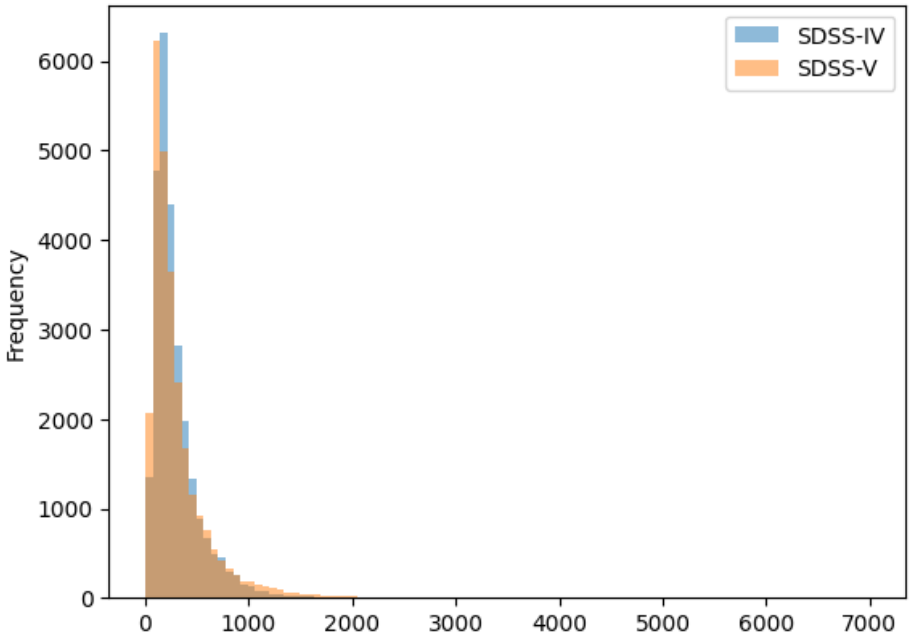
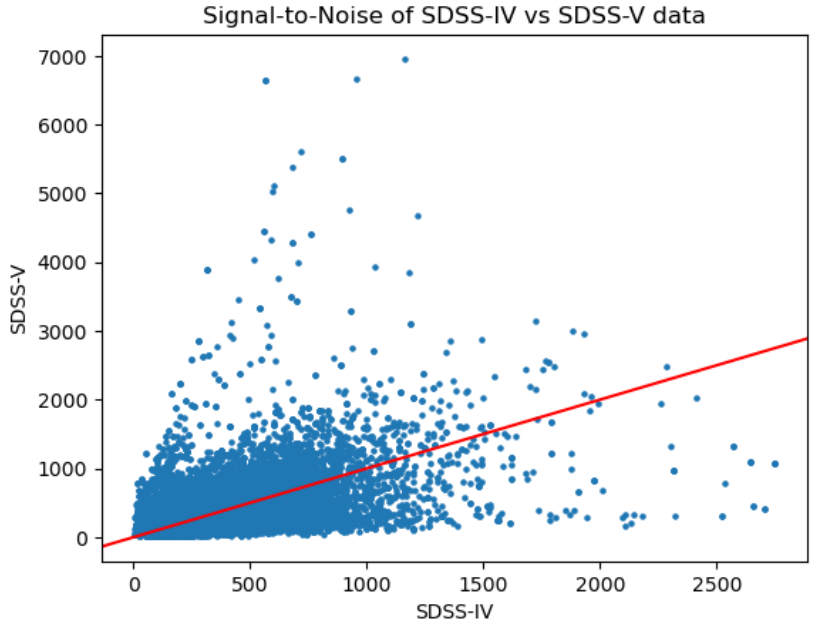
Started with X-match files which consist of about 27k star matches between SDSS-IV DR-17 and SDSS-V IPL-2 with about 21k unique stars (many many-to-one matches) between the two surveys.

The distribution of observation signal-to-noise ratio (SNR) is quite similar

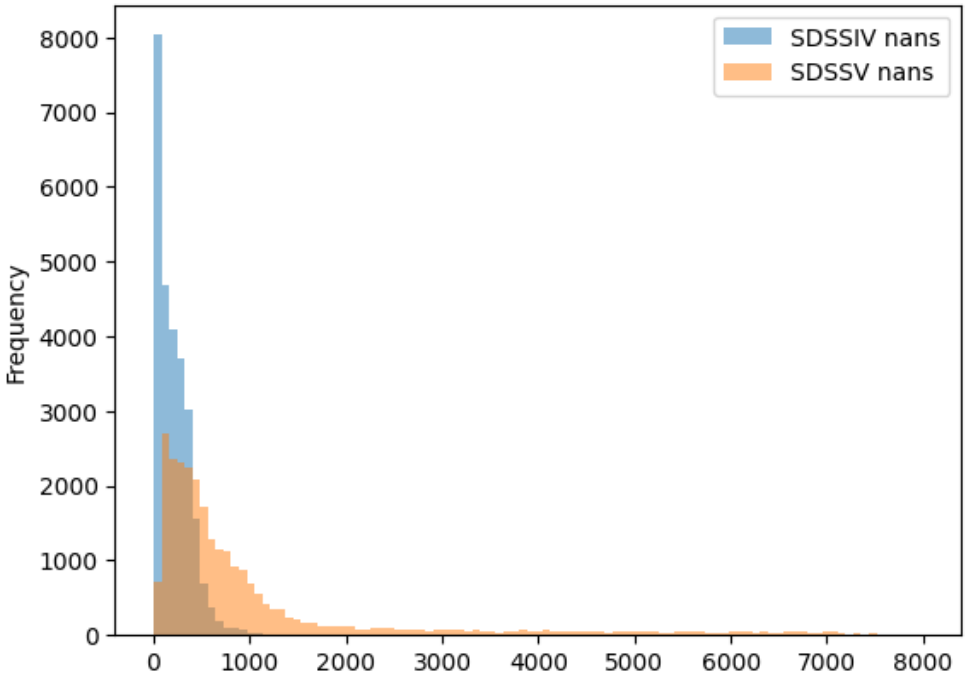


However, the correlation of the SNR between the two surveys in this sample is only about 0.5

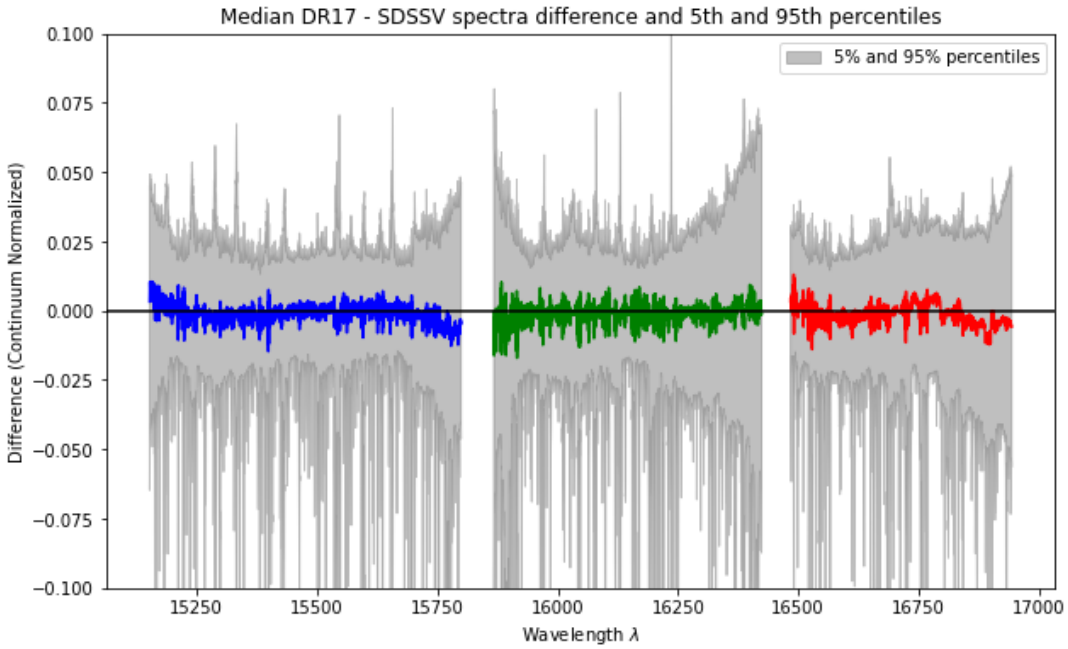


Another notable difference between the observations is in the number of bad pixels. As can be seen in the exhibits below the SDSS-V data has far more bad pixels than the SDSS-IV data (nan = bad pixel in this coding).

