Detecting CIV absorption lines in SDSS by Gaussian Processes

Reza Monadi[1], Simeon Bird[1], Ming-Feng Ho[1], Kathy Cooksey[2]

[1]University of California Riverside, Riverside, USA.
[2]University of Hawaii at Hillo, Hillo, USA.

Introduction

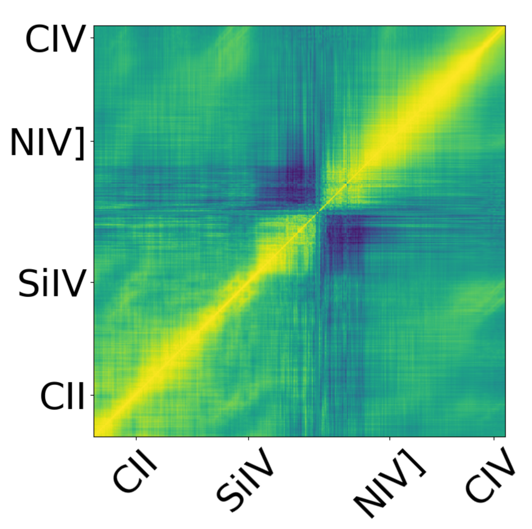
- Metal absorption lines in quasar spectra give us information about the history of Universe.
- The spatial distribution of CIV (triply ionized carbon) absorbers in the Inter Galactic Medium (IGM) tells us about the effect of feed-back and star formation.
- IGM clouds with CIV may be the main ingredient of future galaxies.
- CIV is a doublet outside of Lyman-alpha forest which is good for automatic detection.
- Here we introduce the method and preliminary results of our CIV catalog, which will be the *largest* one so far.

Overview of our method

- Our team at UCR has developed an automated technique for detecting damped Lyman-alpha absorbers[2, 3, 4] which can be used for detecting other absorption lines including CIV.
- We use a novel tailored Gaussian process to model quasar spectra that are clear from CIV absorption line.
- The Gaussian process is trained on a set of spectra that were assessed for CIV absorption but no CIV absorption was detected.
- Then, we model a CIV absorption by a Voigt profile convoluted with a learned absorption-free spectrum in the previous step.
- Finally, using Bayesian model selection, we obtain the probability of existing a CIV absorption in an unseen spectrum.
- With the CIV absorption probabilities, a user can get a catalog with any level of desired certainly.
- The main advantage is getting the most of information out of a spectrum, even from the noisy and low-resolution spectra like the SDSS.

