

# CTA200: Final Project

Rosayla Coulthard

May 2024

**Brief Note:** Since the DESI data is private and cannot be posted on GitHub, mock DESI data has been generated for the purpose of this assignment. This data can be found in student8's home directory on the CITA computers under the name 'DESI\_EDR\_DDPAINE MOCK.csv'. The name of the file path may need to be modified.

## 1 Part 1: Preparing The Data

### 1.1 Skim through Zhang et al. and read in the DESI DD-Payne catalog. What stellar atmospheric parameters and elemental abundances did they measure? How many stars with good measurements are included in this catalog?

In Zhang et al. 3 stellar atmospheric parameters are being measured:

- $T_{eff}$ : The effective temperature
- $logg$ : The surface gravity
- $v_{mic}$ : The micro turbulence velocity

The paper also measures 12 elemental abundances:  $[Fe/H]$ ,  $[C/Fe]$ ,  $[N/Fe]$ ,  $[Mg/Fe]$ ,  $[O/Fe]$ ,  $[Al/Fe]$ ,  $[Si/Fe]$ ,  $[Ca/Fe]$ ,  $[Ti/Fe]$ ,  $[Cr/Fe]$ ,  $[Mn/Fe]$ ,  $[Ni/Fe]$ . Within the DESI data, there are 22583 stars, none of which appear to have missing values or values set as NA. There are also no additional notes on the measurements themselves, implying that all 22583 stars have good measurements.

**1.2 Look up the following quantities of Sextans in the Local Volume Database: Right Ascension (RA), Declination (Dec), and half-light radius (rh). From the DESI catalog, create a subset of stars that are within 20 half-light radii of the center of Sextans. How many stars are included in this sample?**

Within the Local Volume Database ([https://github.com/apace7/local\\_volume\\_database](https://github.com/apace7/local_volume_database)), the row for Sextans can be found in data: dwarf\_all.csv: line 50.

Right ascension: 153.2628 deg

Declination: -1.6133 deg

Half-light-radius: 16.5 pc

In order to find stars within 20 half-light radii, a conical shape can be constructed outwards from the Earth to Sextans. Where the height of the cone is the distance from the Earth-Sextans (85.9 kpc), and the radius of the cone is 20 half-light radii of Sextans (0.33 kpc). Then by performing simple trigonometry, a maximum angle of separation between the Earth-star line and the Earth-Sextans line of 0.0038 radians can be obtained. All stars whose angle from the Earth-Sextans line is less than or equal to 0.0038 radians will be within 20 half-light radii and will be considered part of the subset. The SkyCoord package contains a function called separation that calculates this angular separation when two sky coordinates are provided. The code written utilizes this to evaluate the angular separation of each star and, provided it is less than or equal to 0.0038 rads, adds it to a subset database. The number of stars in this database is 74.

**1.3 Using the Astroquery python package, download and combine the data in Tables 2, 6, and 10–12 of Theler et al. (2020) from VizieR. How many stars does this catalog have?**

Upon downloading, merging, and then removing duplicate columns, there are 112 stars in the catalogue.

**1.4 Using the Astropy python package, cross-match the two catalogs. Justify your separation criteria. How many stars are in both catalogs?**

To be able to use the Astropy package to cross-match the catalogues, both need to be converted from pandas tables to Astropy tables. From there, the RA and DEC coordinates for each catalogue are put into usable form for Astropy. Then, based on an arbitrary separation criteria, the catalogues are cross-matched to become one catalogue containing only overlapping stars. Typically a separation criteria of 0.3 - 1.0 arc seconds is ideal. When 0.3 is used, only 3 stars remain.

Thus, to have enough data points to work with for further questions, a separation of 1.0 is used; 6 stars remain in the cross-matched catalogue. This is a large jump from the original 22583 stars in DESI and 112 in the Theler paper.

- 1.5 Visualize your final sample by creating a figure illustrating the spatial distribution (RA vs. Dec) of each catalog. Choose marker shapes and colors that make it easy to see which stars come from each catalog and which stars are included in both.