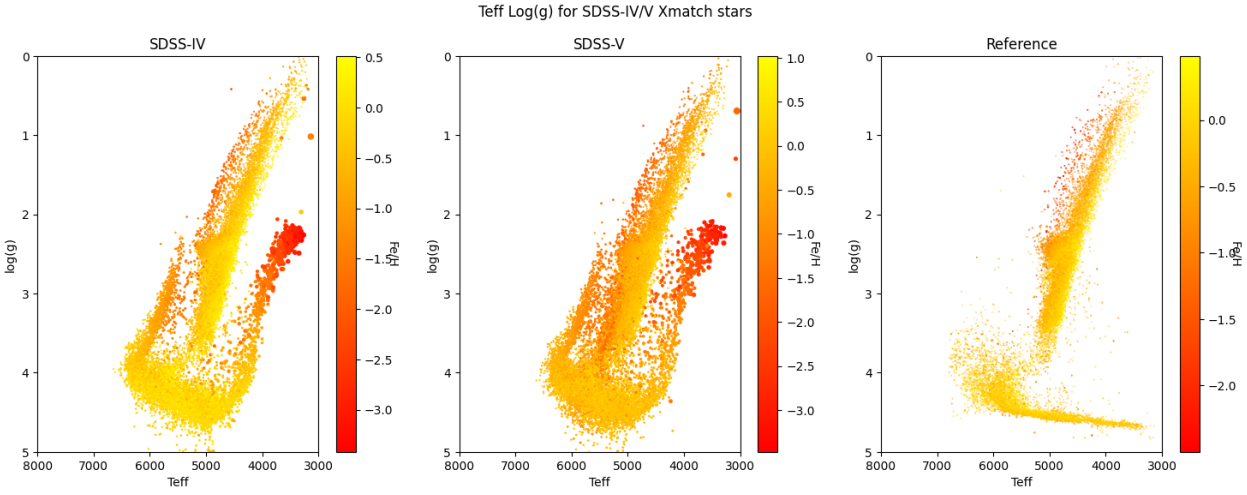
	SDSSIV nans	SDSSV nans
count	26714.000000	26714.000000
mean	204.598899	1052.532417
std	180.206379	1430.652731
min	0.000000	8.000000
25%	57.000000	273.000000
50%	174.000000	526.000000
75%	312.000000	1032.000000
max	2293.000000	7514.000000

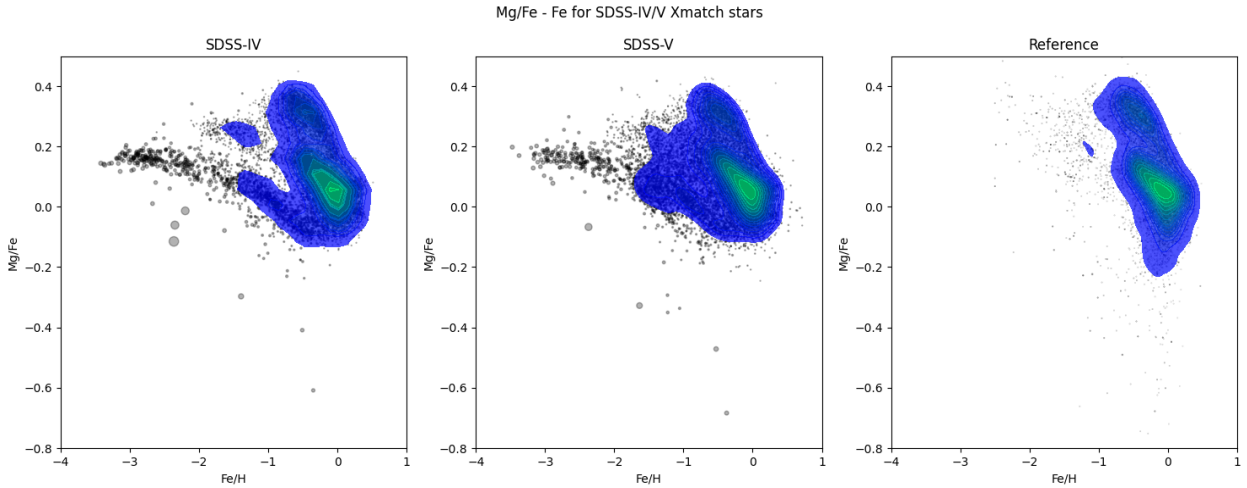


If bad pixels are coded to 1 as is the case when the spectra is used for inference, there can still be large differences in the spectra. Even though the mean difference across all channels is centered at 0, the 5th and 95th percentiles in the difference can be quite significant (up to 0.05 dex for 95th percentile and 0.1 dex for 5th percentile)

