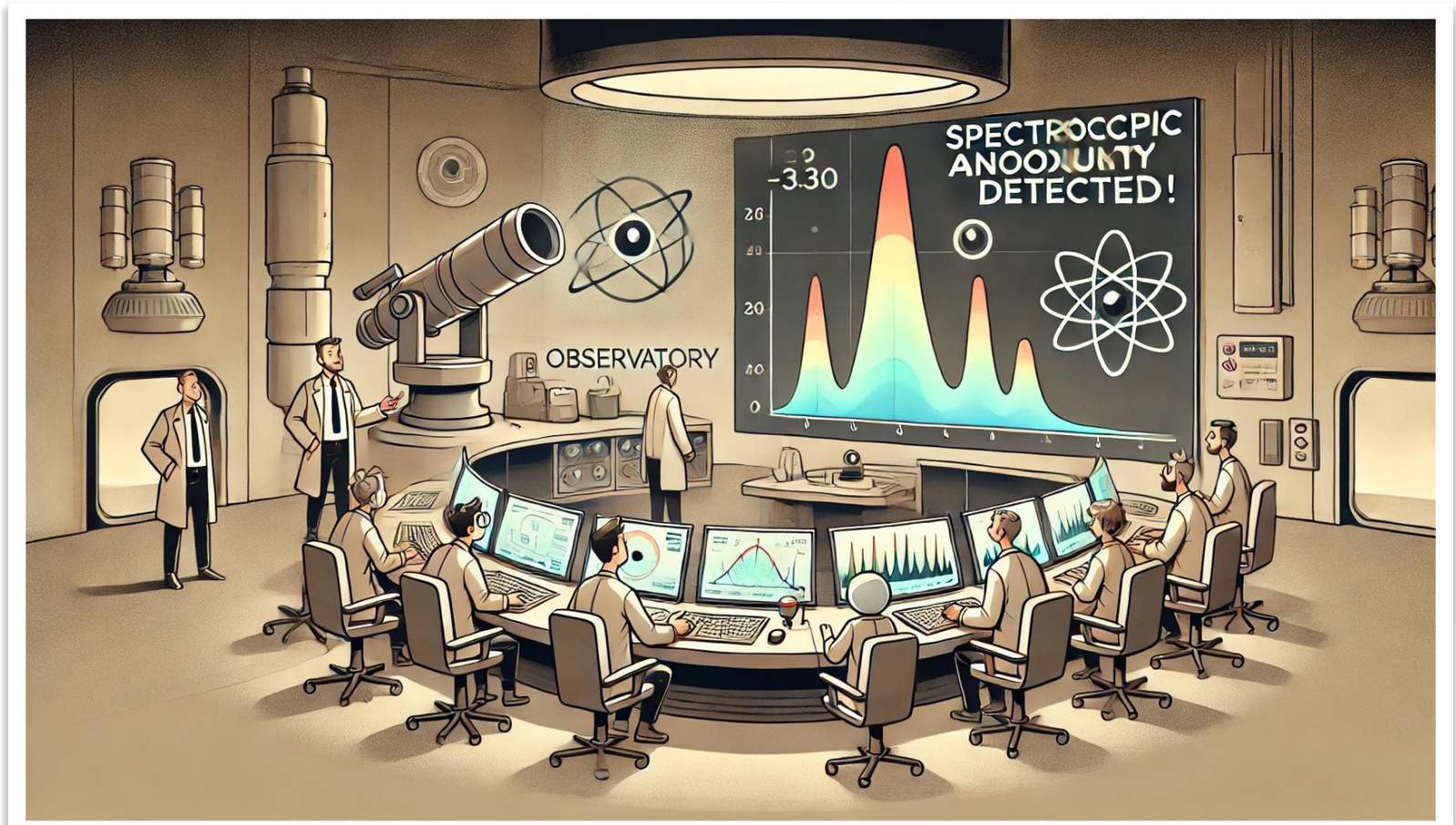
Spectroscopic Quasar Anomaly Detection

SDSS DR16 Quasar Catalog



DALL.E prompt

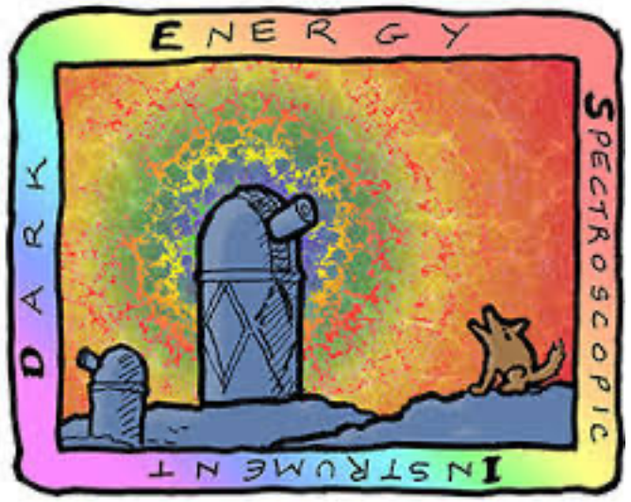
make a cartoonish image to be used for the title slide of the presentation about spectroscopic anomaly detection project; neutral tone

Vivek M (IIA) 08/01/2025

Data Revolution in Spectroscopy

SDSS DR16Q catalog which contains 750,414 quasar spectra

SDSS



4MOST

MEAVE

GAMA

PFS
Prime Focus Spectrograph
SuMIRe

Spectroscopic Quasar Anomaly Detection (SQuAD) Project

Arihant Tiwari
IIA VSP intern

Goal is to identify and characterise anomalous quasars from these large datasets

arXiv:2411.16858

Follow-up the interesting ones with bigger facilities

https://vivekastro.github.io/SQuAD.github.io

Spectroscopic Quasar Anomaly Detection (SQuAD) Project

Home | About the Project | The Team | Download Catalogs | Contact Us

Welcome to the SQuAD Project!

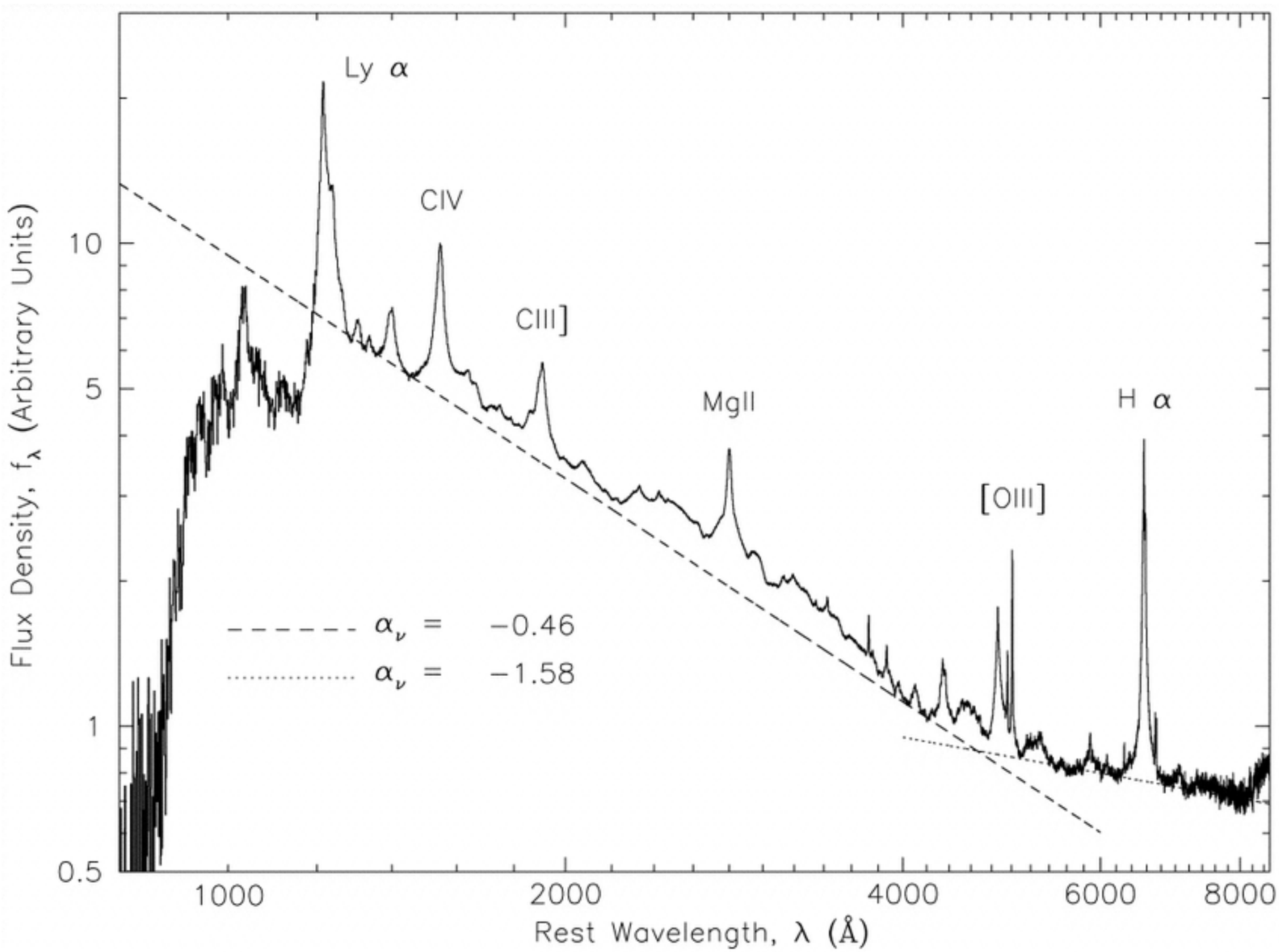
The **Spectroscopic Quasar Anomaly Detection (SQuAD) Project** is dedicated to identifying anomalous quasars from large-scale spectroscopic surveys and conducting follow-up studies on these intriguing sources. Our mission is to uncover rare quasar phenomena to enhance our understanding of supermassive black holes and the extreme environments they inhabit.

About the Project

Quasars are among the brightest and most energetic objects in the universe, powered by supermassive black holes at the centers of distant galaxies. While many quasars have been studied extensively, anomalous quasars—those with unusual spectral characteristics—offer unique insights into astrophysical processes.

Project 1 : Spectroscopic Quasar Anomaly Detection (SQuAD) I: Rest-Frame UV Spectra from SDSS DR16 : The Spectroscopic Quasar Anomaly Detection (SQuAD 1) project is designed to identify and categorize spectroscopically anomalous quasars from large-scale astronomical surveys, such as the SDSS DR16 catalog. By utilizing advanced techniques like Principal Component Analysis (PCA) for dimensionality reduction and hierarchical K-Means clustering, the project focuses on finding rare or unusual quasars with distinct properties. In this first phase, SQuAD 1 concentrates on rest-frame UV spectra to detect anomalies in quasar spectra.