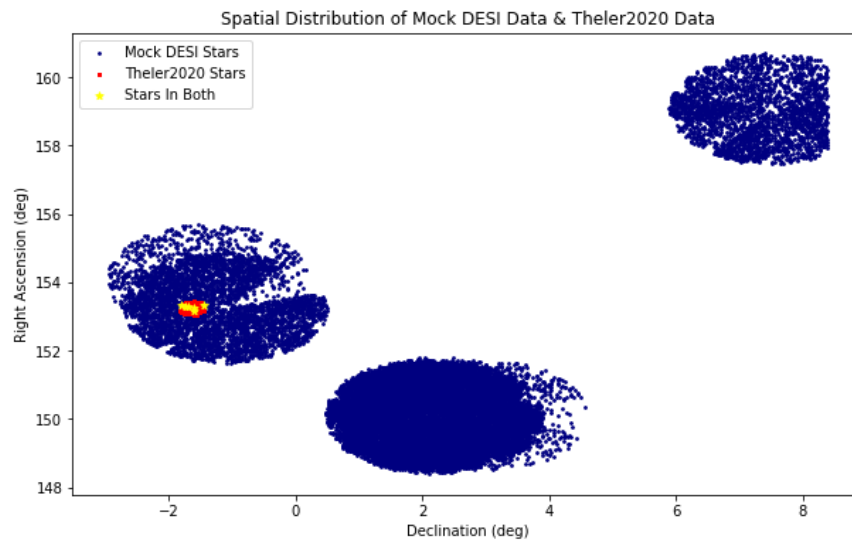


Figure 1: Spatial Distribution of Mock DESI Data and Theler2020 Data

## 2 Part 2: Validation and Calibration

**2.1 For each stellar parameter and elemental abundance that the two catalogs have in common, plot the differences between the two catalogs for the cross-matched sample using  $[\text{Fe}/\text{H}]$  from the Theler et al. catalog as the x-axis. Be sure to include appropriate error bars. Add a horizontal line at a value of zero for reference. At first glance, how does the agreement between the catalogs look for each quantity? Is there an obvious offset? A linear trend? Large scatter?**

Firstly, the common elemental abundances can be determined by examining the columns of both catalogues:

Theler2020 abundances:  $[\text{Fe}/\text{H}]$ ,  $[\text{Mg}/\text{Fe}]$ ,  $[\text{Ca}/\text{Fe}]$ ,  $[\text{Sc}/\text{Fe}]$ ,  $[\text{TiI}/\text{Fe}]$ ,  $[\text{TiII}/\text{Fe}]$ ,  $[\text{Cr}/\text{Fe}]$ ,  $[\text{Mn}/\text{Fe}]$ ,  $[\text{Co}/\text{Fe}]$ ,  $[\text{Ni}/\text{Fe}]$ ,  $[\text{Ba}/\text{Fe}]$ ,  $[\text{Eu}/\text{Fe}]$

DESI abundances:  $[\text{Fe}/\text{H}]$ ,  $[\text{C}/\text{Fe}]$ ,  $[\text{N}/\text{Fe}]$ ,  $[\text{Mg}/\text{Fe}]$ ,  $[\text{Al}/\text{Fe}]$ ,  $[\text{Si}/\text{Fe}]$ ,  $[\text{Ca}/\text{Fe}]$ ,  $[\text{Ti}/\text{Fe}]$ ,  $[\text{Cr}/\text{Fe}]$ ,  $[\text{Mn}/\text{Fe}]$ ,  $[\text{Ni}/\text{Fe}]$

The abundances in common are:  $[\text{Fe}/\text{H}]$ ,  $[\text{Mg}/\text{Fe}]$ ,  $[\text{Ca}/\text{Fe}]$ ,  $[\text{Ti}/\text{Fe}]$ ,  $[\text{Cr}/\text{Fe}]$ ,  $[\text{Mn}/\text{Fe}]$ ,  $[\text{Ni}/\text{Fe}]$ . While both catalogues have some form of  $[\text{Ti}/\text{Fe}]$  since DESI does not specify the specific isotopes, these abundances will not be compared. To create a table of differences between the two catalogues, the difference of the absolute values of each abundance was taken. Additionally, to properly propagate the errors in the measurements, the errors of the differences were found by taking the sum of the errors from each catalogue. Lastly, the plots were created. At a glance, the quantities seem to agree between the catalogues with the exception of  $[\text{Fe}/\text{H}]$  and  $[\text{Ca}/\text{Fe}]$ . For these abundances, stars with lower metallicities appear to be responsible for the disagreement between the catalogues. As the metallicity increases, the differences follow a downward trend, with  $[\text{Fe}/\text{H}]$  decreasing linearly and  $[\text{Ca}/\text{Fe}]$  decreasing almost exponentially. While some of the other quantities do appear to be different, this may be explained by their error bars.