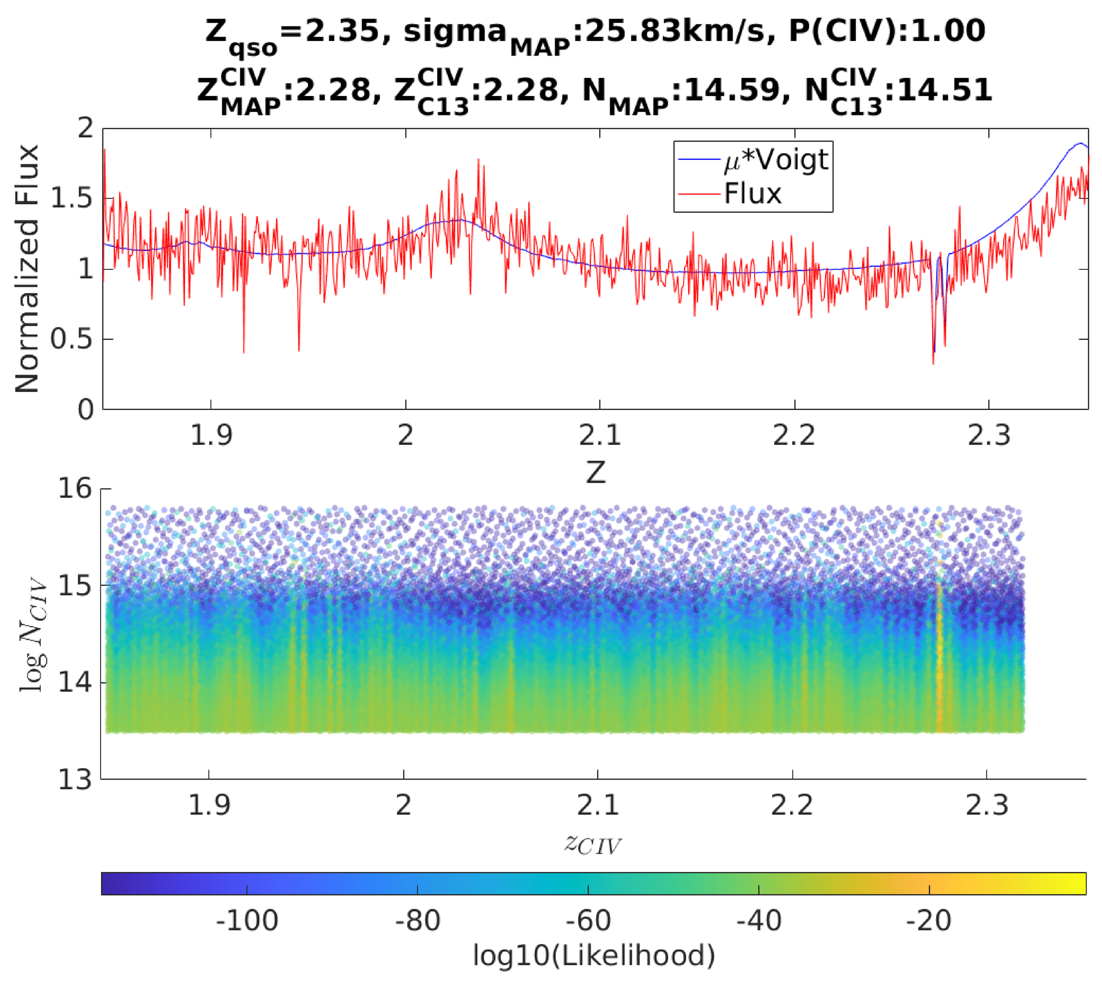
Learning the null model

- Our training set consists of $\sim$ 26000 SDSS DR7 spectra which were assessed for CIV absorbers in [1].
- The absorption free model is trained on $\sim$12000 spectra in [1] with confirmed no CIV absorption.
- We find the flux co-variance matrix that maximized the likelihood of our data.



Above figure shows the co-variance matrix labeled by different emission lines.