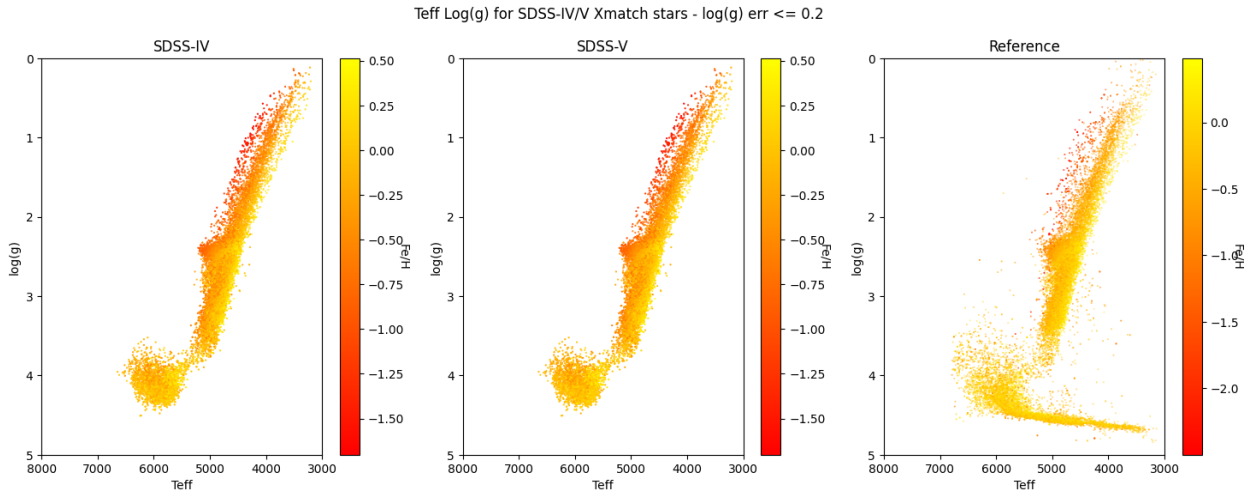


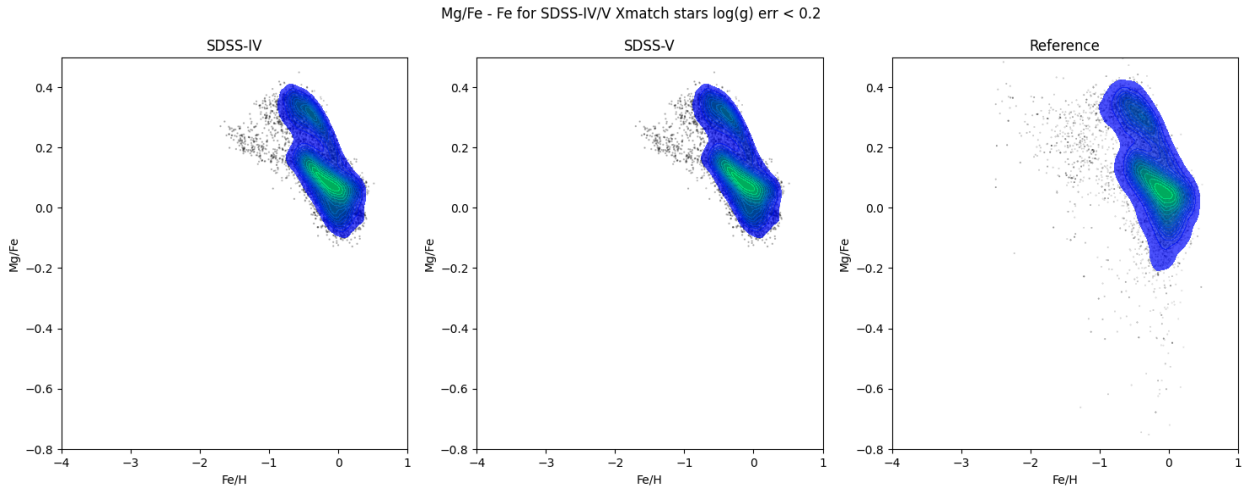
However, despite these differences the inference results (using `astroNN_0512_run002` which is trained on DR=17 data for inference) are quite similar when measured by teff - logg - Fe/H and Mg/Fe - Fe



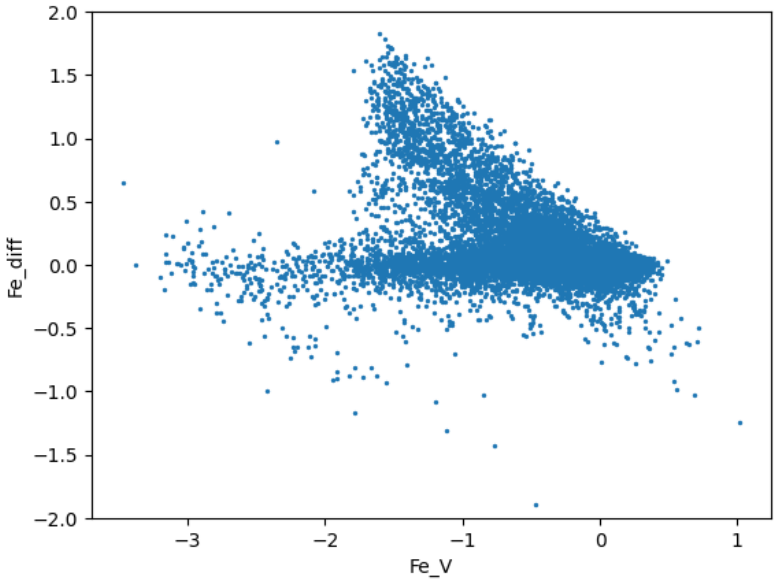


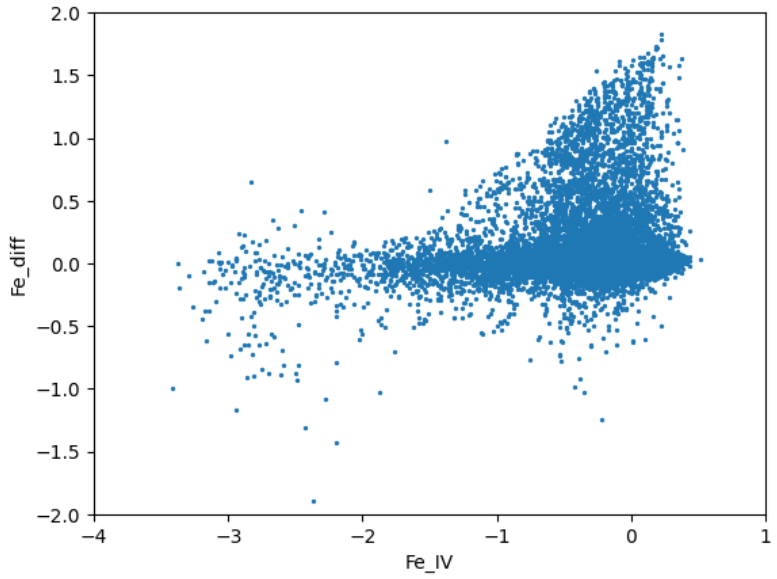
Though there are notable differences between astroNN predictions and the ASPCAP reference data, problematic predictions are largely removed when filtering on model certainty (logg error < 0.2). In particular, the trained model performs poorly on the main sequence and thus valid observations are mostly confined to the red giant branch.



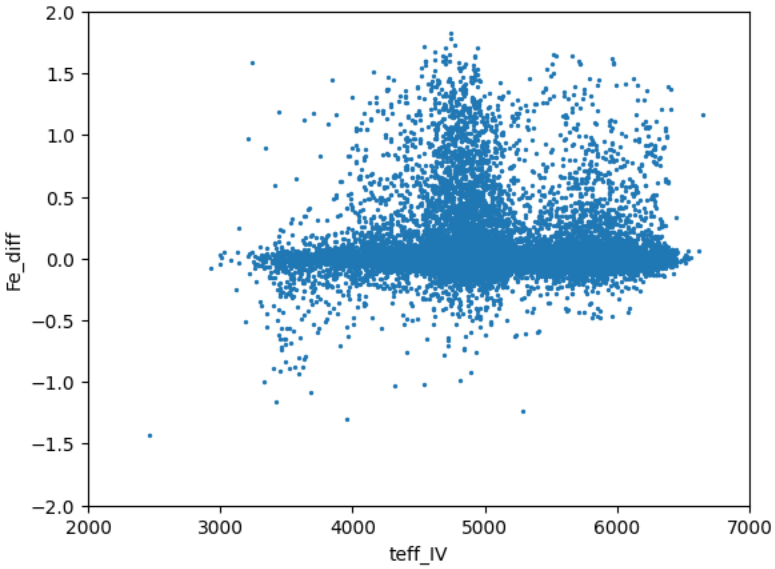


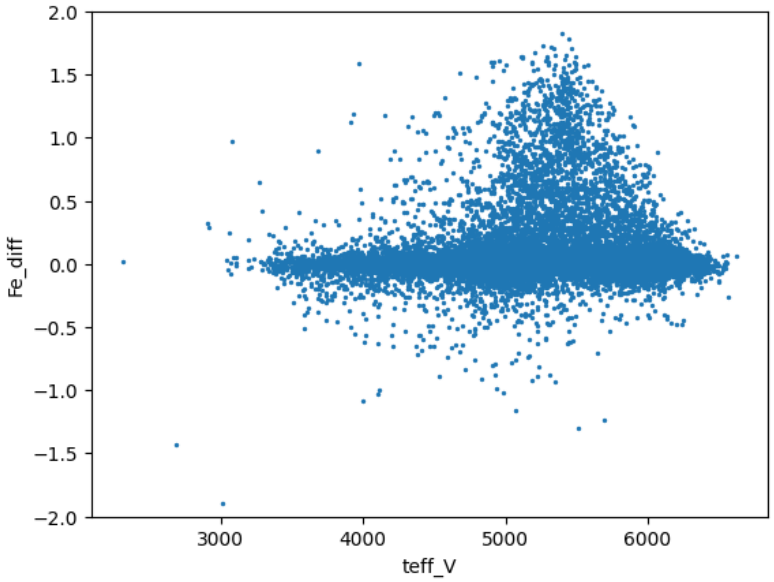
However, prediction differences between SDSS-IV and SDSS-V do exhibit some heteroskedasticity, particularly in Fe/H difference - Fe/H