Overall, since there are only 5 stars being displayed, the trends may not be representative of all of the data if all 112 stars were cross-matched.

**$[\text{Fe}/\text{H}]$  differences:** The differences between the catalogues appears to disagree for stars with lower metallicity; the data is quite scattered. The differences decrease linearly as metallicity increases.

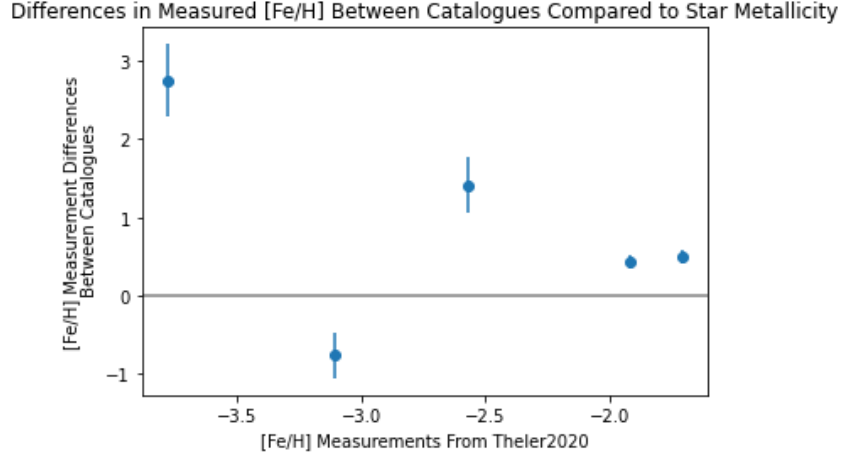


Figure 2: [Fe/H] Differences / Measured Metallicity

**[Mg/Fe] differences:** The differences between catalogues does not appear to follow a trend and remains somewhat constant around the zero line. The data, however, does appear to be quite scattered around this line.

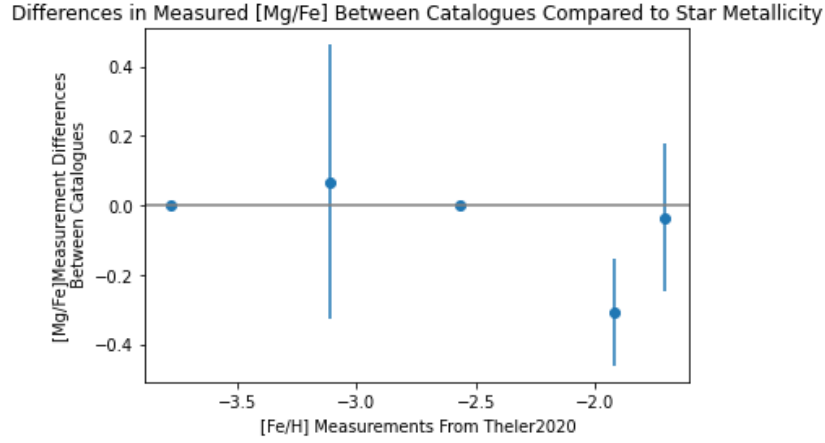


Figure 3: [Mg/Fe] Differences / Measured Metallicity

**[Ca/Fe] differences:** Differences between the catalogues appears to decrease exponentially as metallicity increases.

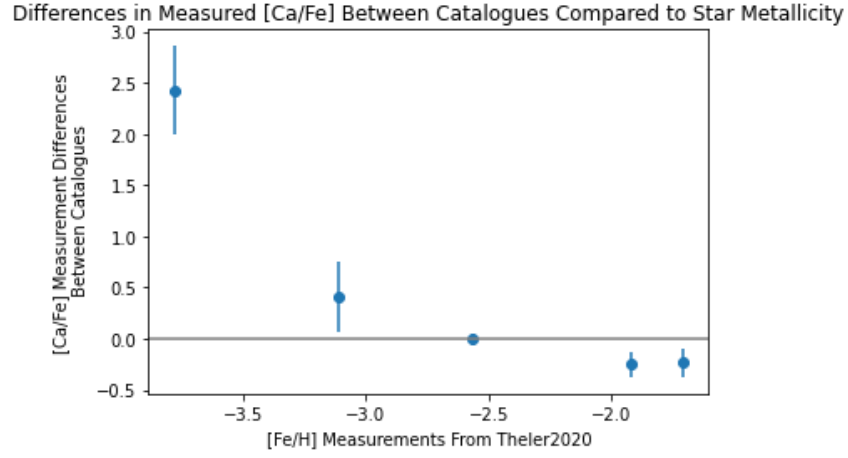


Figure 4:  $[\text{Ca}/\text{Fe}]$  Differences / Measured Metallicity

**$[\text{Cr}/\text{Fe}]$  differences:** The differences appear to be centered very closely around the zero line, with the exception of higher metallicity stars. However, for those stars, their error bars still pass through the zero line.