Quasar spectrum

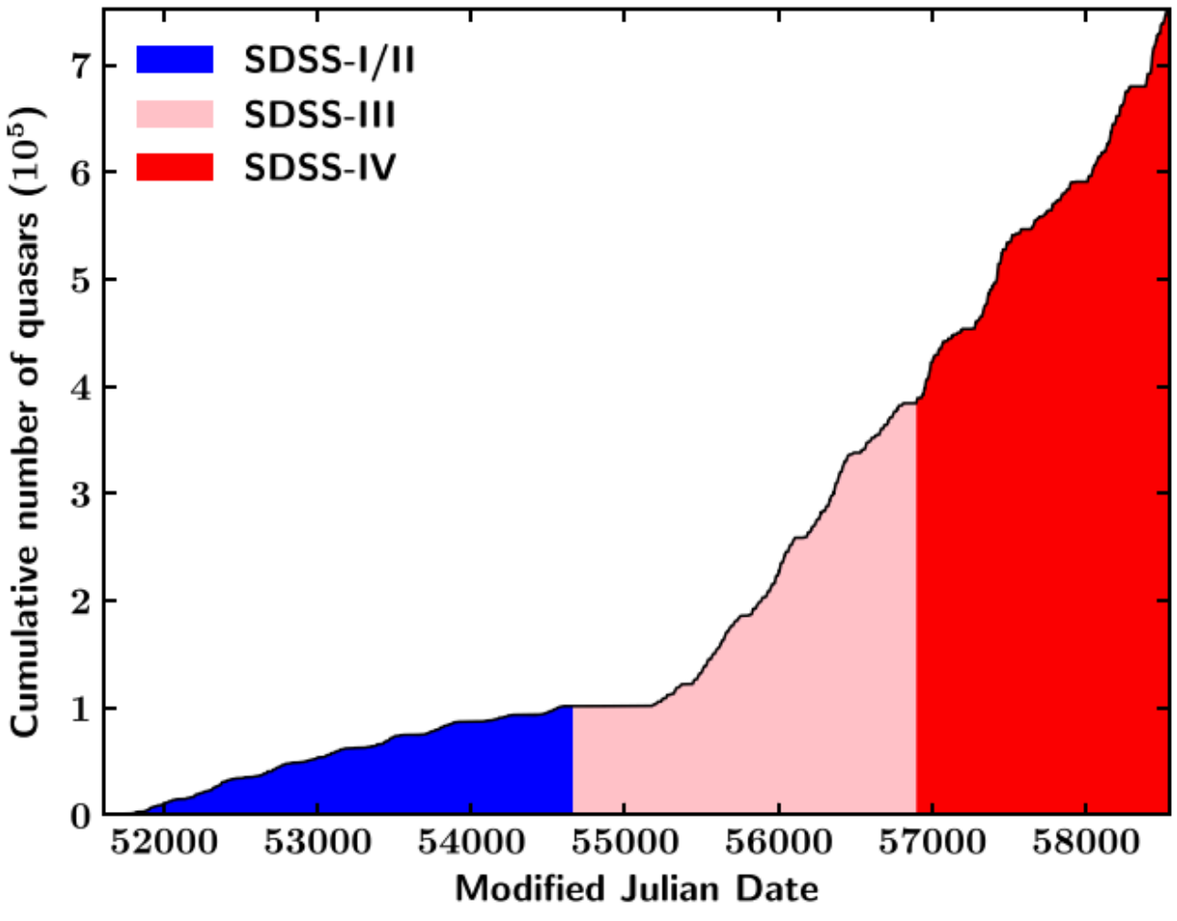


vanden berk et al 2001

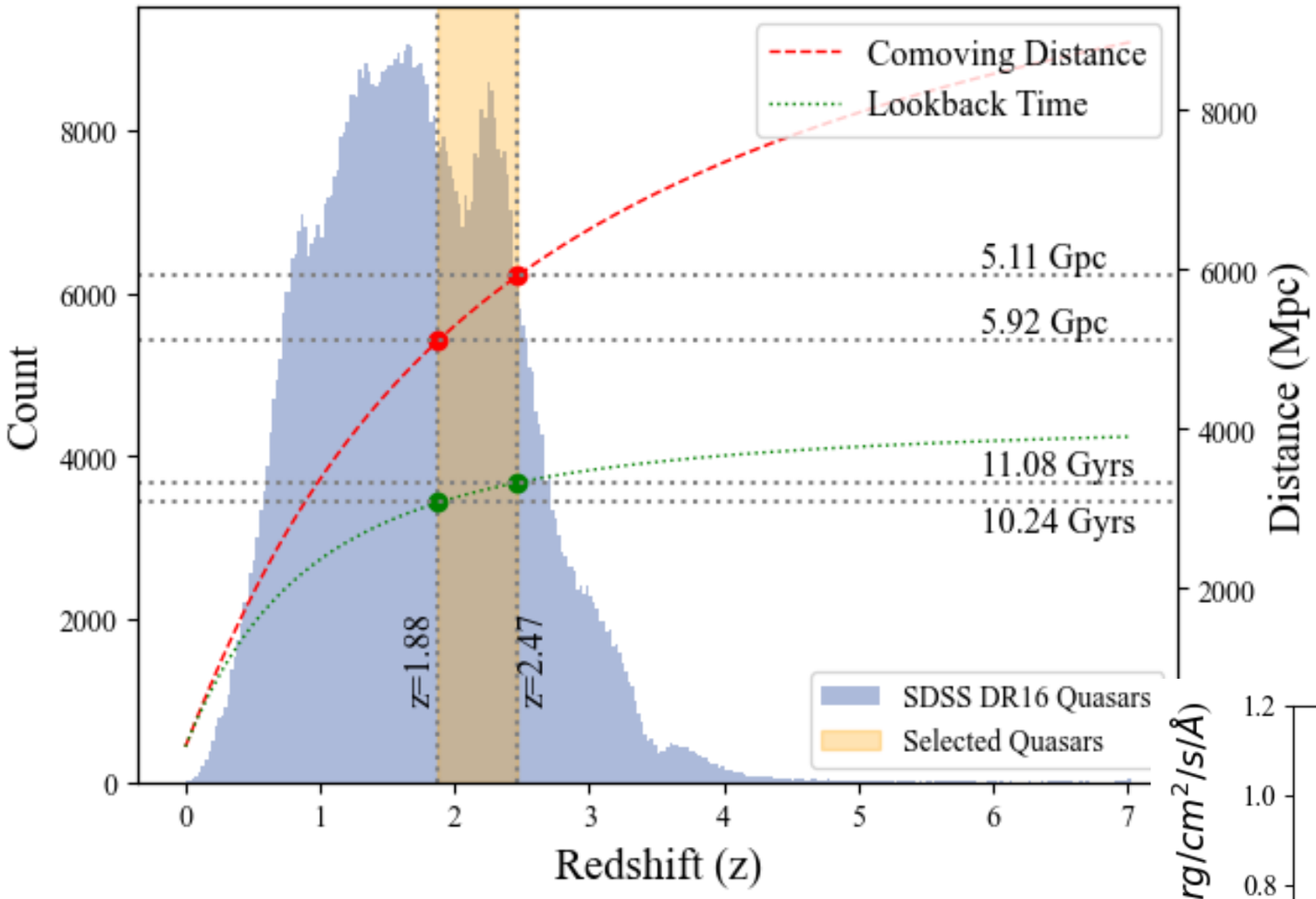
SDSS DR16 QSO Catalog Sample

- Quasar catalog from Data Release 16 of the extended Baryon Oscillation Spectroscopic Survey (eBOSS).
- The largest selection of spectroscopically confirmed quasars to date. (750,414 quasars)
- 99,856 broad absorption-line quasars and 35,686 damped Lyman alpha quasars.