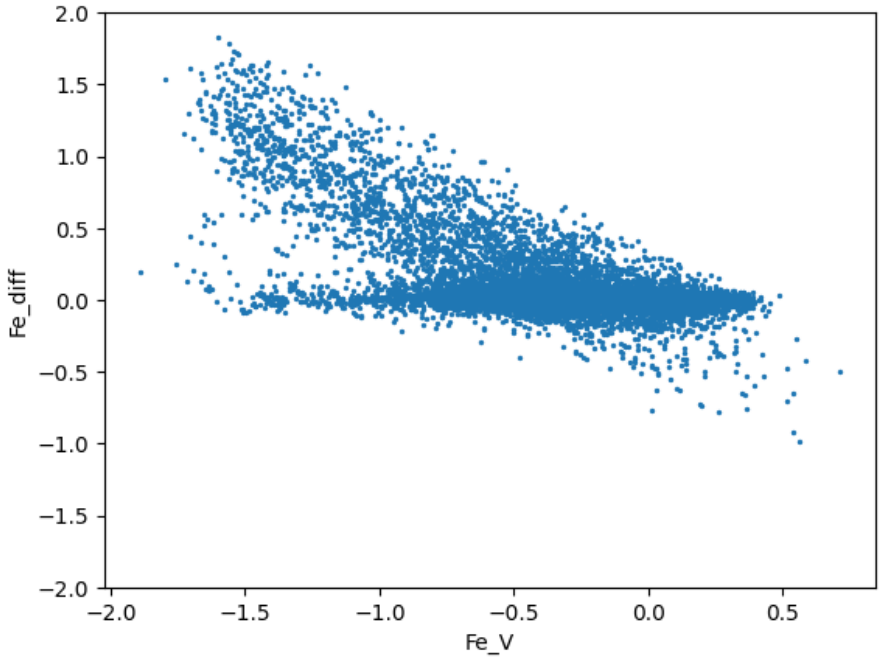


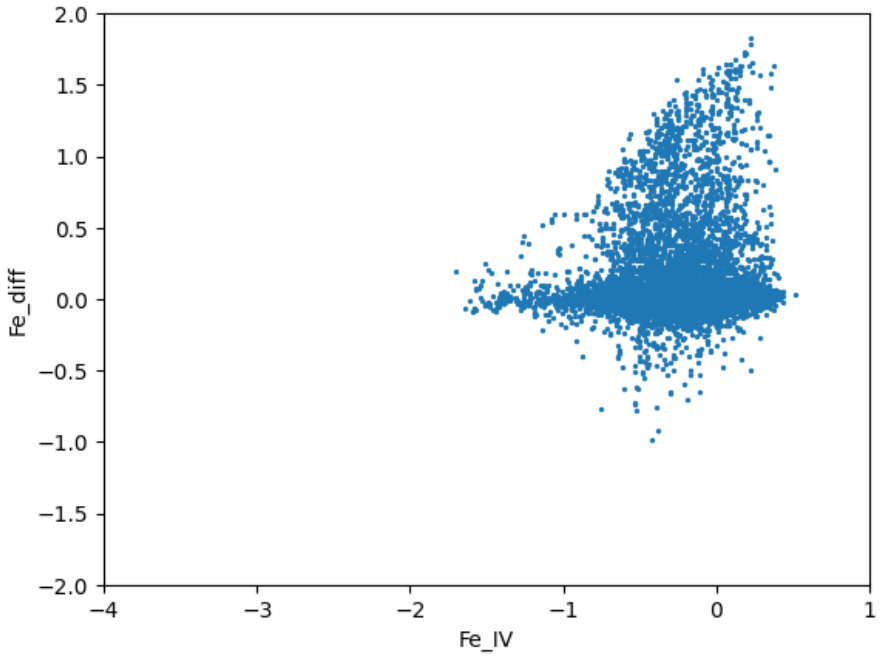
And Fe/H difference - Teff



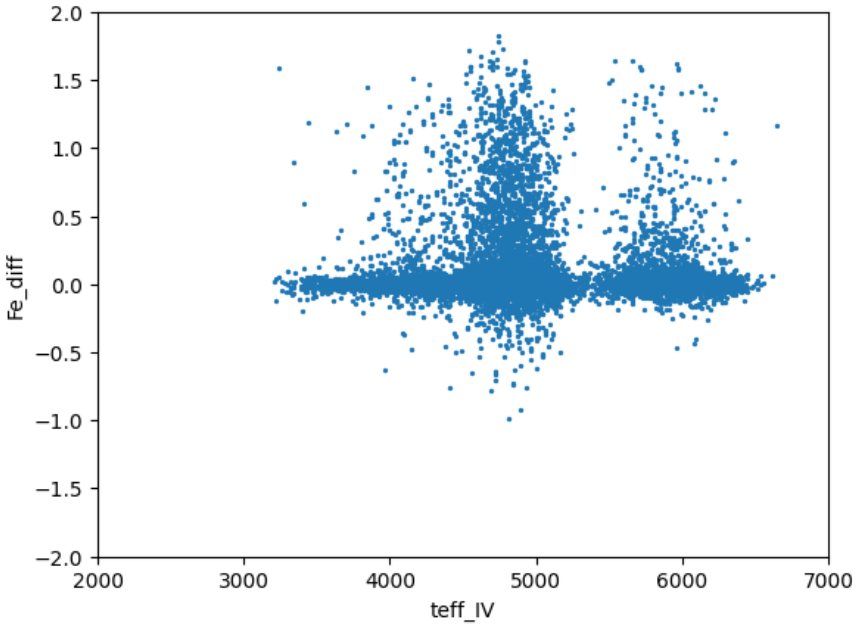


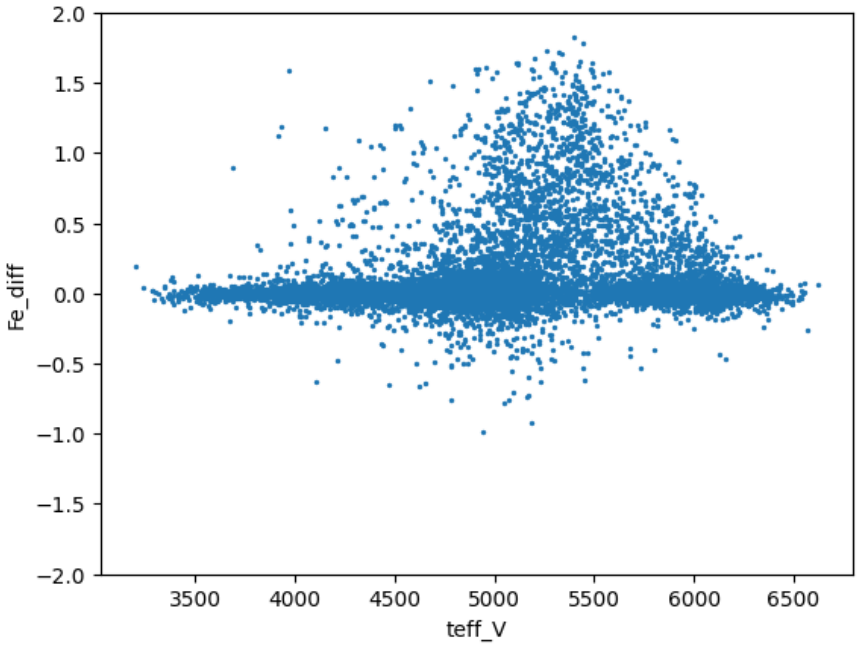
Even when filtered on a union of logg errors on predictions for both SDSS-IV and SDSS-V stars the heteroskedasticity still persists for Fe diff - Fe



