, May 08, 2019. https://www.atnf.csiro.au/outreach/education/senior/astrophysics/stellarevolution_hrintro.html
- [2] N. Langer and R. P. Kudritzki, “The spectroscopic Hertzsprung-Russell diagram,” *Astronomy & Astrophysics*, vol. 564, p. A52, Apr. 2014, doi: <https://doi.org/10.1051/0004-6361/201423374>.
- [3] T. Angelov, “1996BABel.154...13A Page 13,” adsabs.harvard.edu, Sep. 24, 1996. <https://adsabs.harvard.edu/full/1996BABel.154...13A>
- [4] SDSS, “Mapping the Sky – Voyages,” SDSS Voyages. <https://voyages.sdss.org/preflight/locating-objects/ra-dec/#:~:text=Reference%20Lines%20in%20the%20Sky&text=Lines%20of%20longitude%20are%20now>
- [5] SDSS Press Releases, “Serving up the Universe on a plate — SDSS — Press Releases,” press.sdss.org, Jul. 14, 2021. <https://press.sdss.org/serving-up-the-universe-on-a-plate/>
- [6] US Dept of Energy Office of Science, “spectrograph,” www.desi.lbl.gov. <https://www.desi.lbl.gov/spectrograph/>
- [7] SDSS, “Table of Spectral Lines Used in SDSS,” classic.sdss.org. <https://classic.sdss.org/dr6/algorithms/linestable.php>