Lyke et. al. 2020



SDSS DR16 QSO Catalog Sample



$1.88 \leq z \leq 2.47$

SNR ≥ 5

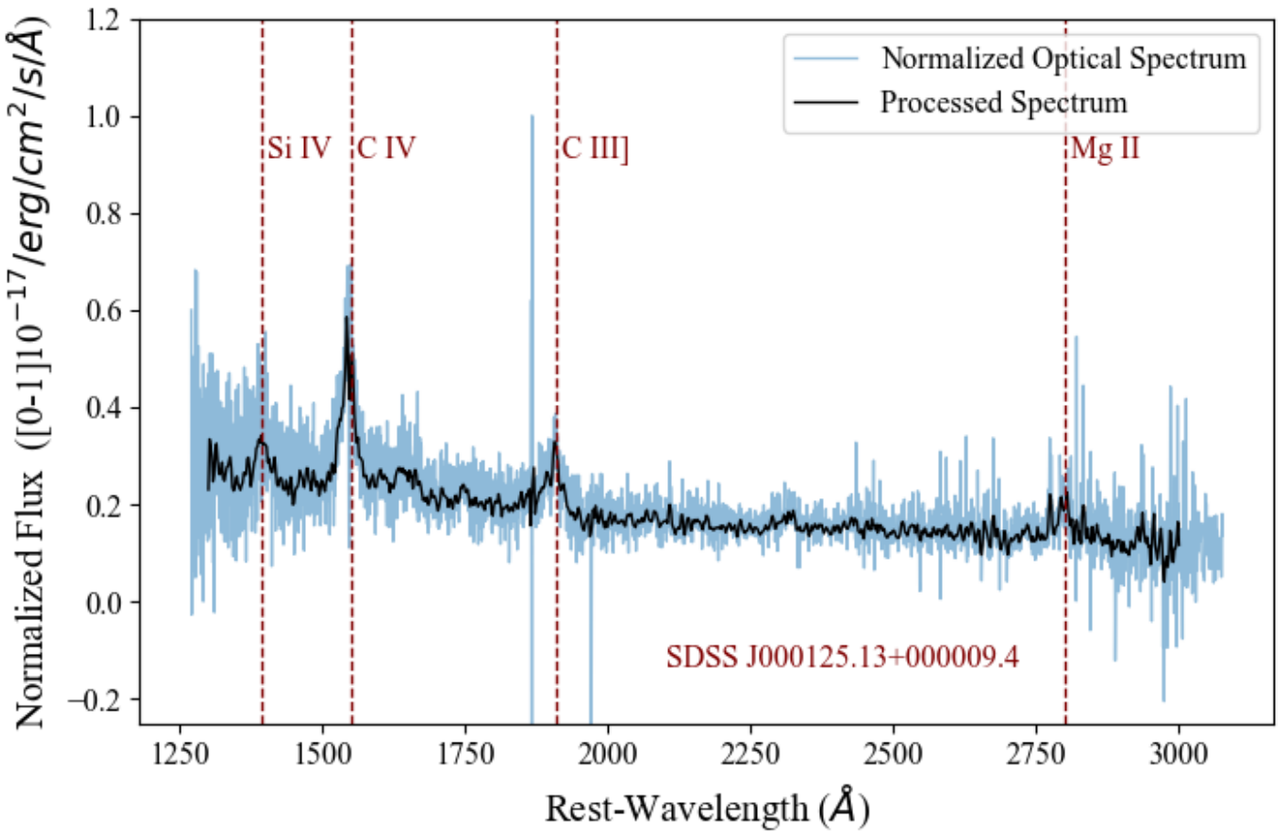
ZWARNING = 0

IS QSO FINAL = 1

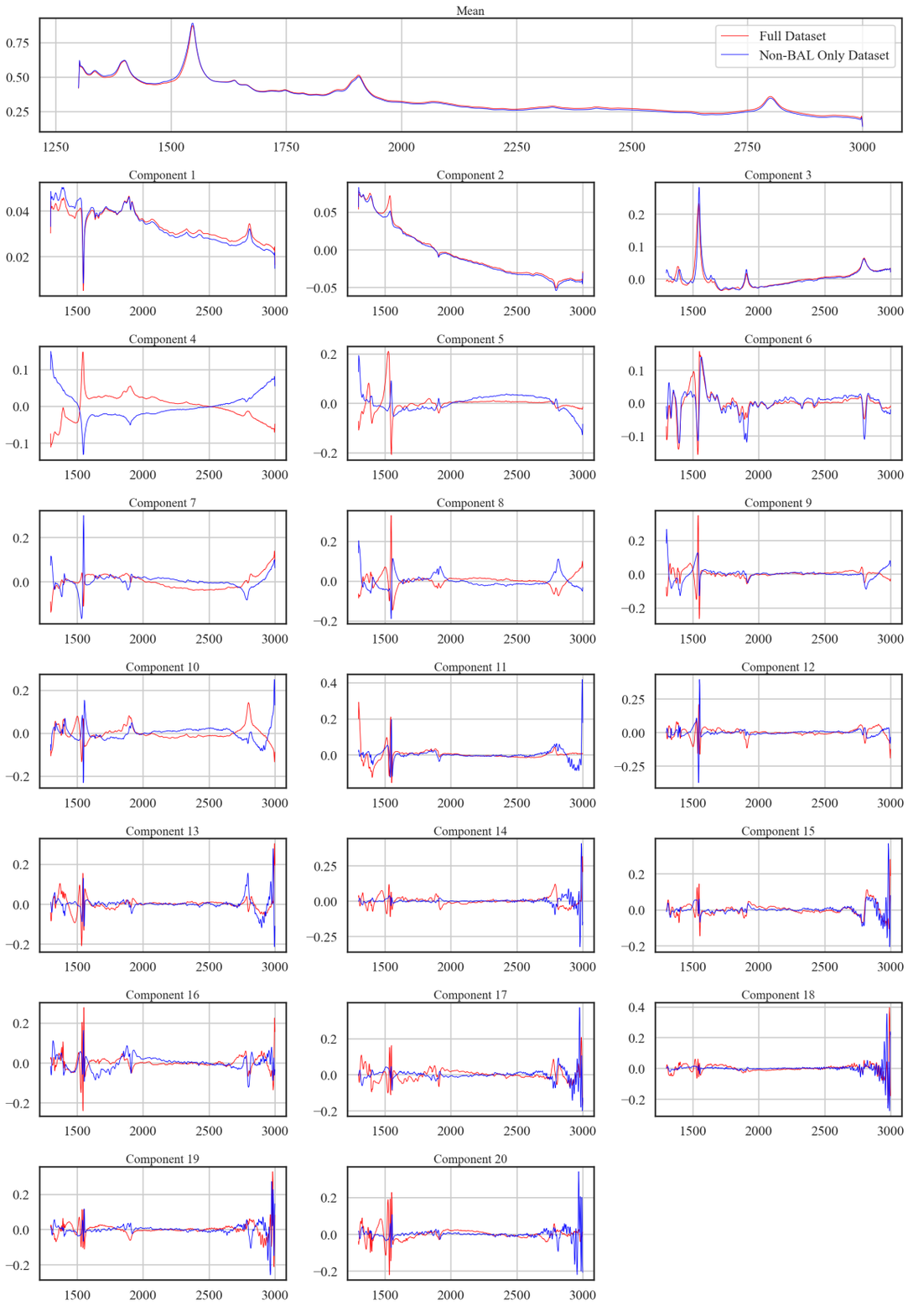
Full : 81,814

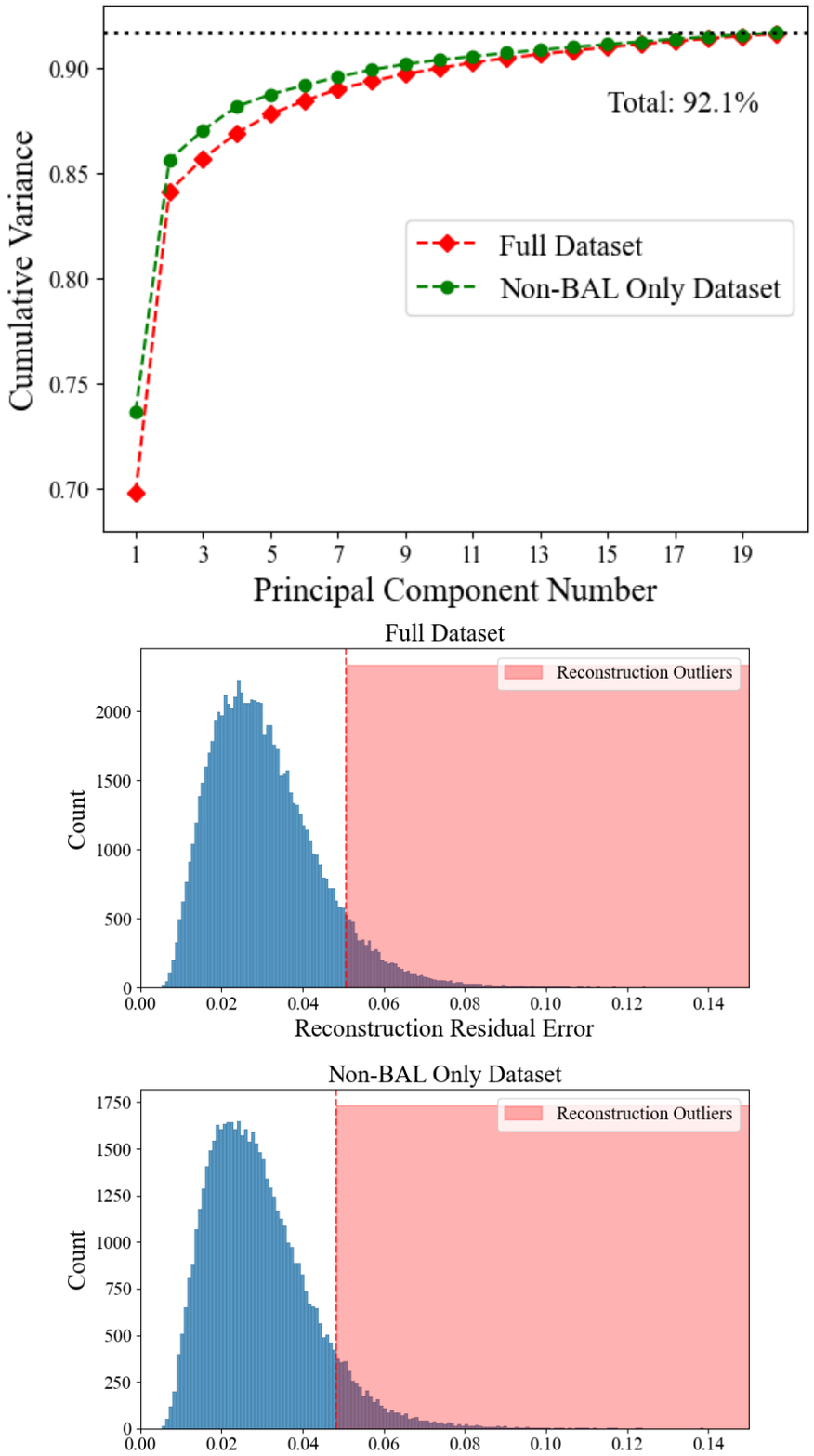
Non-BAL Only : 55,245

Max-normalization
Smoothing
Resampling
Padding

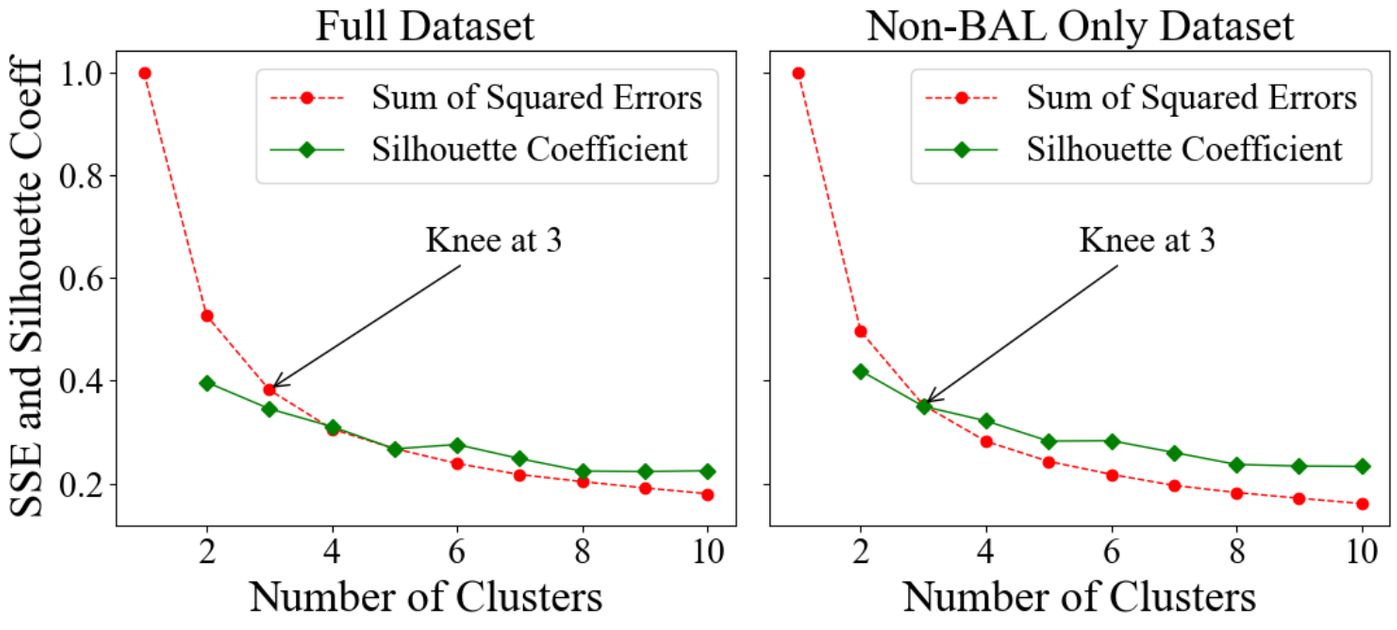


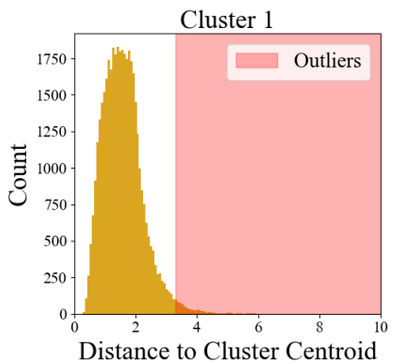
Dimensionality Reduction (PCA)