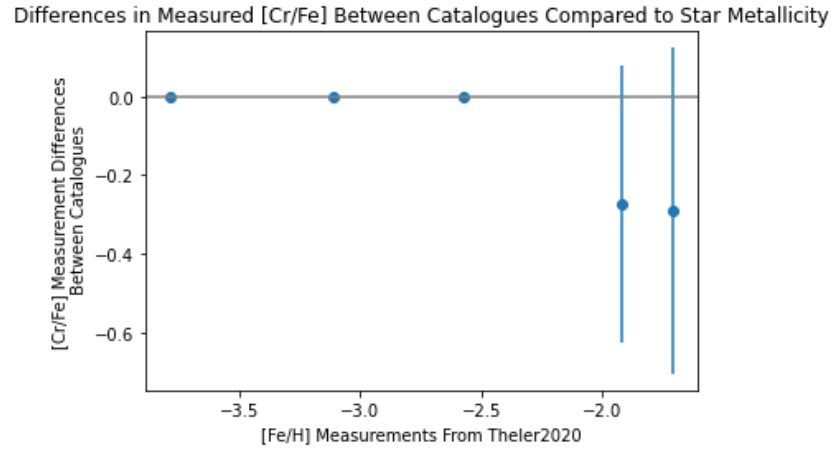


Figure 5:  $[\text{Cr}/\text{Fe}]$  Differences / Measured Metallicity

**$[\text{Mn}/\text{Fe}]$  differences:** Differences very closely centered around zero.

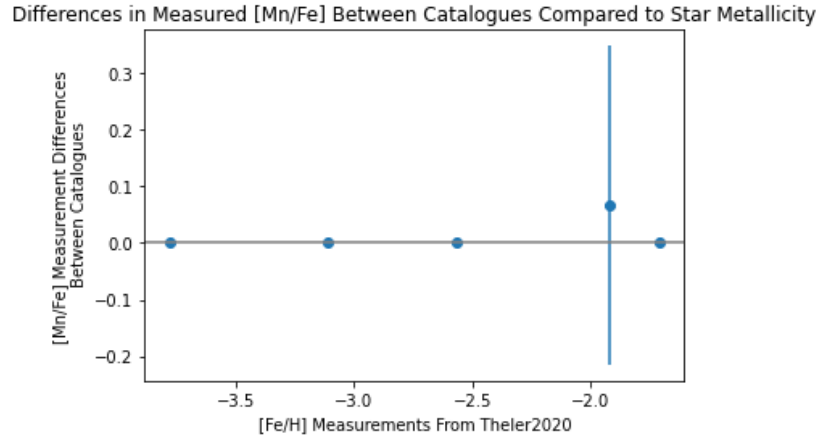


Figure 6:  $[\text{Mn}/\text{Fe}]$  Differences / Measured Metallicity

**$[\text{Ni}/\text{Fe}]$  differences:** Differences very closely centered around zero.

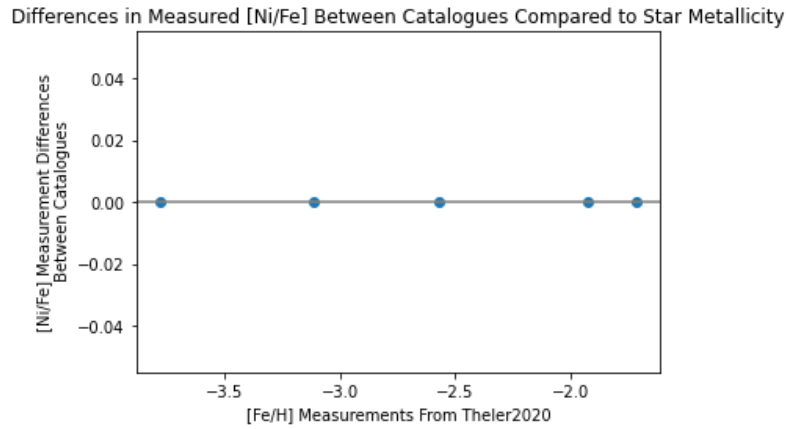


Figure 7:  $[\text{Ni}/\text{Fe}]$  Differences / Measured Metallicity

## 2.2 Create a linear model answering the following questions:

Now consider a linear model to describe the potential systematic offset between these two catalogs of a similar form as presented in the emcee “Fitting a model to data” tutorial. Write a function that evaluates the log-probability of this model. The model should have the following properties:

- The model should have three parameters describing the intercept and slope of the line as well as a scatter about the line.