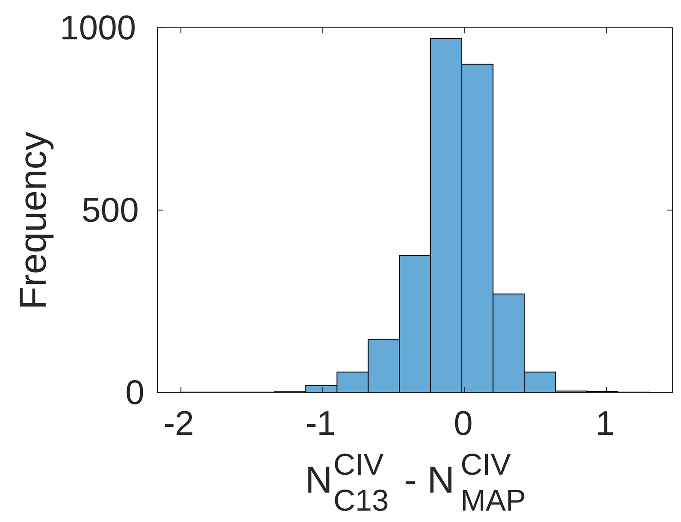
Testing the line detection



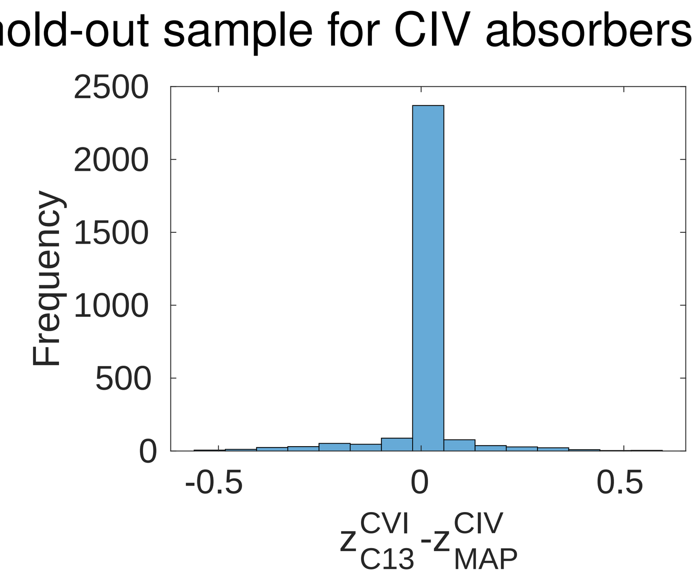
- We trained on 80% of spectra in [1] and hold out 20% of those spectra for validation.
- The upper panel of figure below shows the raw flux from SDSS and our fitted flux which is a convolution of a Voigt profile and our learned mean flux.
- The Lower panel of the figure shows the likelihood probability for each pair of our estimating parameters: Column density (N) and redshift (z)
- Our Bayesian approach gives a very high probability (P(CIV)=1) for this example spectrum with confirmed CIV in the hold-out sample.
- Our redshift and column density estimations are in agreement with the measurements in [1] for this example CIV absorber.

Validating the parameter estimation

- The figure below demonstrates the difference between CIV column (N) density and our Maximum a posteriori estimation for N in the detected CIV absorbers in our hold-out sample consisting 5200 spectra.