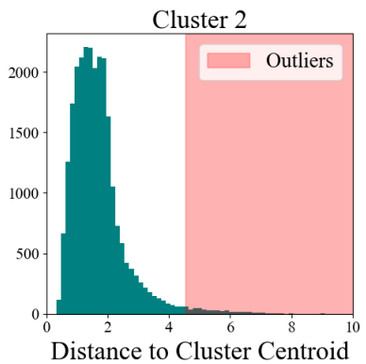


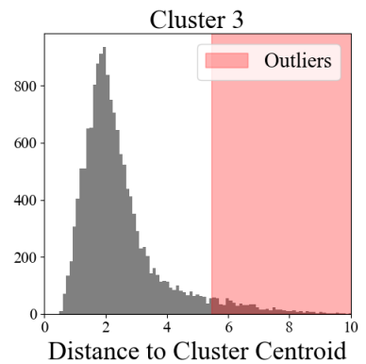


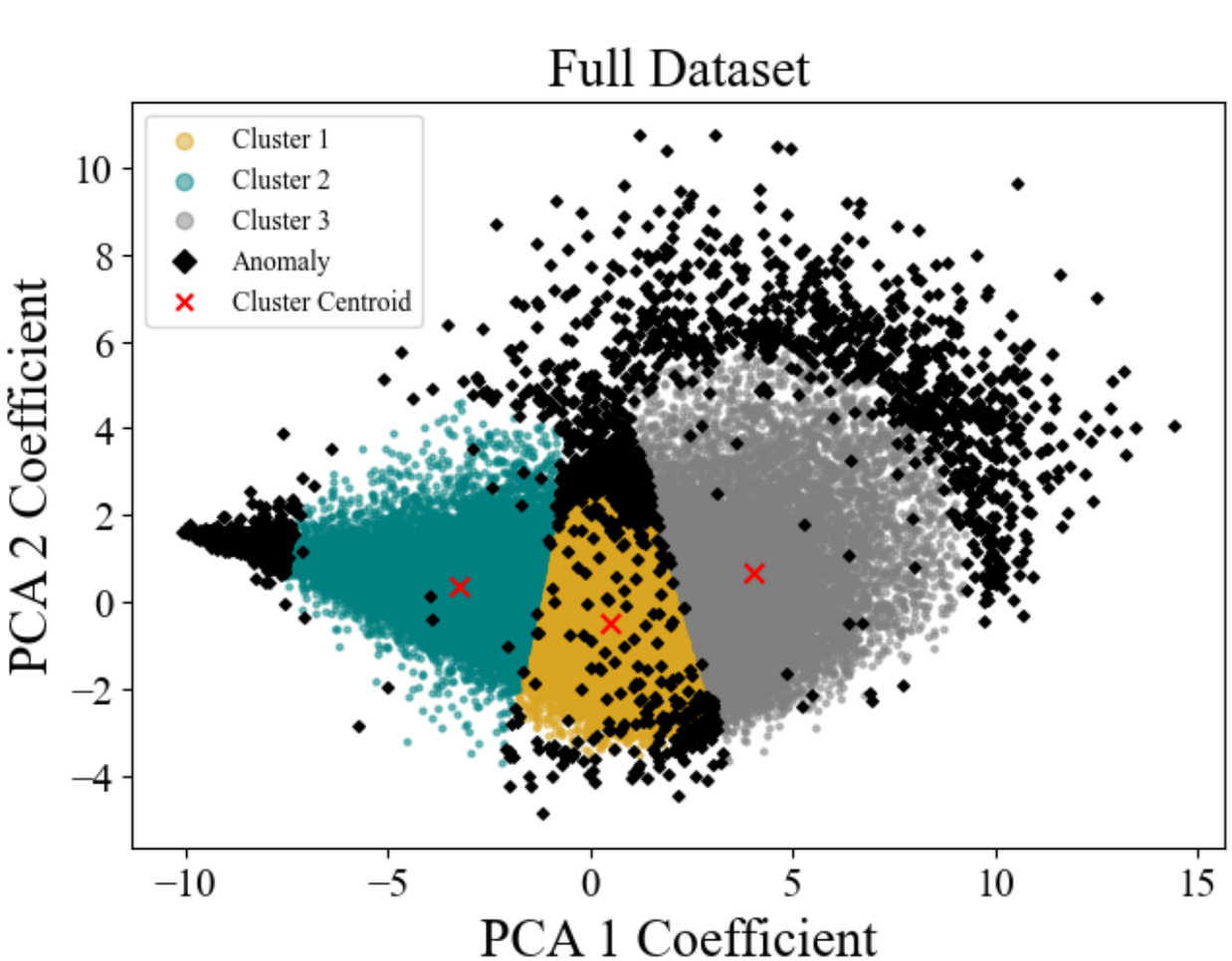
k-Means Clustering

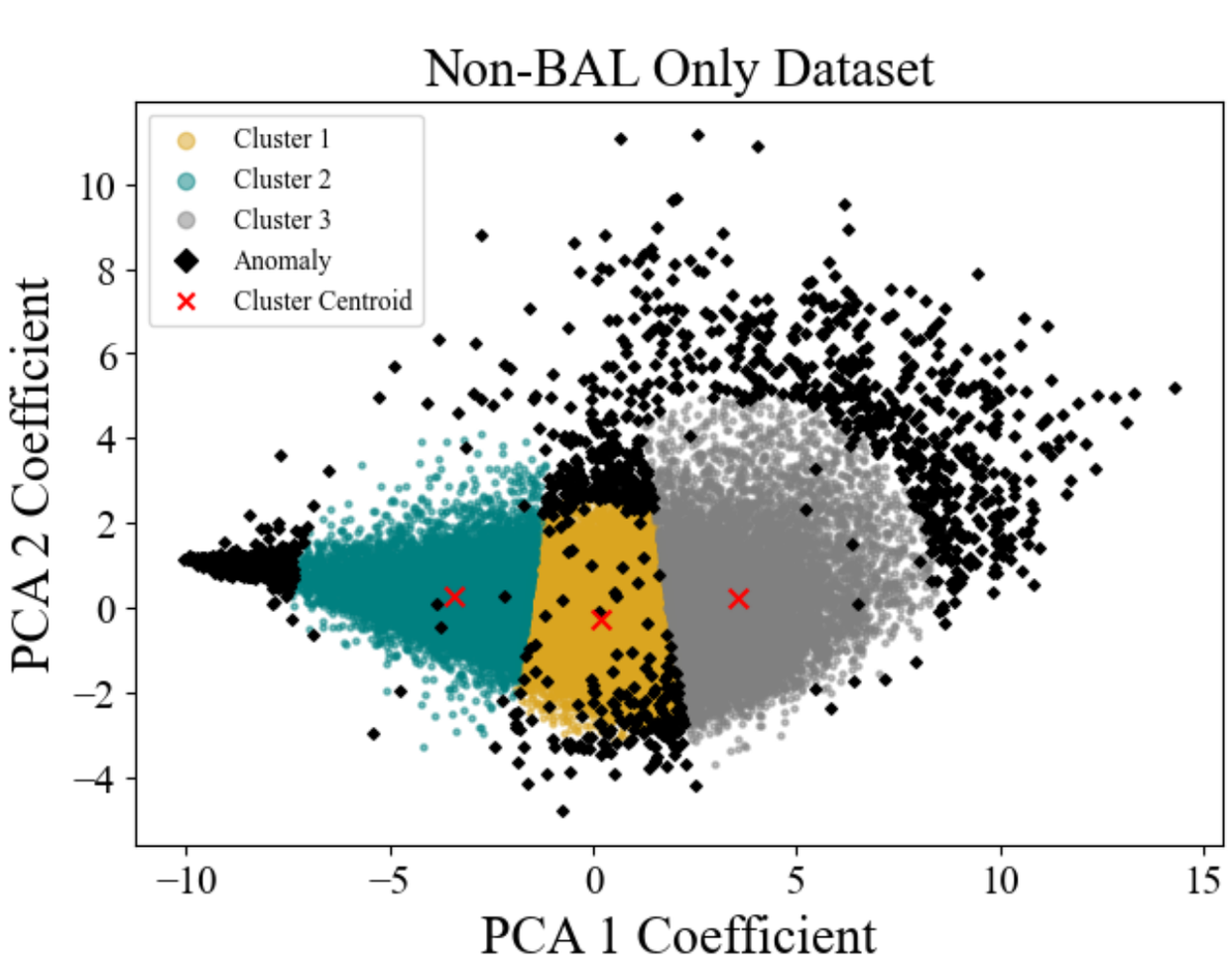




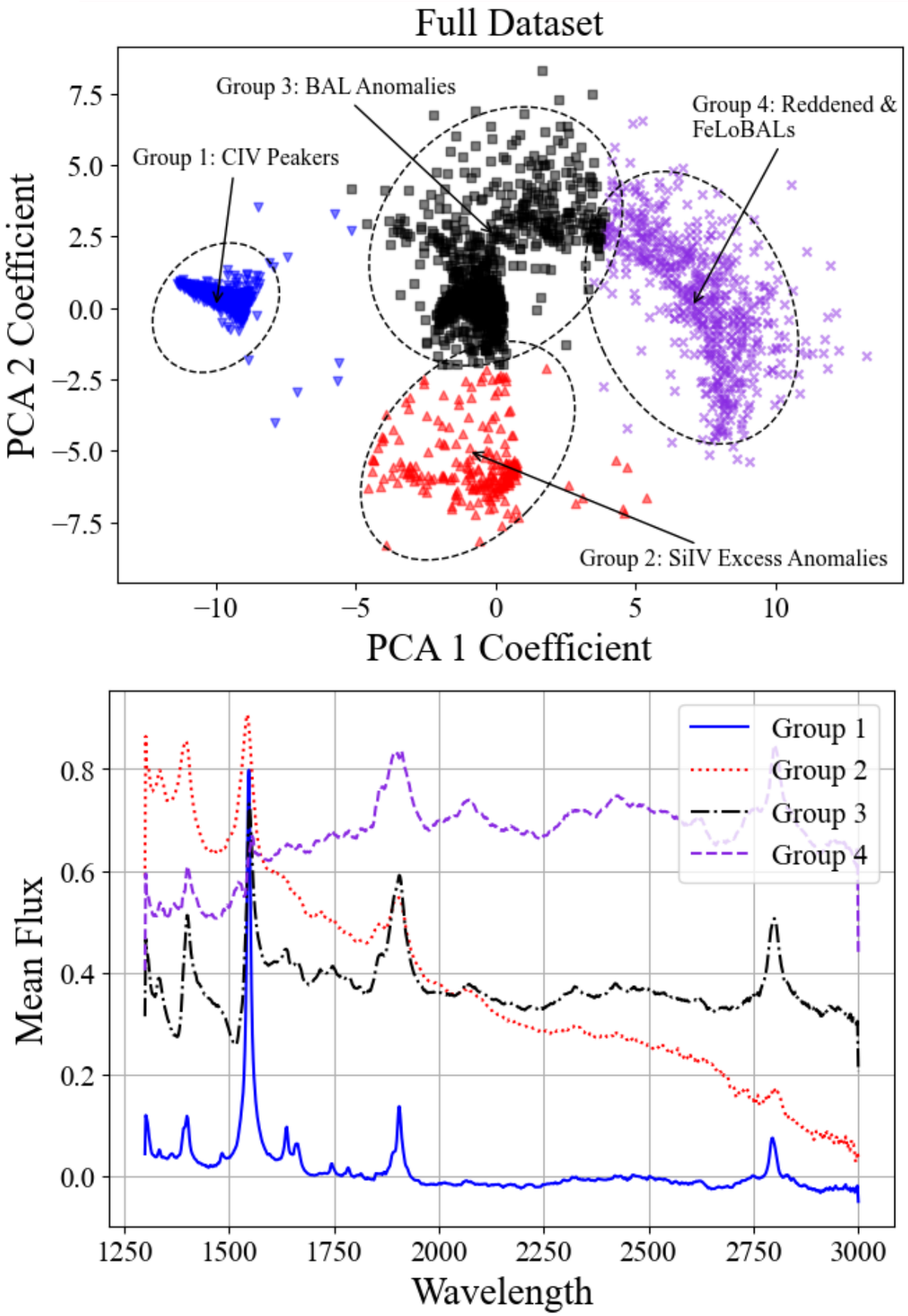


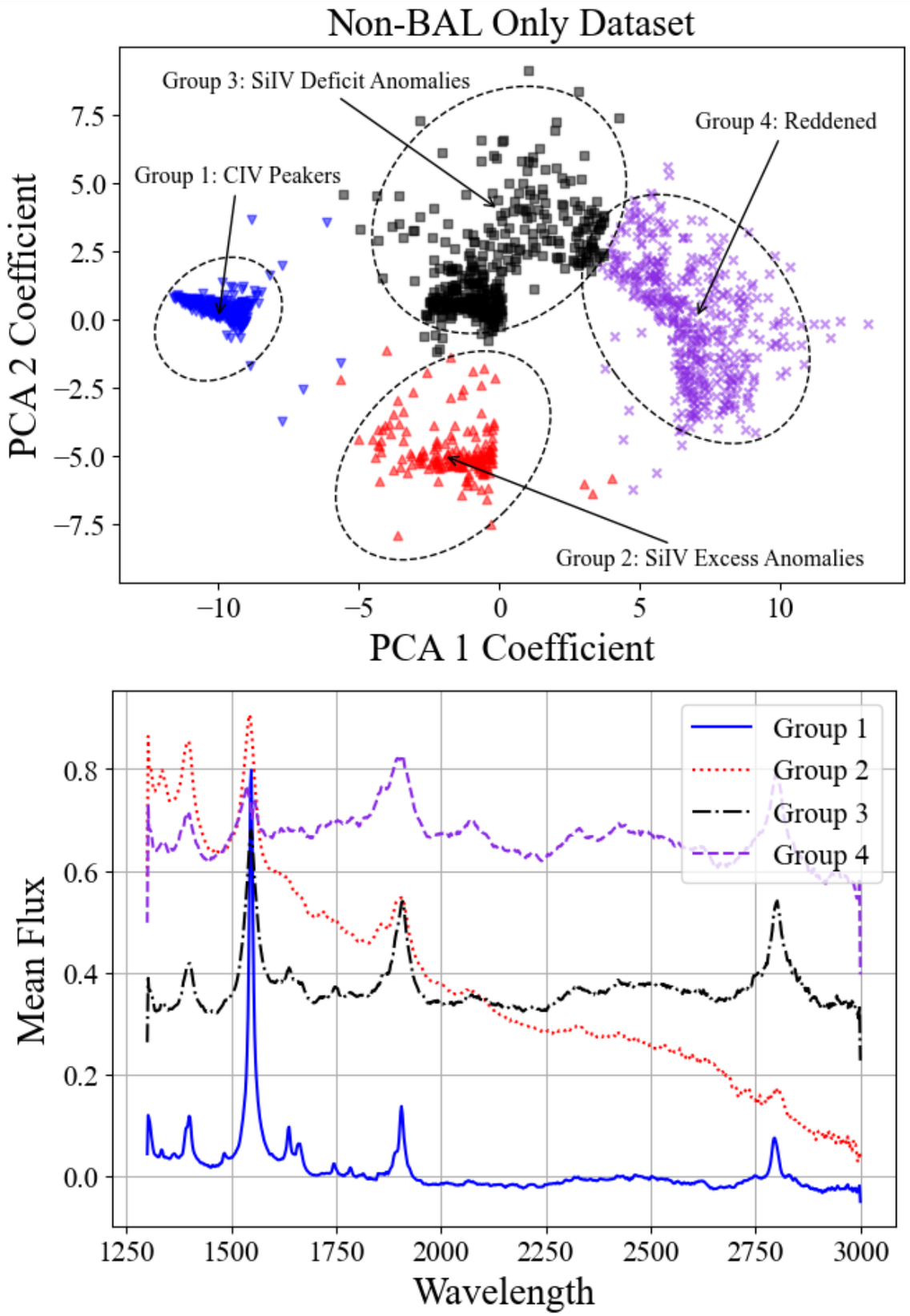




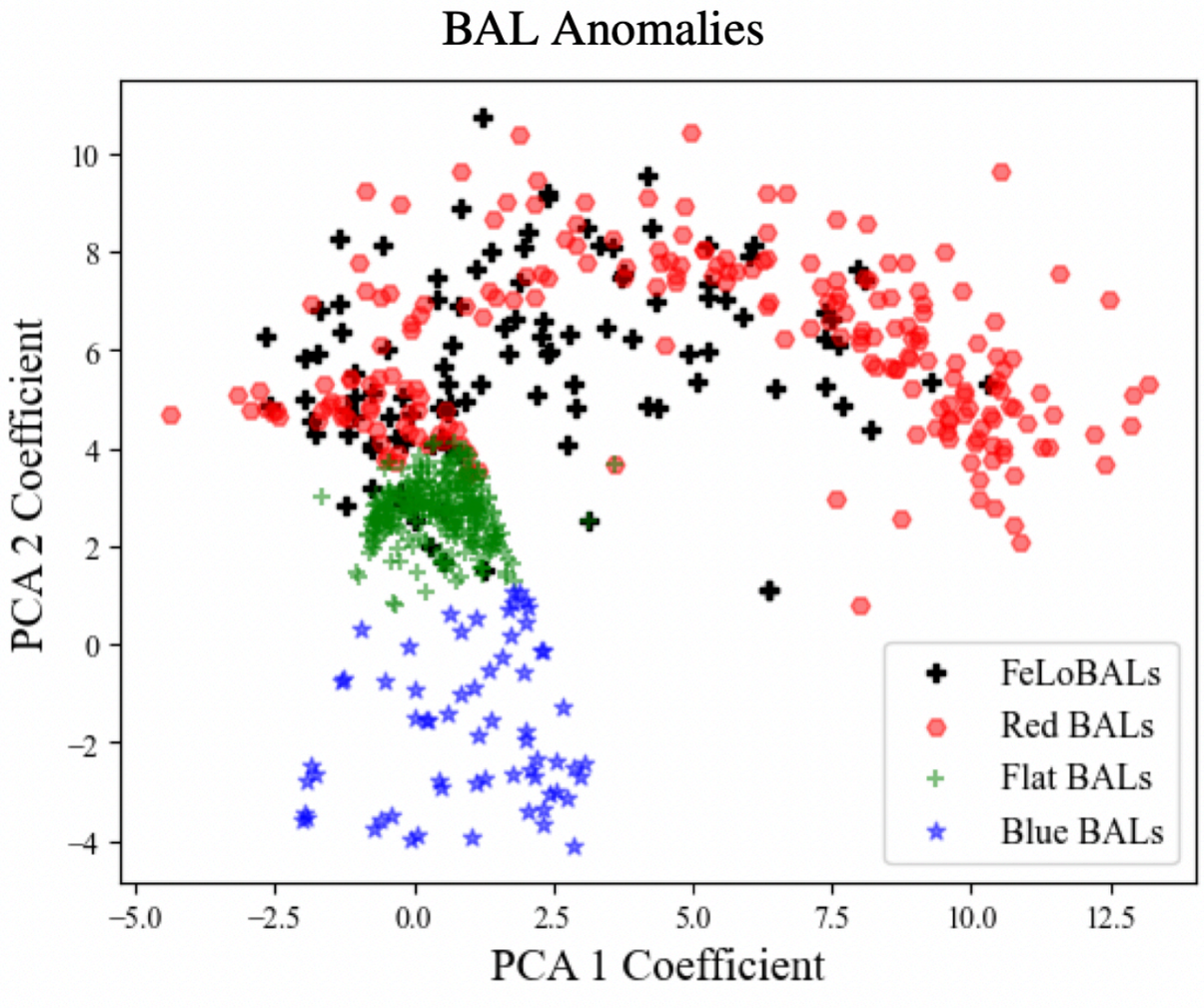


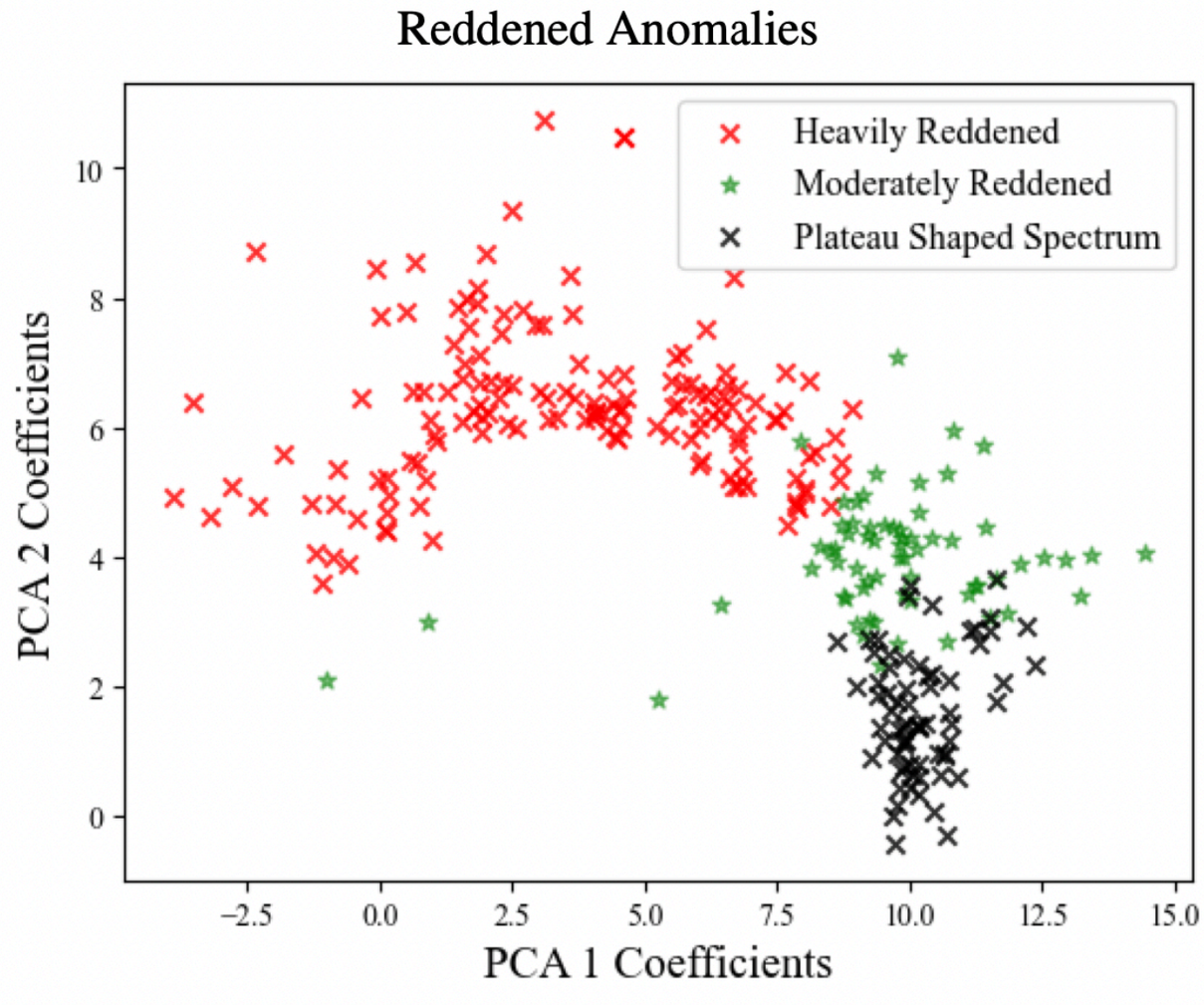
Outliers


Full Dataset
Group 3: BAL Anomalies
Group 4: Reddened & FeLoBALs
Group 1: CIV Peakers
Group 2: SiIV Excess Anomalies
PCA 2 Coefficient
PCA 1 Coefficient
Group 1
Group 2
Group 3
Group 4
Mean Flux
Wavelength