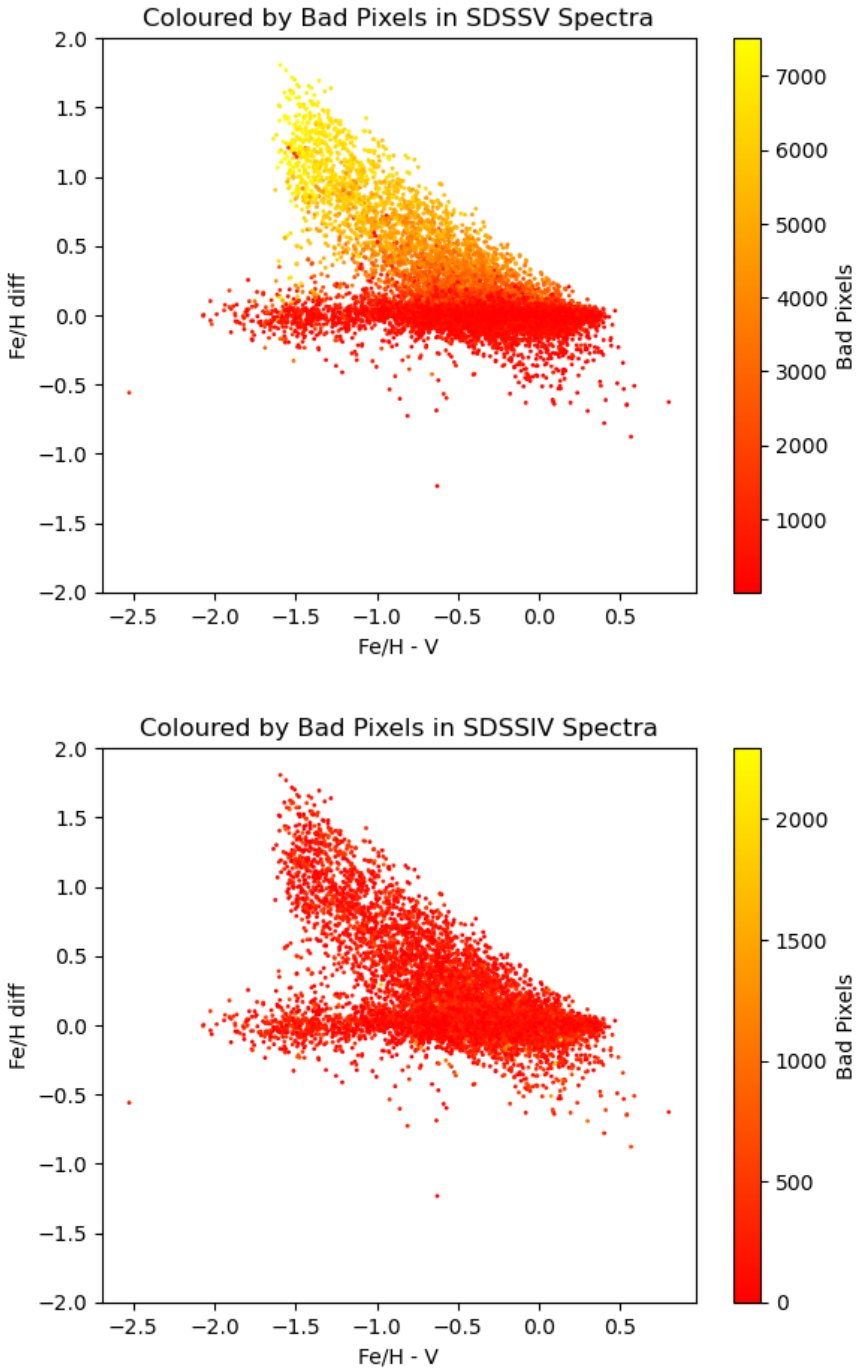


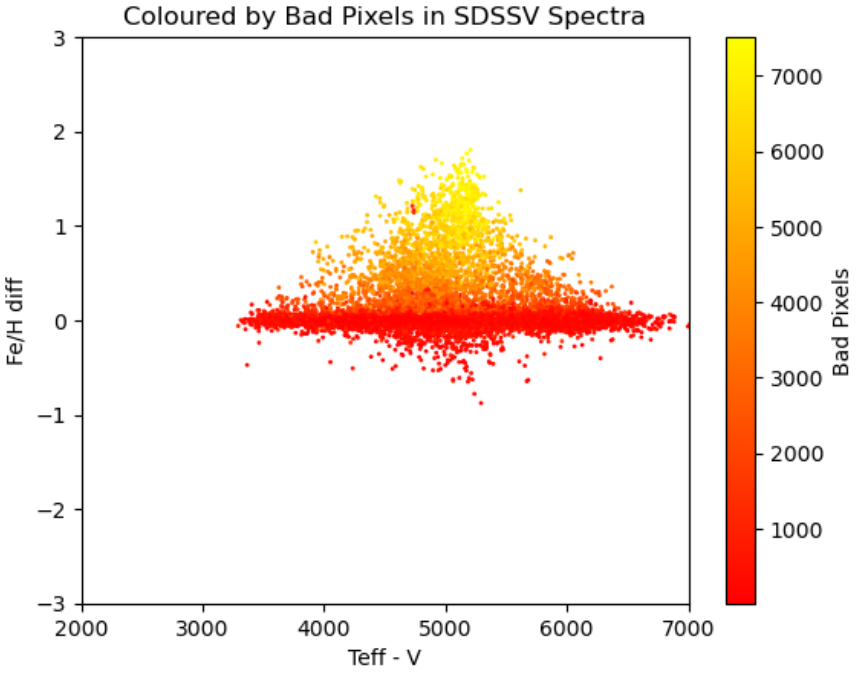
And Fe diff - teff

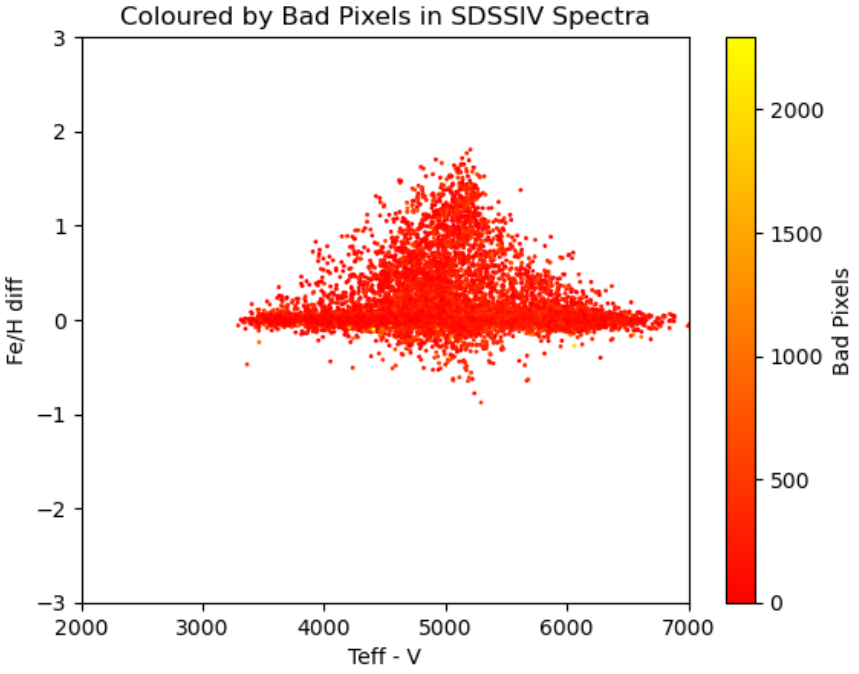


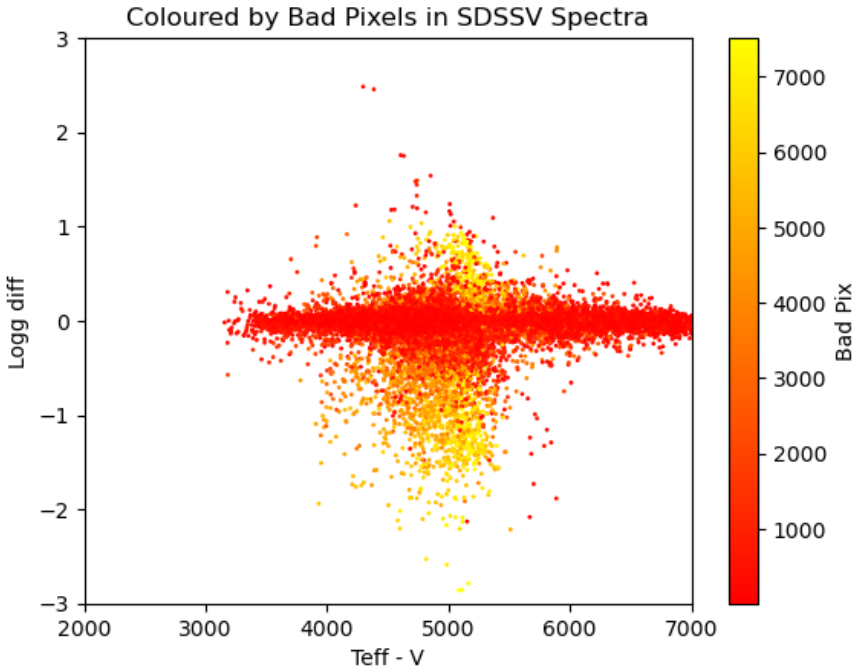


It does seem like a lot of the large errors come from missing pixels in the SDSS-V data and not SNR

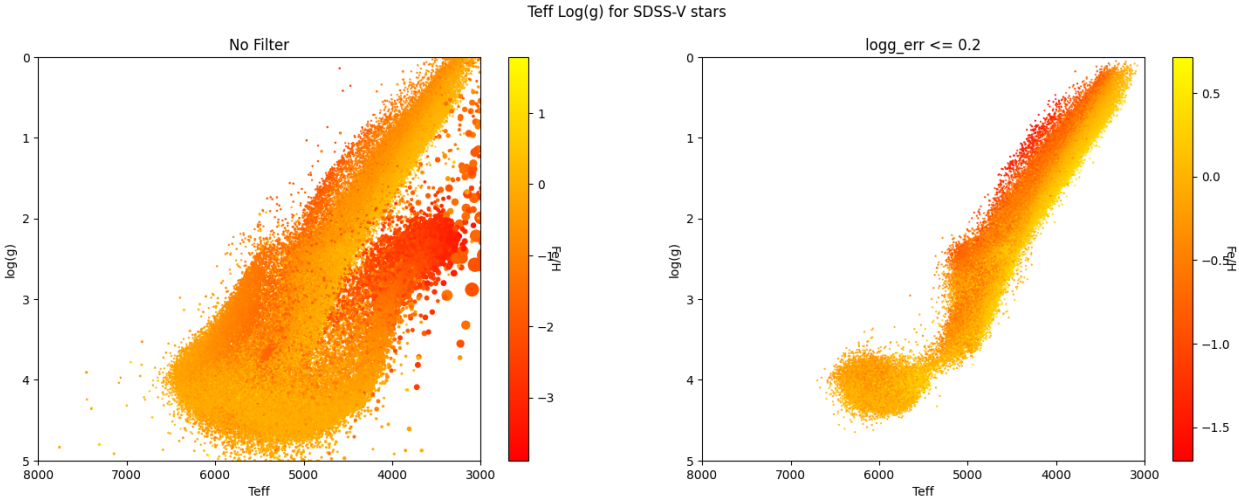


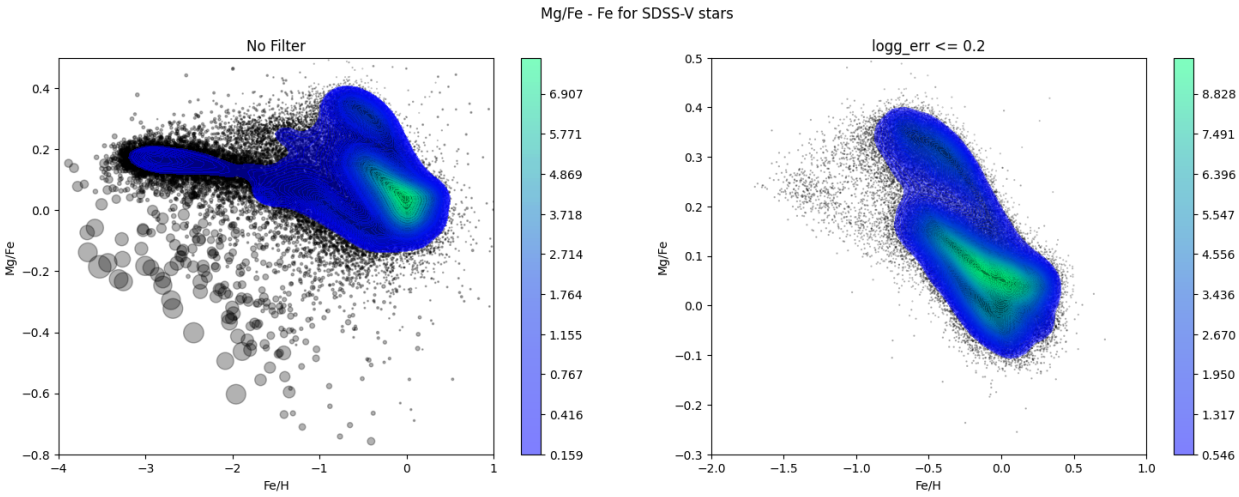

Coloured by Bad Pixels in SDSSV Spectra
Fe/H diff
Teff - V
Bad Pixels
3
2
1
0
−1
−2
−3
2000
3000
4000
5000
6000
7000
7000
6000
5000
4000
3000
2000
1000


Coloured by Bad Pixels in SDSSIV Spectra
Fe/H diff
Teff - V
Bad Pixels
3
2
1
0
−1
−2
−3
2000
3000
4000
5000
6000
7000
2000
1500
1000
500
0



When applied to the entirety of SDSS-V data the results are as expected given the results from the crossmatch data.