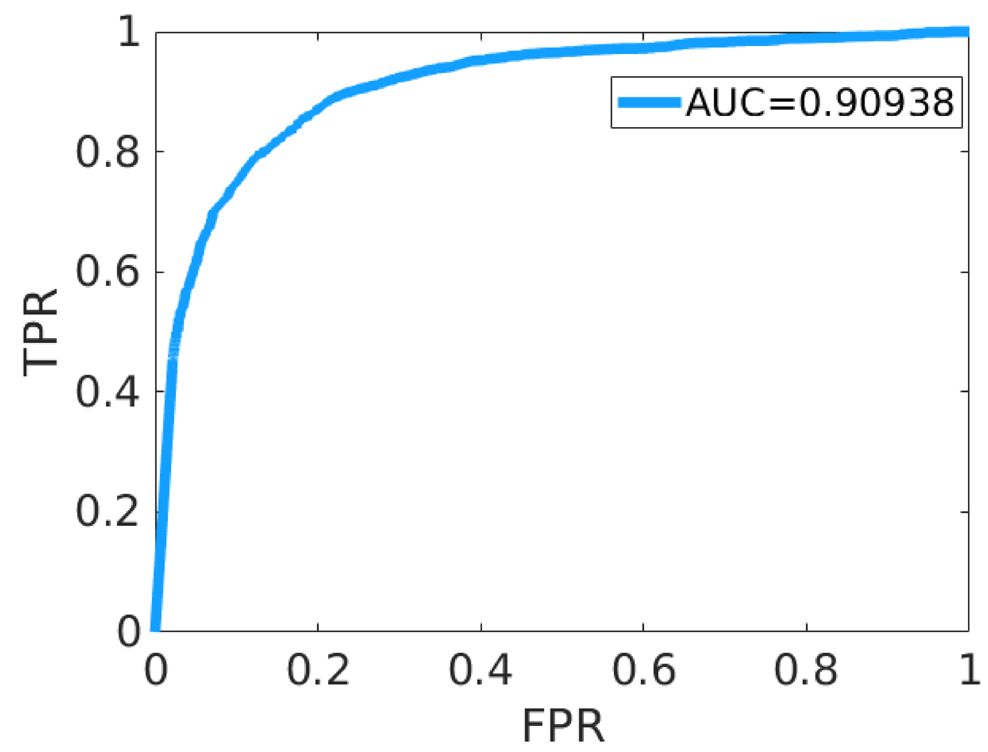


- Figure below shows the difference between our redshift estimation and the redshift measurements in the hold-out sample for CIV absorbers in [1].

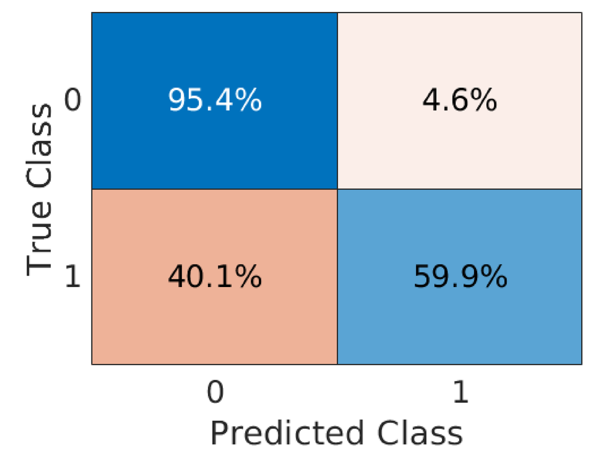


Performance

- We present ROC curve and confusion matrix to show the performance of our approach.



- Figure above shows the ROC curve for our validation set having the true labels from [1].



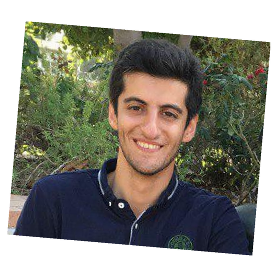
- The confusion matrix for a probability threshold of 85% for accepting a CIV absorber as a true detection is shown above. Class 0 is CIV absorber and class 1 is the absorber free class.
- We believe that we some of our false positives are real absorbers missed by [1].

Some important considerations

- **CIV column density**: The difference between our estimations and the measurements in [1] is spread around 0 with a little skew, probably because many N measurements are lower limits in [1].
- **CIV absorption redshift**: There is a very good agreements between our estimations and the redshift measurements in [1] as in seen by a narrow and symmetric distribution around 0 for the redshift differences between us and [1].
- **ROC curve**: The area under curve (AUC) for the ROC curve is more than 0.90 which shows a good performance.
- **Confusion matrix**:
  - Despite a very high true positive rate of 95%, we have 40% False Positive Rate for the probability threshold of 85%.
  - False positives can be real absorbers missed by [1] because of their sensitivity limitation.
  - Assuming the CIV absorbers are correlated with Lyα absorption lines, we may incorporate this information to constrain our posterior probabilities for CIV absorption.

References

[1] Kathy L Cooksey, Melodie M Kao, Robert A Simcoe, John M O'Meara, and J Xavier Prochaska. *ApJ*, 763(1):37, 2013.
[2] Roman Garnett, Shirley Ho, Simeon Bird, and Jeff Schneider. *MNRAS*, 472(2):1850–1865, 2017.
[3] Simeon Bird, Roman Garnett, and Shirley Ho. *MNRAS*, 466(2):2111–2122, 2017.
[4] Ming-Feng Ho, Simeon Bird, and Roman Garnett. *MNRAS*, 496(4):5436–5454, 2020.



Email: reza.moandi@email.ucr.edu
Website: https://rezamonadi.github.io