Non-BAL Only Dataset
Group 3: SiIV Deficit Anomalies
Group 4: Reddened
Group 1: CIV Peakers
Group 2: SiIV Excess Anomalies
PCA 2 Coefficient
PCA 1 Coefficient
Group 1
Group 2
Group 3
Group 4
Mean Flux
Wavelength

BAL & Reddened Anomalies

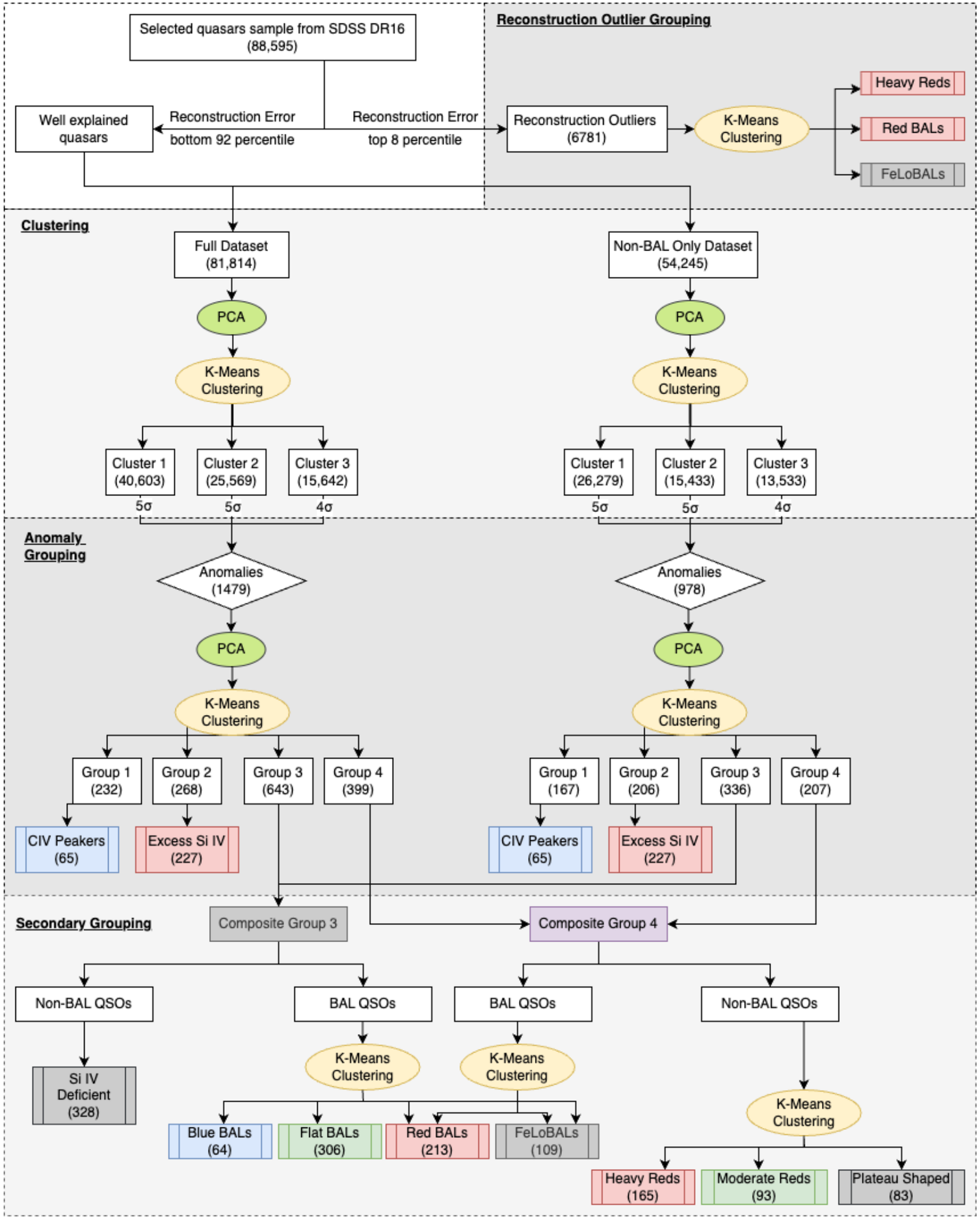




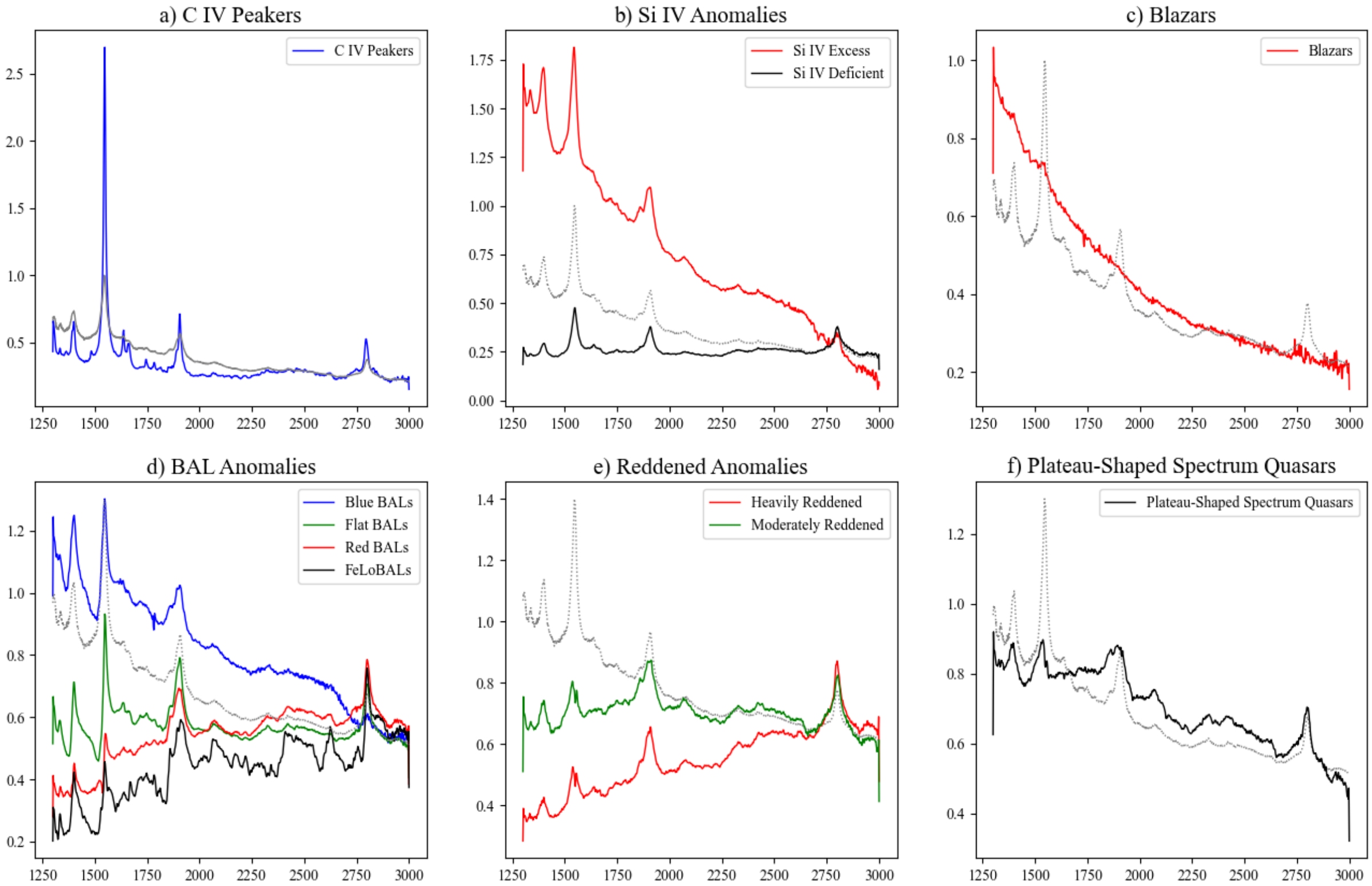
Algorithm

1,888 anomalous quasars, categorised into 10 broad groups.

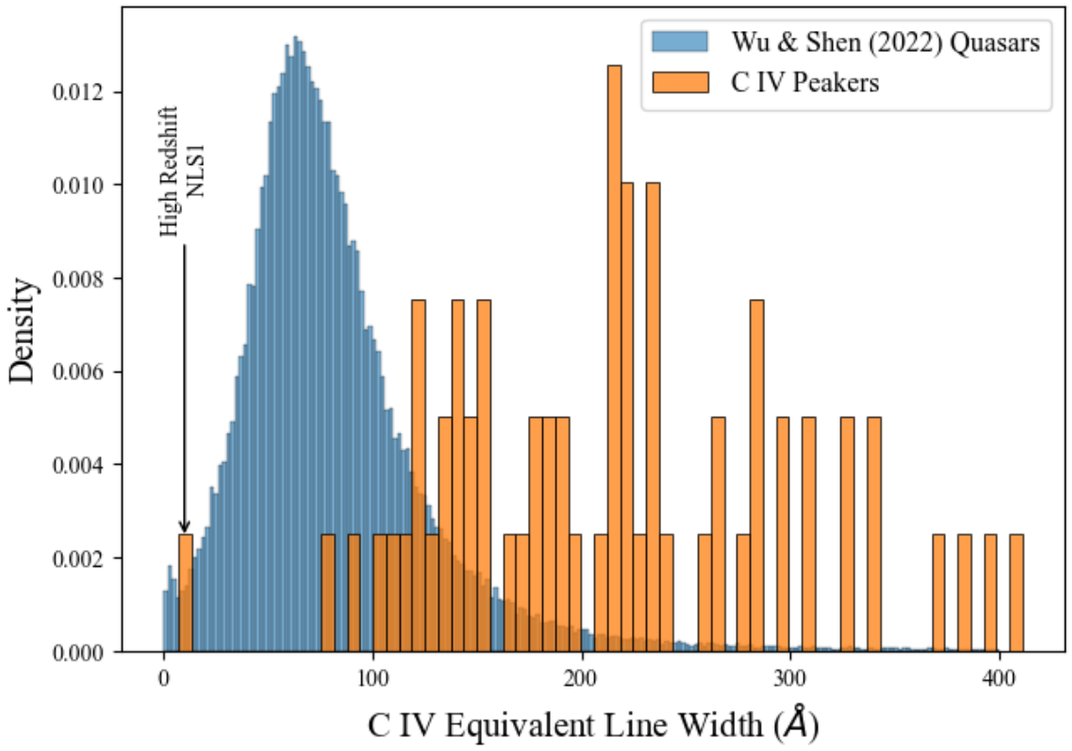
- C IV Peakers : 65
- Excess Si IV emitters : 227
- Si IV Deficient Anomalies : 328
- BAL Anomalies : 692
  - Blue BALs : 64
  - Flat BALs : 306
  - Red BALs : 231
  - FeLoBALs : 109
- Reddened Anomalies : 341
  - Heavily Reddened : 165
  - Moderately Reddened : 93
  - Plateu-shaped Spectrum : 83