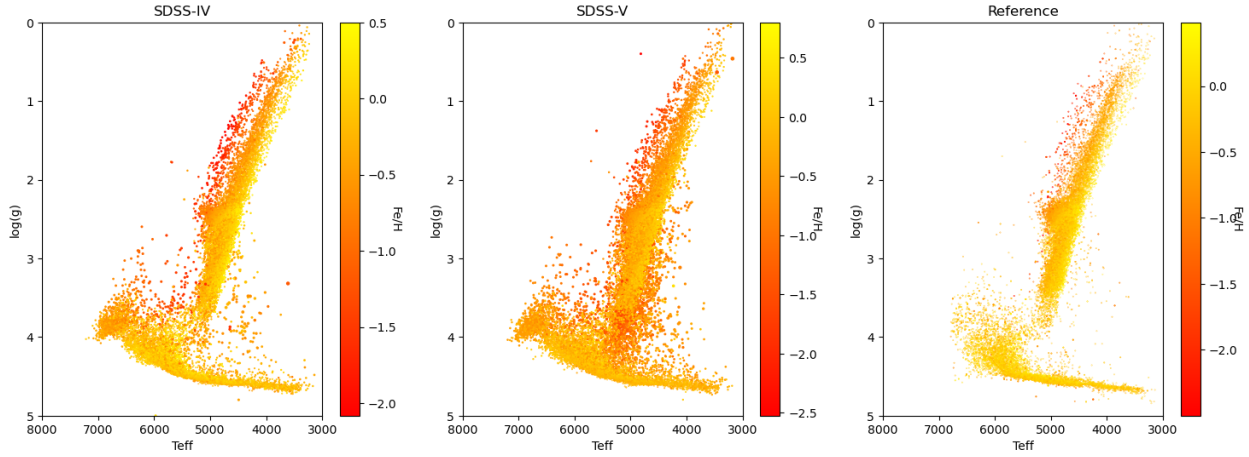


In the interest of improving performance across a wider portion of stars, we are considering training a model on a less conservative cut of DR-17 stars which includes a greater number of main-sequence stars. However, this may introduce issues with the data quality of the spectra and ground truth of the stellar parameter and abundance estimations (relaxing filter on ASPCAPFLAG).

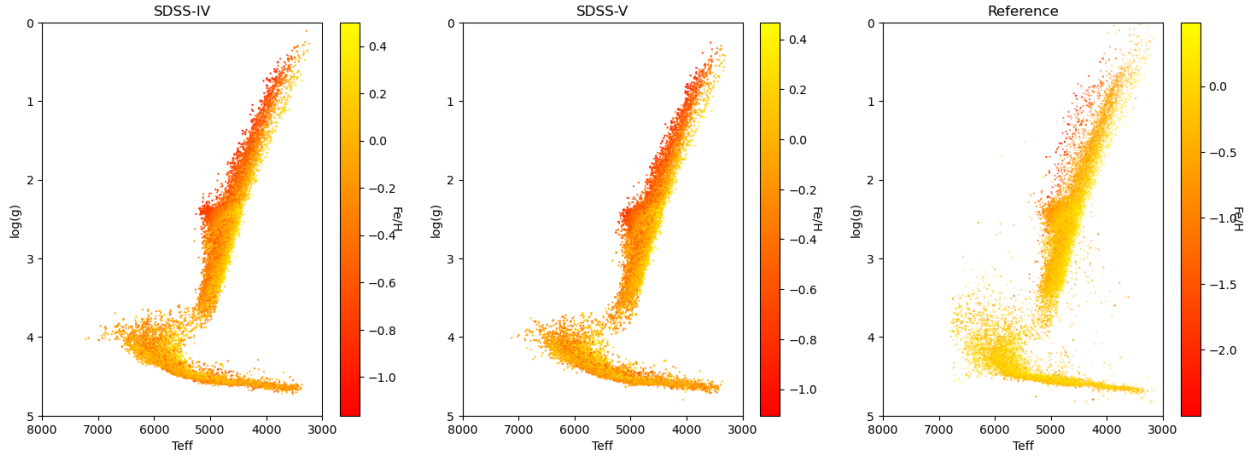
The retraining model used a filter on all DR17 stars of M_H_BAD, STAR_BAD, NO_ASPCAP_RESULT, MISSING_APSTAR, NO_GRID, and PROBLEM_TARGET

As we can see in the plots below, retraining the model on a less conservative cut of data greatly improves predictive performance on the main sequence