- The model should predict the offset in the quantities as a function of  $[\text{Fe}/\text{H}]$ .
- Unlike in the emcee example, the additional scatter should be constant and not a function of  $[\text{Fe}/\text{H}]$ .
- The model can be written for a single measured quantity and need not be written for all quantities simultaneously.

You may find it helpful to write separate functions for the log-prior, and the log-likelihoods of each component. Be sure to clearly state what your priors are. Use MCMC to infer the best fit parameters for your model. You will need to initialize your MCMC walkers appropriately and provide evidence that your walkers have converged. Plot the posterior distributions of your sampling in a corner plot (a.k.a. pair plot). Perform and plot a posterior predictive check by over-plotting lines corresponding to a subset of MCMC samples. Do they match the data reasonably well?

---

Given that there are multiple abundances that could be modelled, the code written and following analysis will only be performed on one but will work for any. While beginning the analysis, one of 6 data values in the cross-matched table was masked and could not be used in the code, so the value was removed and only 5 values were used. If errors following a Gaussian distribution is assumed, then the least-squares estimates can be calculated:

$$m = -0.153 \pm 0.155$$

$$b = 0.179 \pm 0.305$$

Estimating a line this way entails minimizing the least squared errors of the data. This model also assumes the errors are perfectly accurate. If the intrinsic scatter of the data is factored in, the error term will involve both a Gaussian distribution and some intrinsic scatter which will be some arbitrary constant (0.5). The result is:

$$m = -0.800$$

$$b = -1.000$$

$$f = 1.649$$

(From the maximum likelihood estimation)



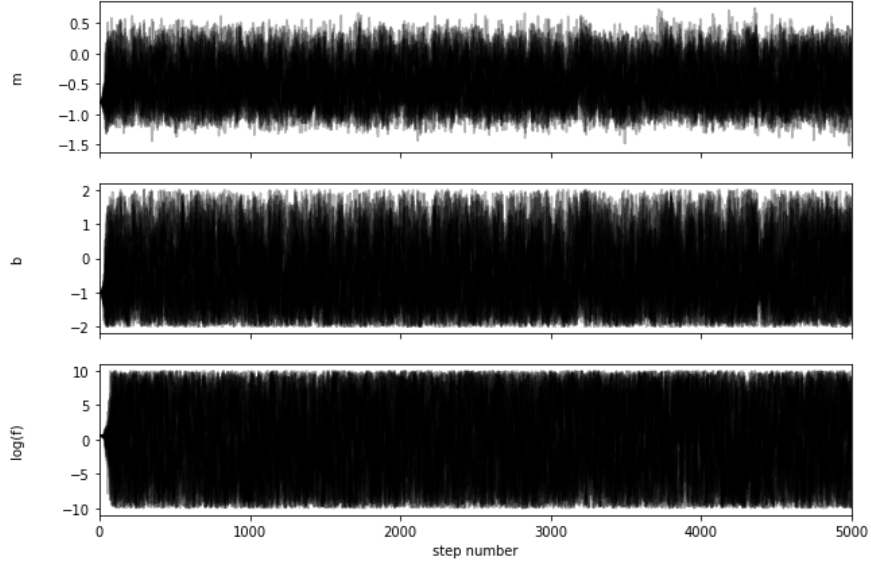


Figure 8: Resultant MCMC plot for each linear equation parameter

Furthermore, another set of candidate terms for the equation can be determined by MCMC. To begin, priors were chosen. If the provided initial slope, intercept and scatter were between -2 and 2, -2, and 2, and -10 and 10 respectively, the initial log prior is 0, otherwise it is negative infinity. Using emcee, 5000 steps of MCMC were done. To determine if the walkers have converged, `sampler.get_autocorr_time()` is used. The result is `[43.50281731 42.23950215 51.26996117]` which indicates that after around 40 steps, they have converged onto the final obtained values. These final obtained values being:

$$\begin{aligned} m &= -0.51 \\ b &= -0.502 \\ f &= 0.034 \end{aligned}$$

In this corner plot, the blue lines indicate what was predicted by the maximum likelihood estimation, since the true values are unknown. From the plot, the MLE are located within one