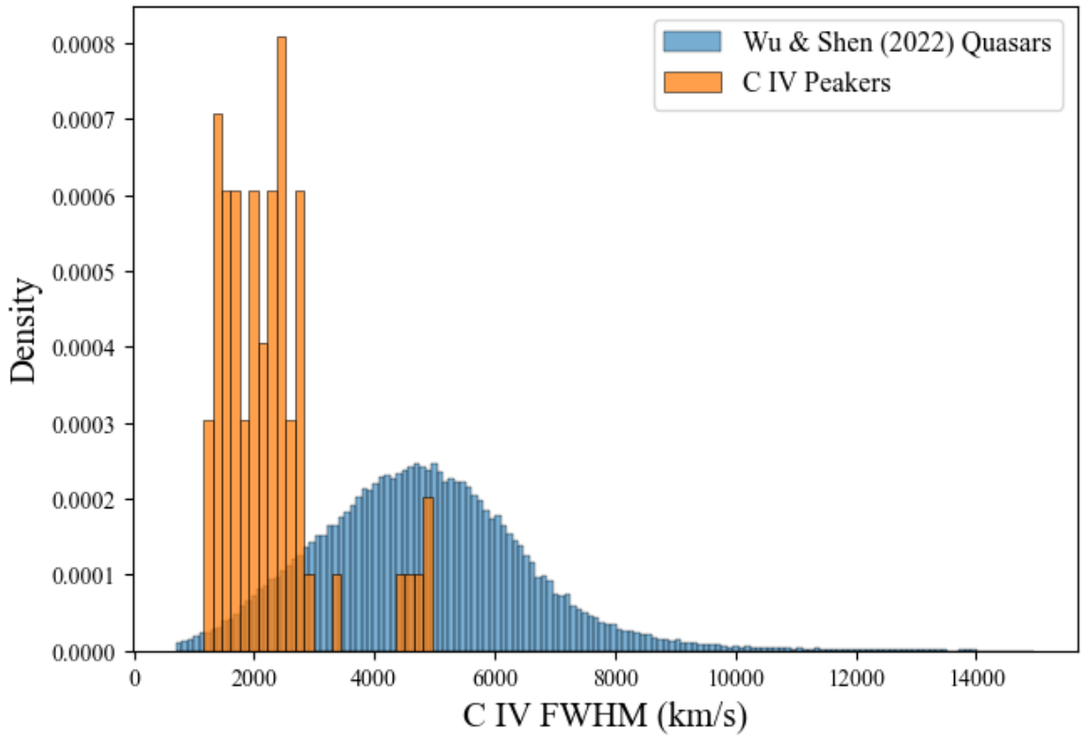


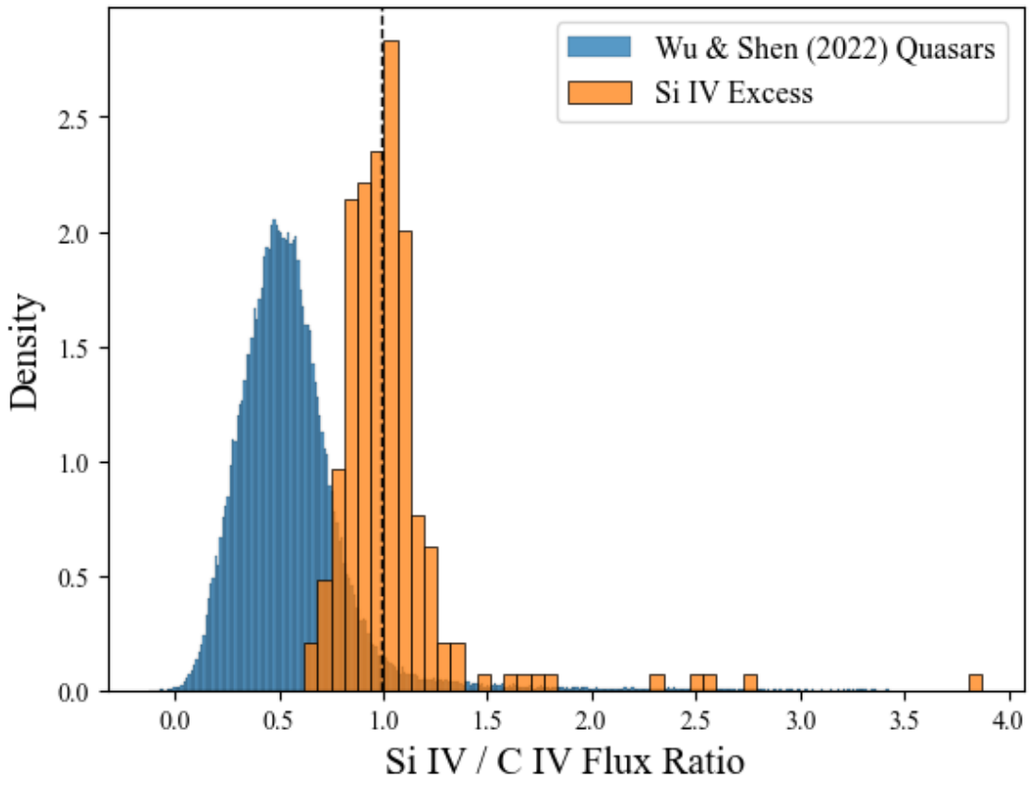
Composite Spectra

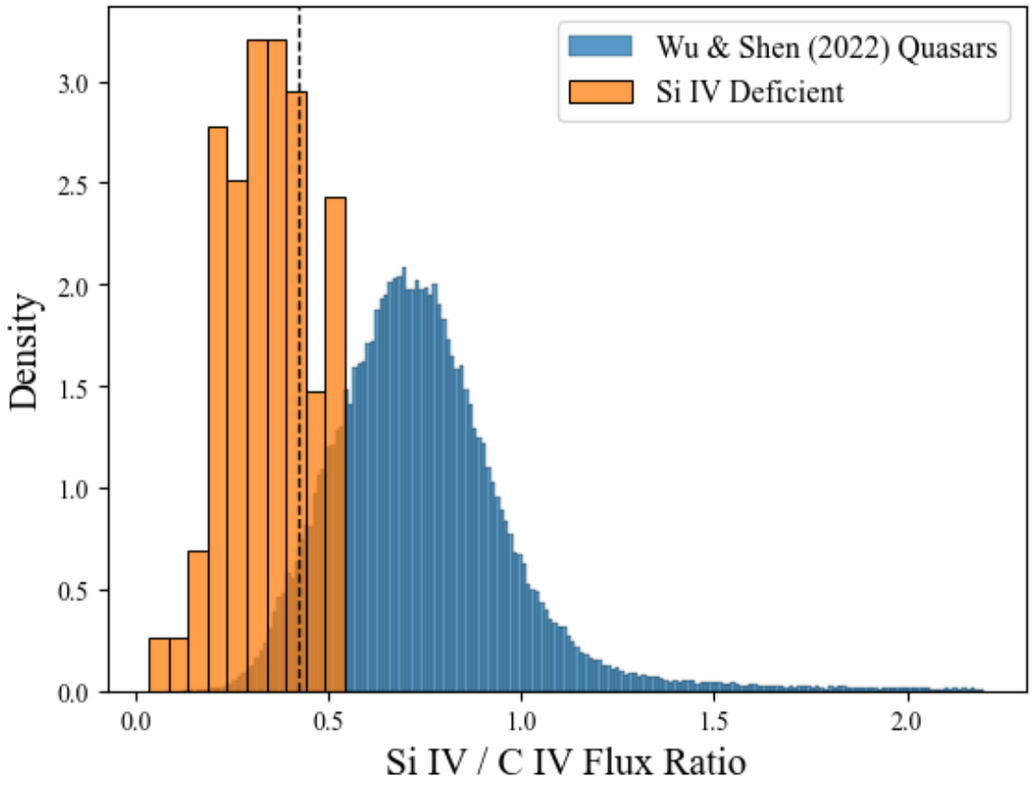


Line Ratios

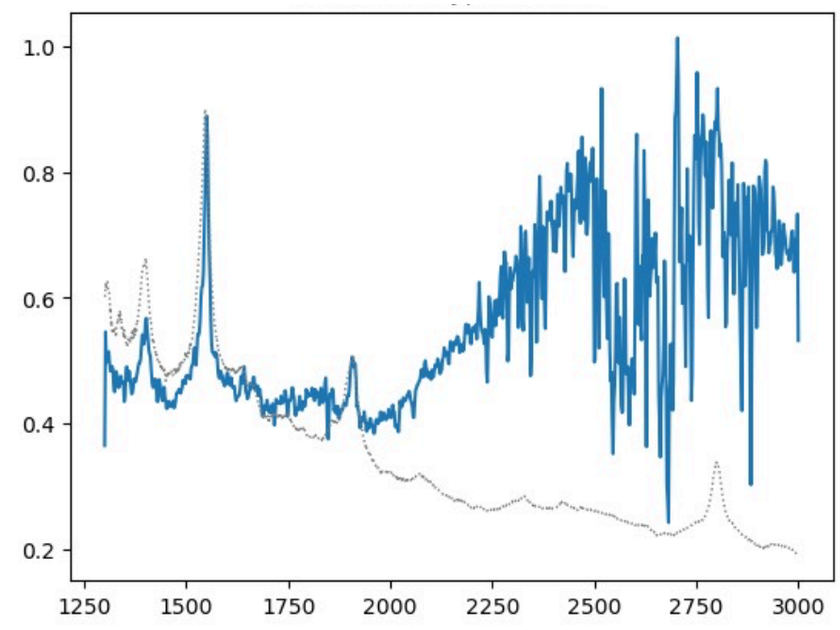

Wu & Shen (2022) Quasars
C IV Peakers
High Redshift
NLS1
Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0
100
200
300
400
C IV Equivalent Line Width (Å)

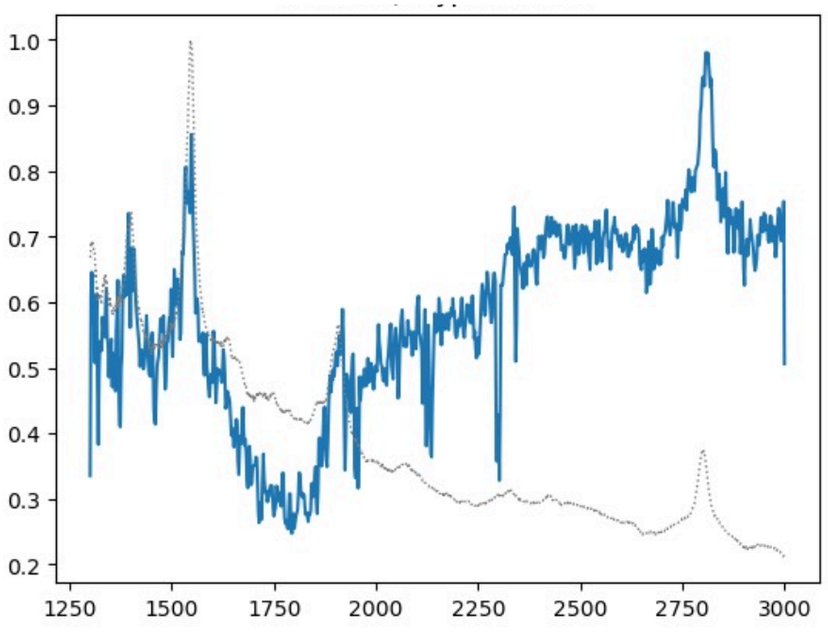

Wu & Shen (2022) Quasars
C IV Peakers
Density
0.0000
0.0001
0.0002
0.0003
0.0004
0.0005
0.0006
0.0007
0.0008
0
2000
4000
6000
8000
10000
12000
14000
C IV FWHM (km/s)

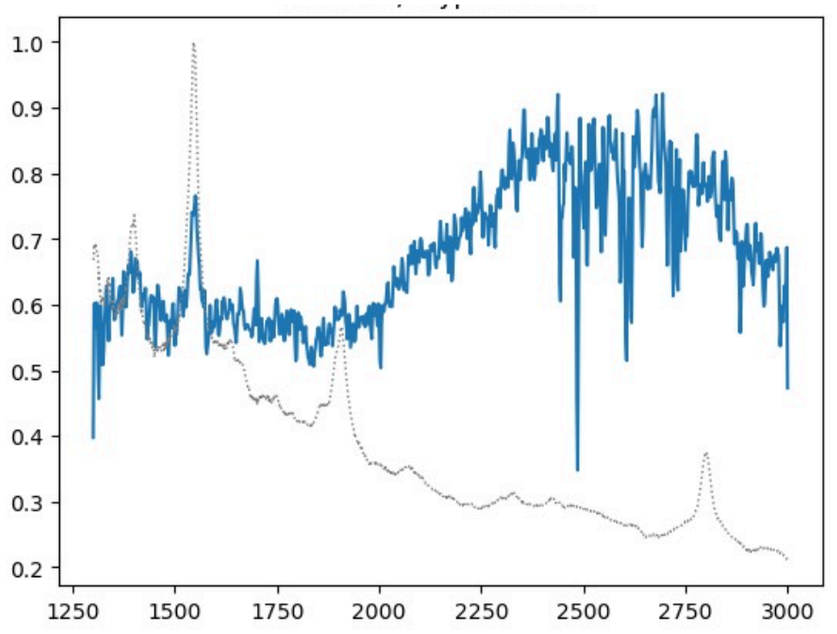

Wu & Shen (2022) Quasars
Si IV Excess
Density
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Si IV / C IV Flux Ratio


Wu & Shen (2022) Quasars
Si IV Deficient
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.5
1.0
1.5
2.0
Si IV / C IV Flux Ratio

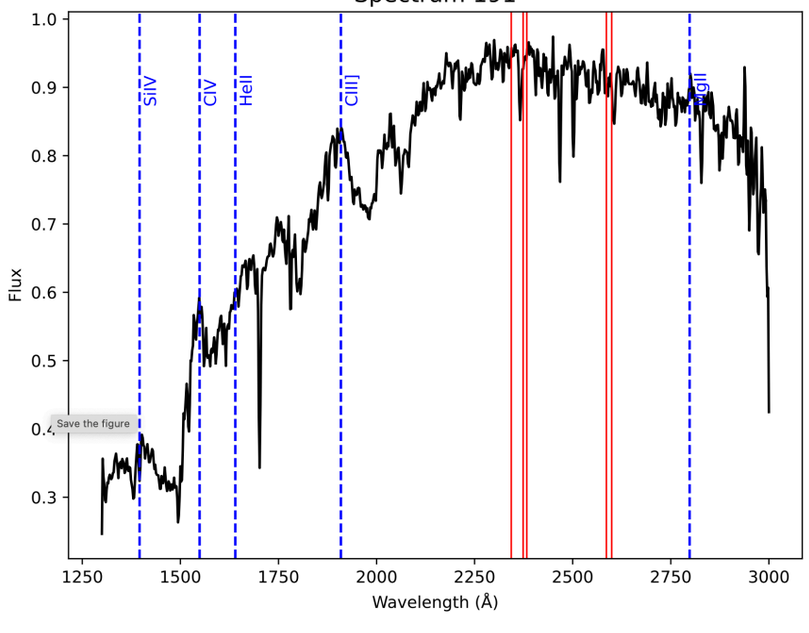
Some examples of really weird spectra !!!



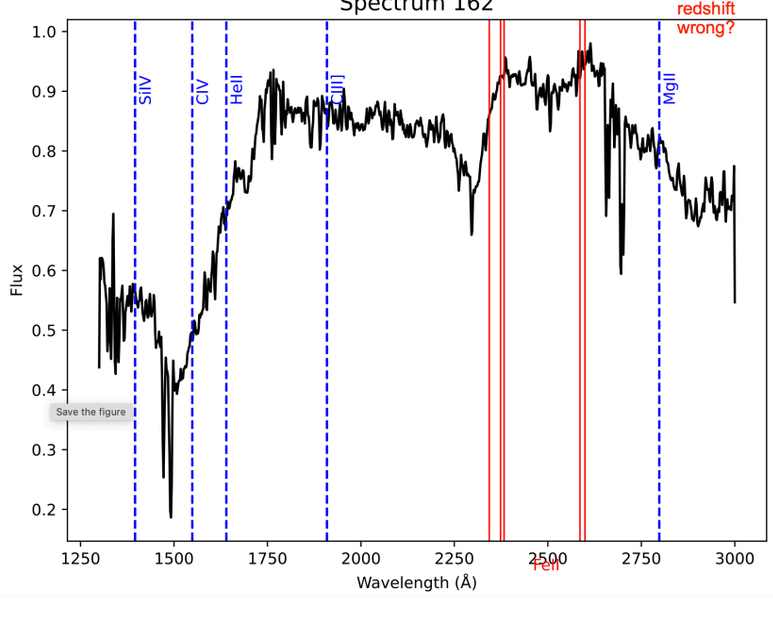




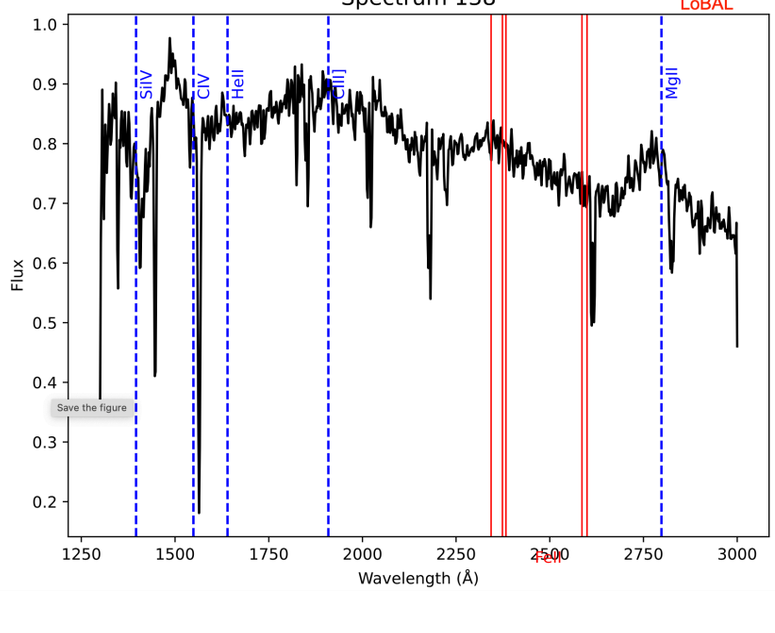
FeLOBAL
But weird



FeLoBAL-
weird
redshift
wrong?