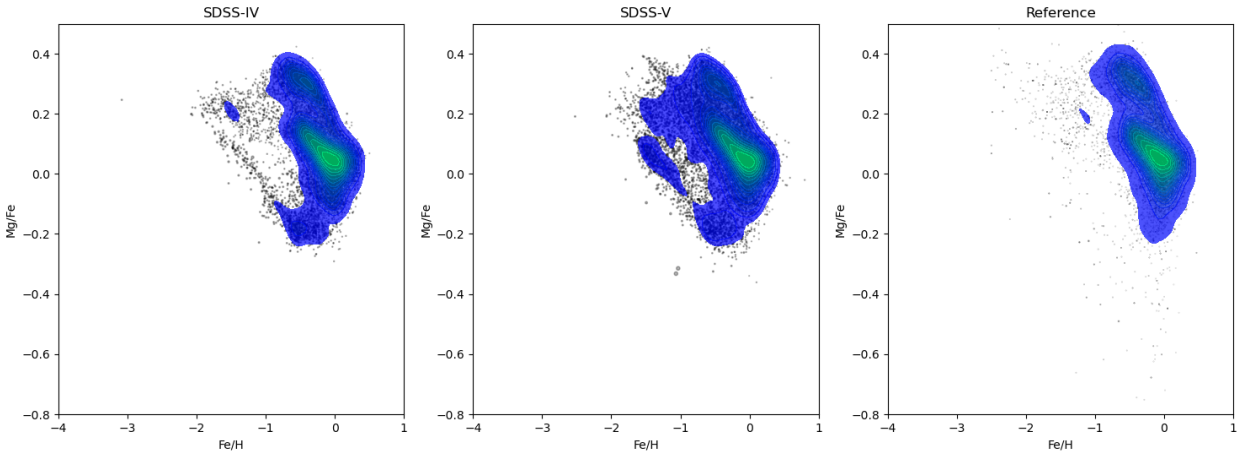
Teff Log(g) for SDSS-IV/V Xmatch stars

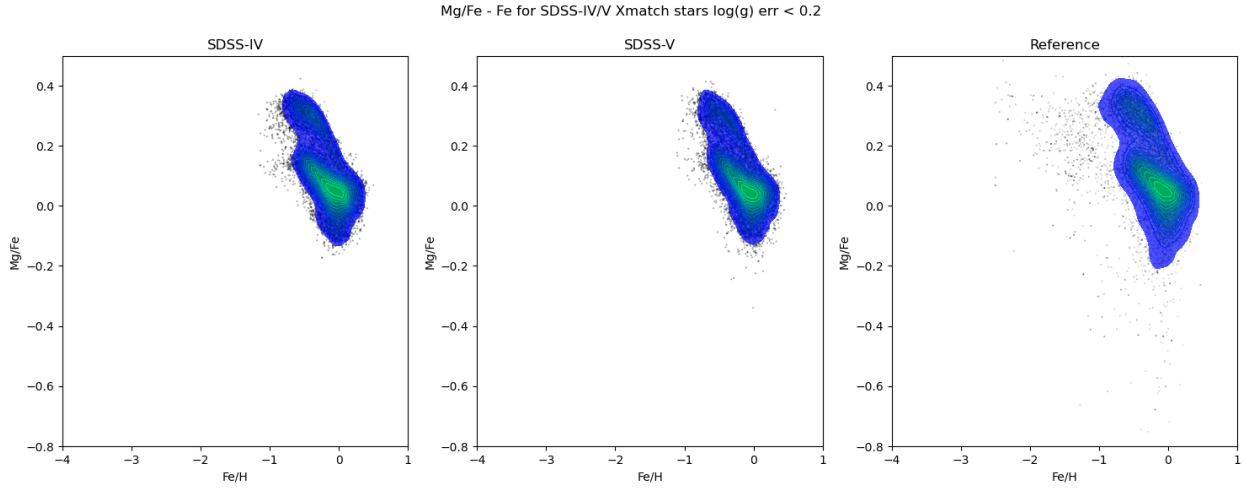


Teff Log(g) for SDSS-IV/V Xmatch stars - log(g) err <= 0.2

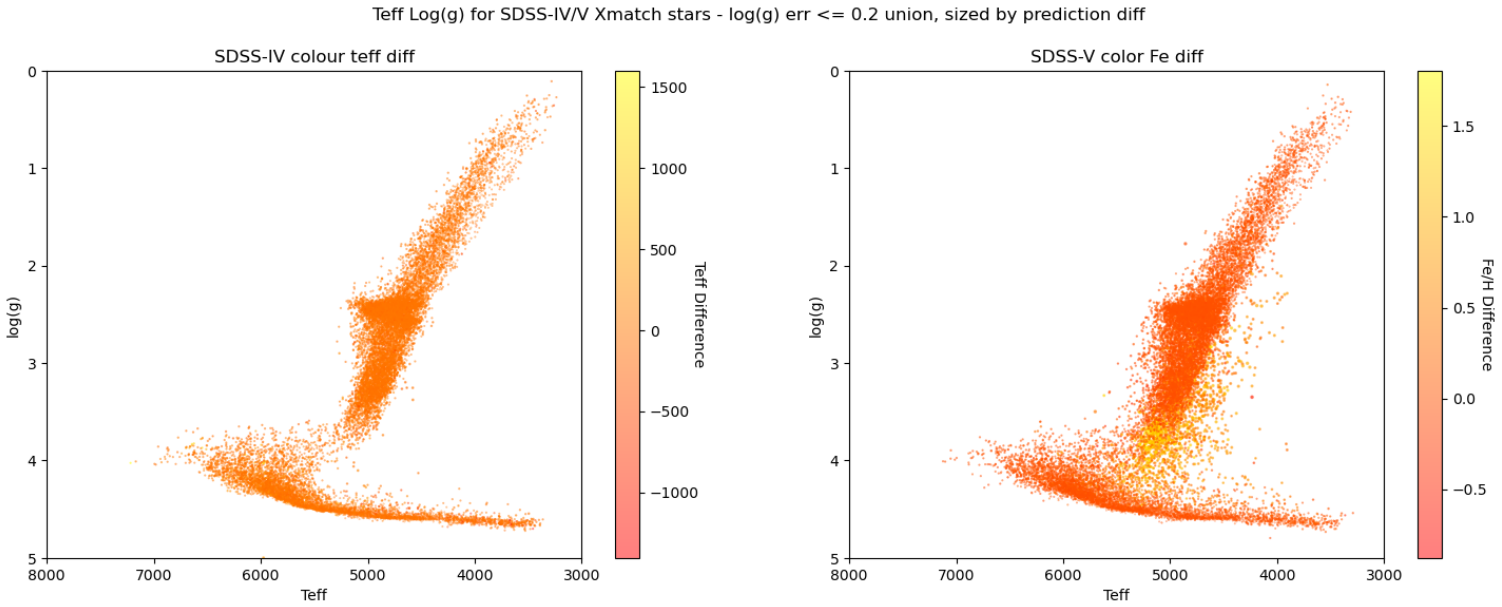


Mg/Fe - Fe for SDSS-IV/V Xmatch stars

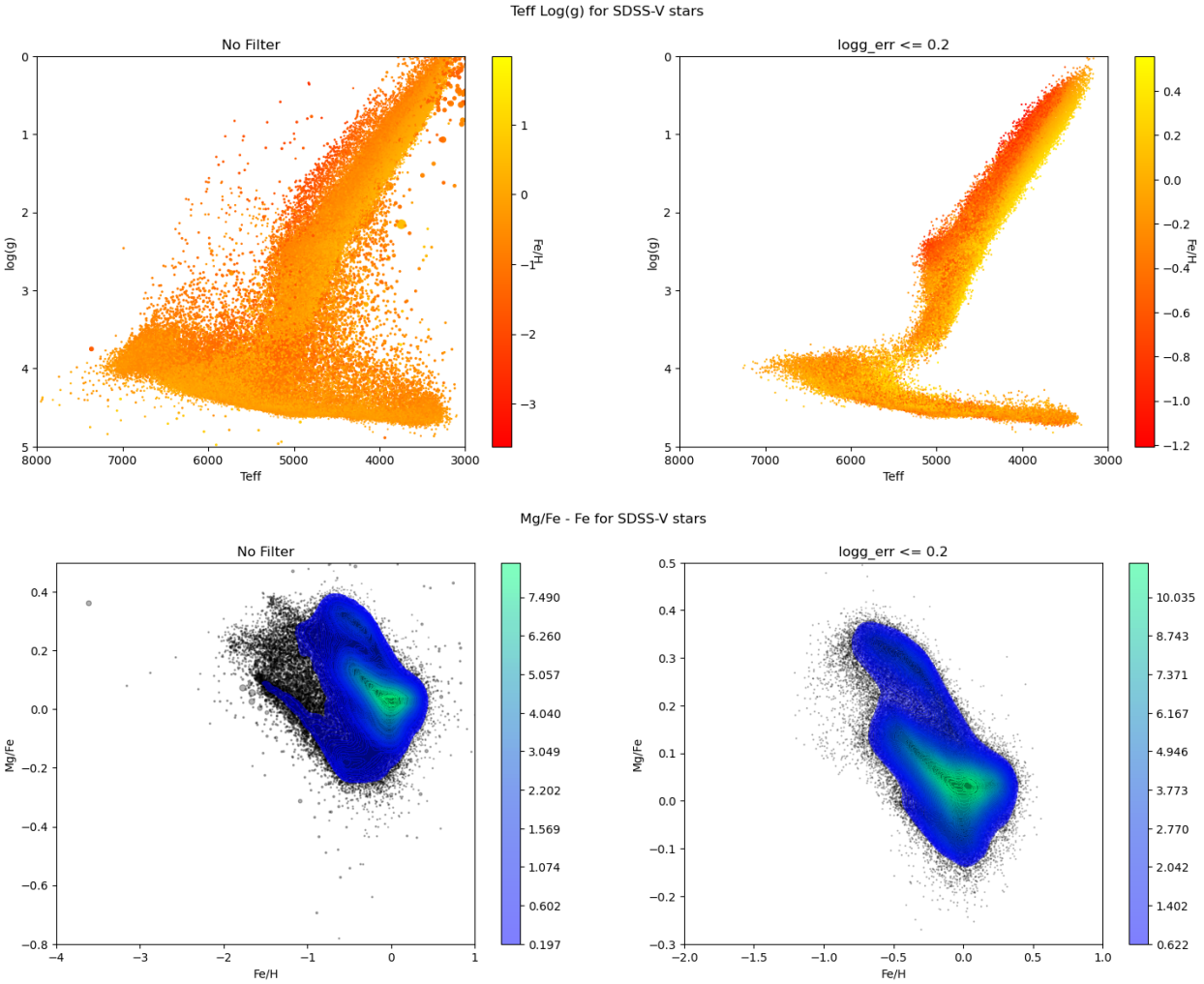




Teff - Logg Plot Coloured by Prediction Difference (teff and Fe) (Logg err <= 0.2) (retrained model predictions)



The retrained model was also applied to the entirety of the SDSS-V data



Just the main sequence (logg > 3.7)