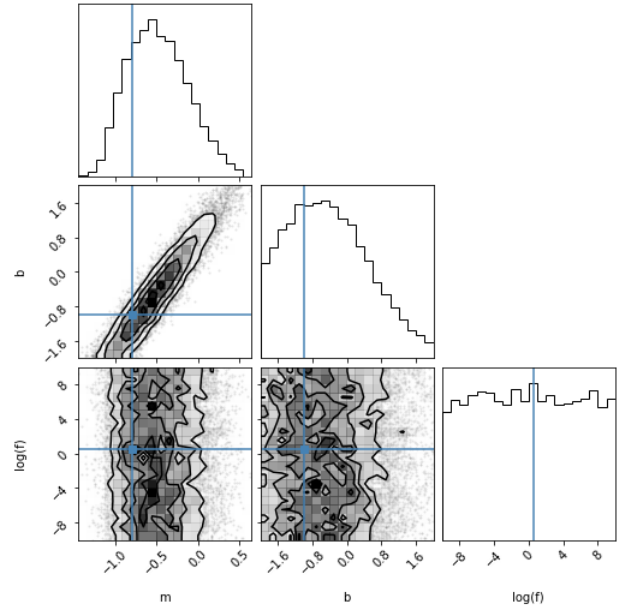


Figure 9: Corner plot for MCMC analysis with MLE comparison

standard deviation of what was predicted from MCMC, suggesting that it is also a reasonable estimate for the data. Furthermore, a sample of 100 steps from the MCMC analysis can be plot over the initial data, MLE line and LSE line.

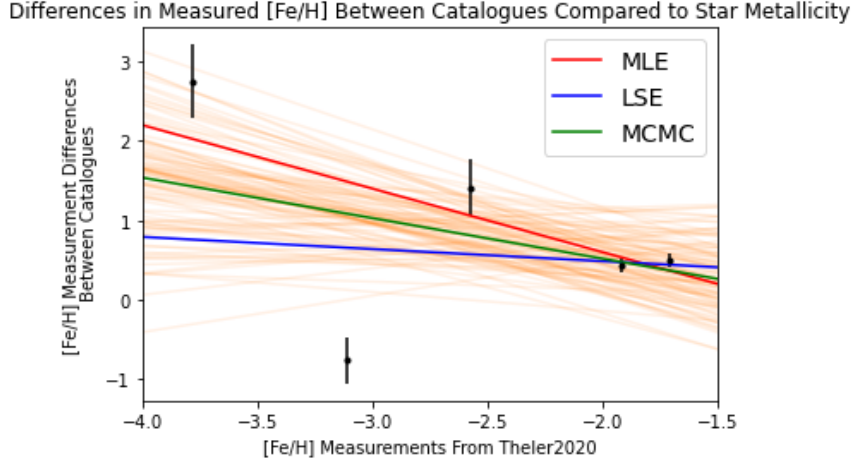


Figure 10: MLE, LSE, MCMC estimations for the  $[\text{Fe}/\text{H}]$ /metallicity differences

Evidently, both the MLE and LSE line are similar to some of the samples of MCMC, they are on the upper and lower ends of the samples respectively. The estimate generated by MCMC itself appears to be between the MLE and LSE line. Another feature to note is the proximity of all the lines to the data with smaller error bars, which is a nice visual note of the factoring in of the size of the error bars. Overall, it appears as though the best equation to represent the difference in  $[\text{Fe}/\text{H}]$  as the metallicity of stars changes is  $y = -0.51x - 0.502$  with an error of about 0.034

### 2.3 Summarize your preliminary findings. For which elements are the DESI measurements in good agreement with the literature? For which elements is the agreement poor? State what metrics have led you to this conclusion.

Overall, it appears as though most abundances are relatively consistent between the two catalogues with the exception of  $[\text{Fe}/\text{H}]$  and  $[\text{Ca}/\text{Fe}]$ . Both these quantities seem to differ more between the catalogues as metallicity decreases. The differences go beyond what can be explained by the error bars, as they can be for other abundances. However, this may be due to the small sample size of stars that remained after cross matching the catalogues. Given the fraction of stars that remained is so small, it is entirely possible that the patterns visible are coincidences caused by 'random' sampling.