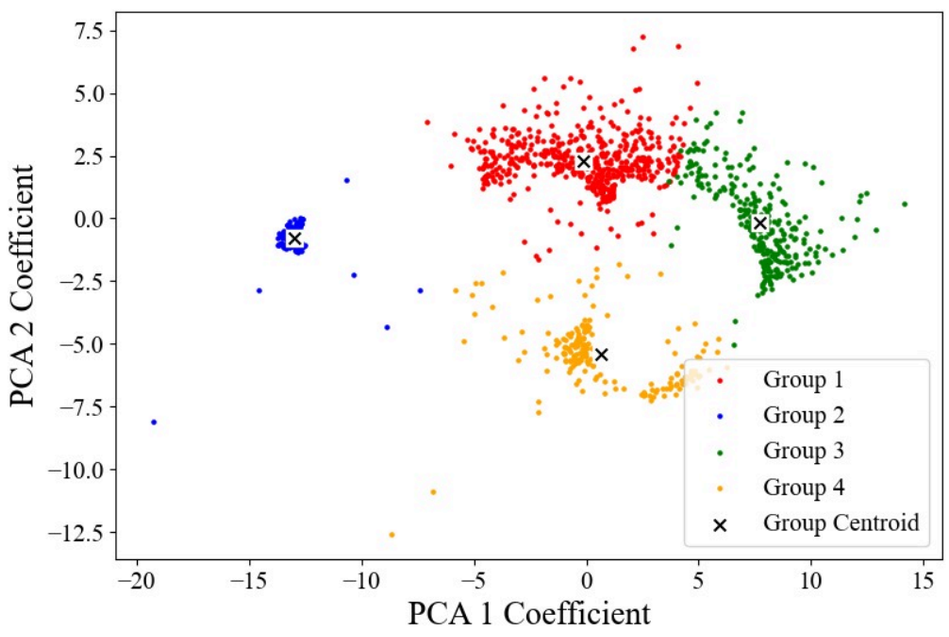


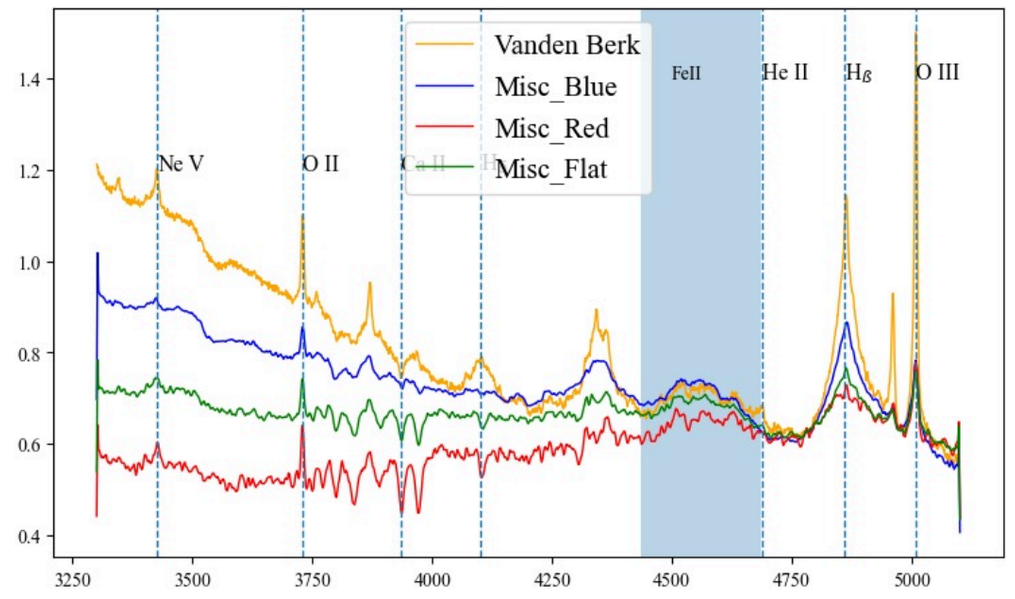
Redshifted
LoBAL



Going Forward

- Photo-ionization simulations to understand what physical conditions lead to the anomalous line ratios.
- Follow-up observations of weird spectra
- Connection between dust and BAL properties ?
- Rest-frame Optical wavelengths : Already implemented
  - Crucial in the Eigen vector 1 context

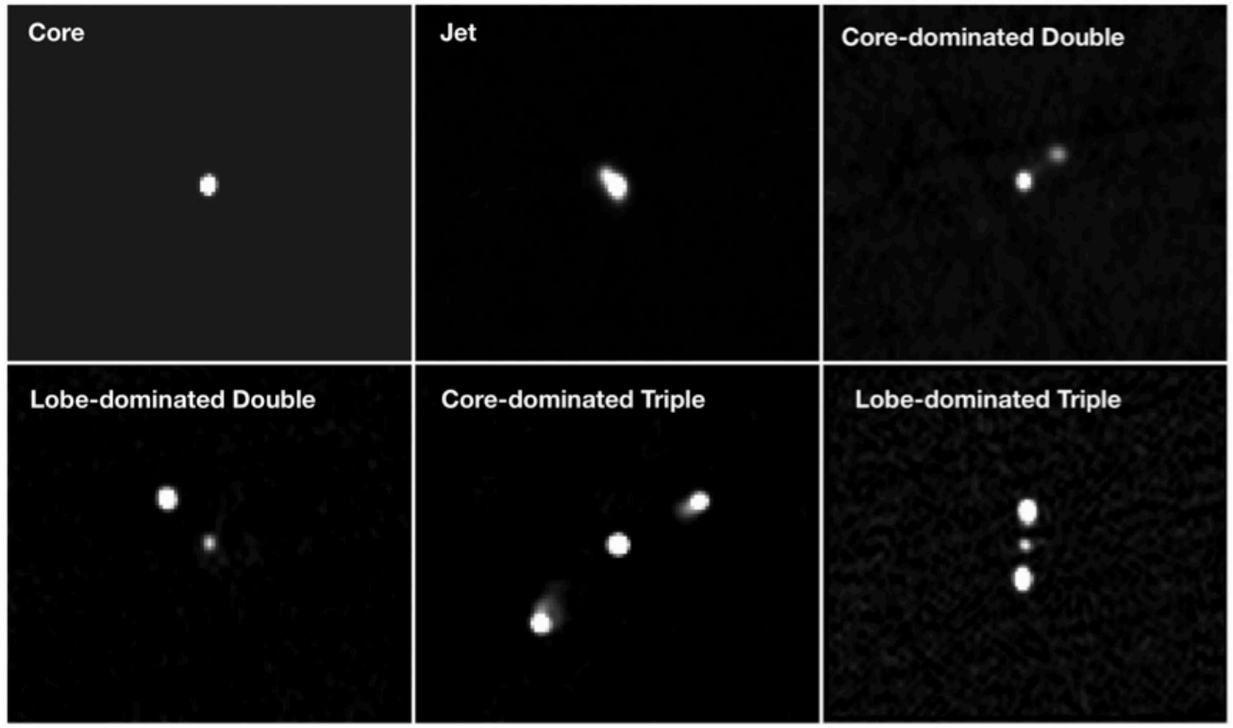


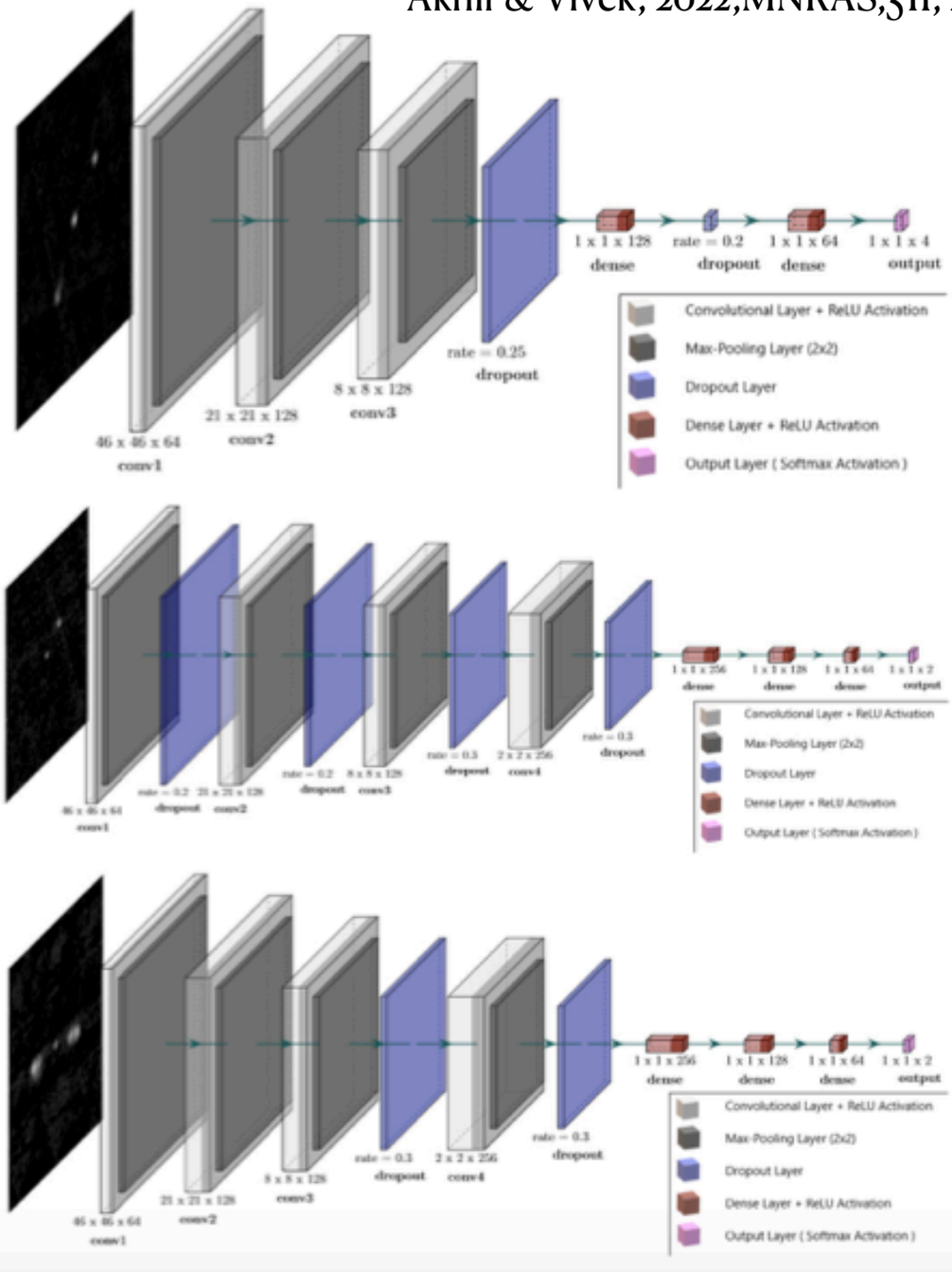


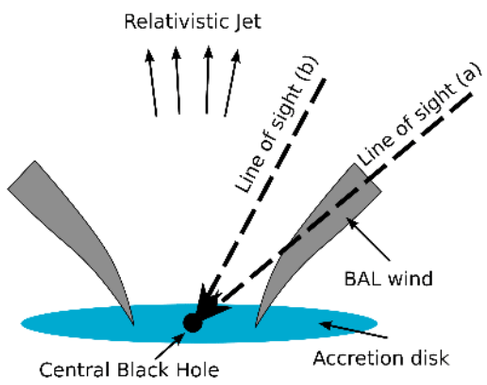
Radio Morphological classification of quasars

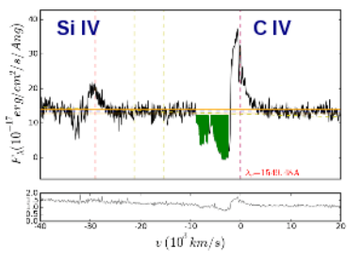
Fraction of broad absorption line quasars in different radio morphologies

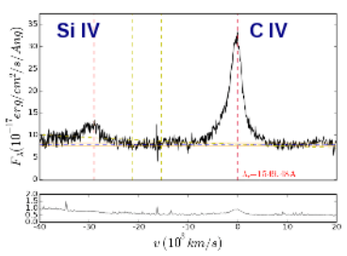
Akhil & Vivek, 2022,MNRAS,511, 4946







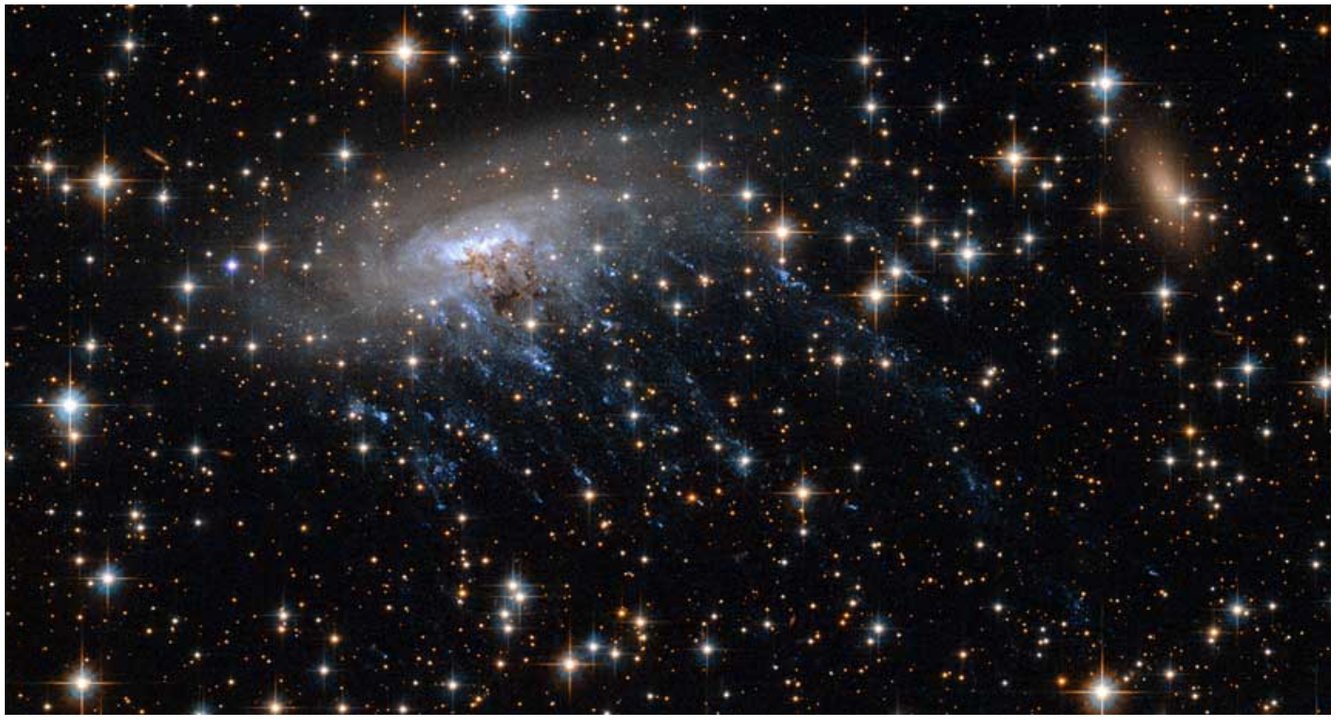


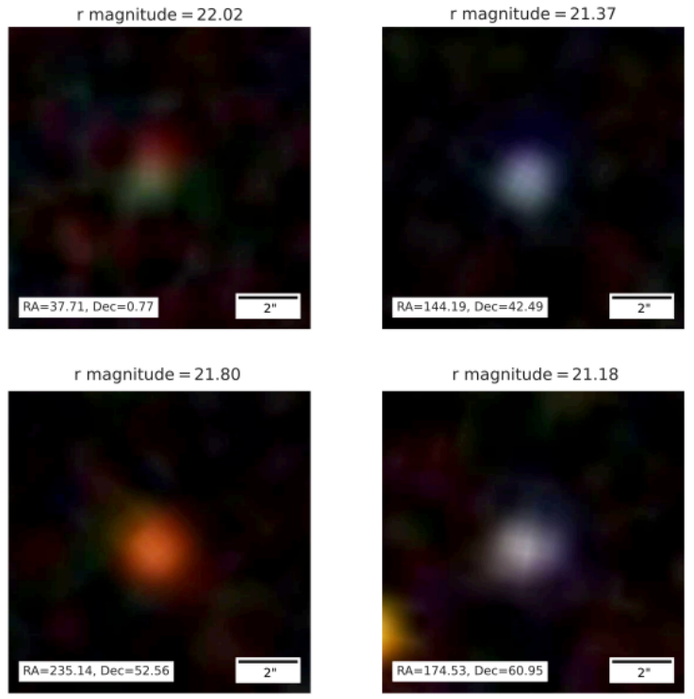


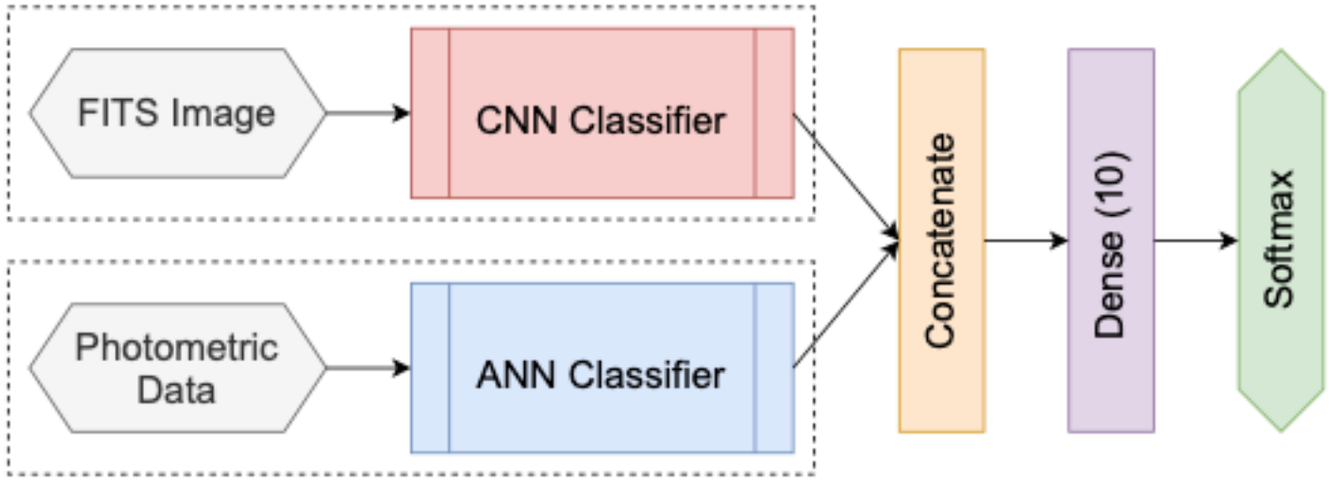
MargNet

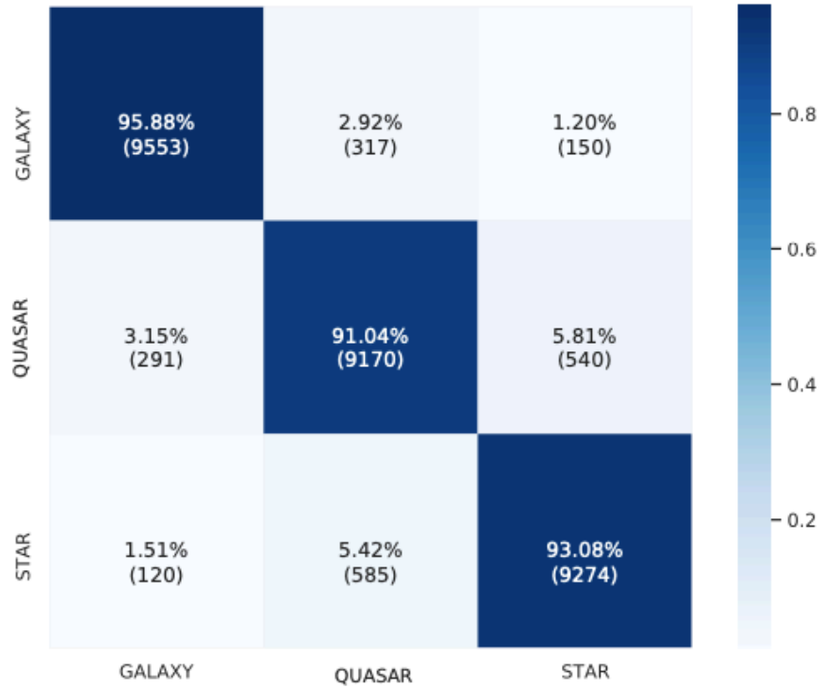
Photometric identification of compact galaxies, stars, and quasars using multiple neural networks

Siddharth Chaini, Atharva Bagul, Anish Deshpande, Rishi Gondkar, Vivek M, 2023, MNRAS, 518, 3123



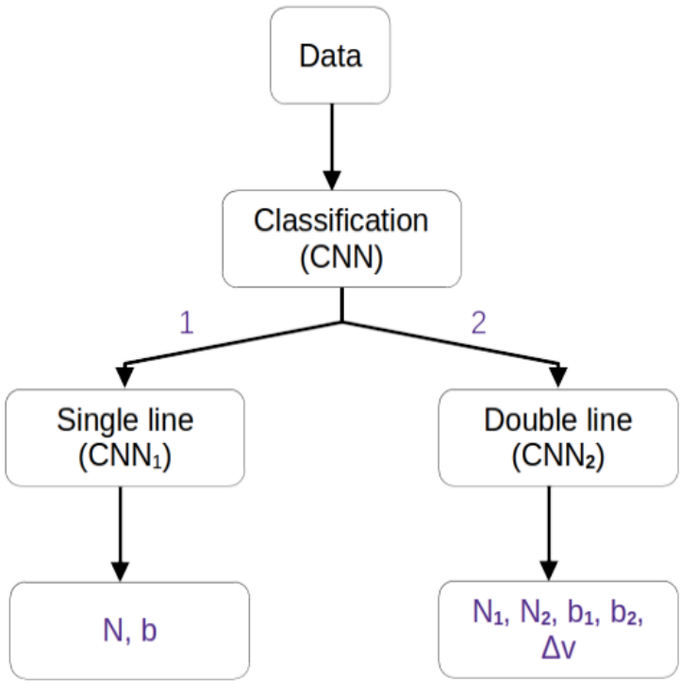


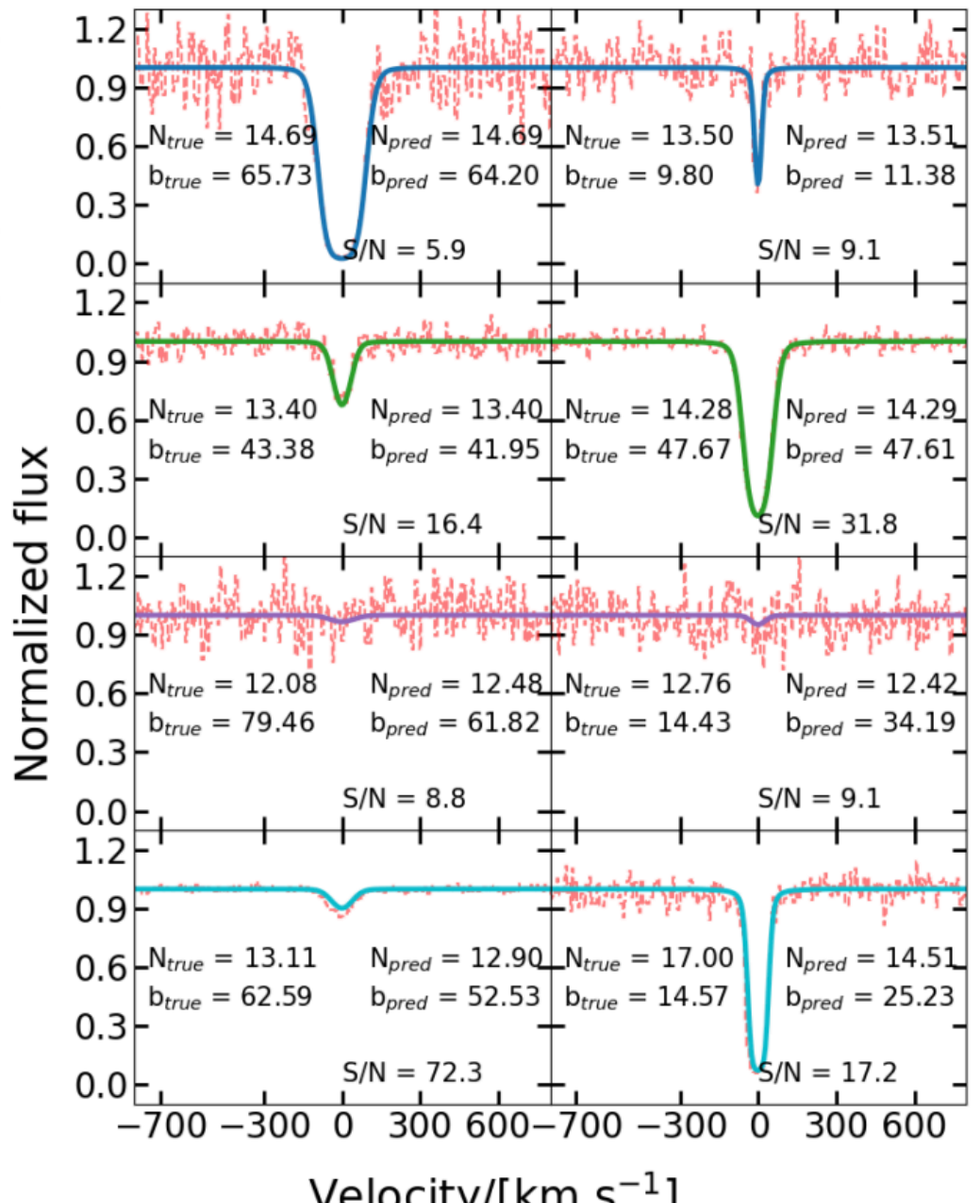


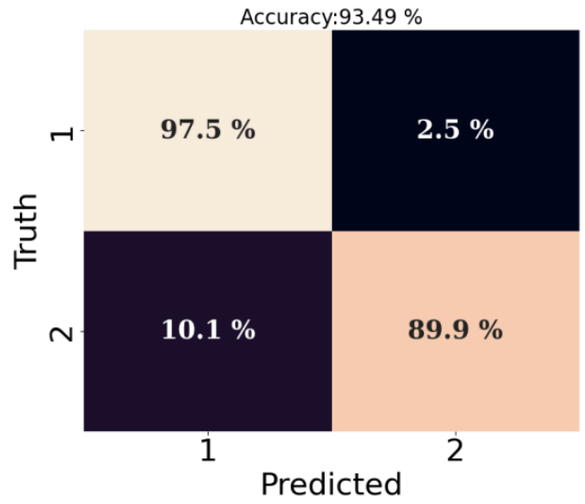


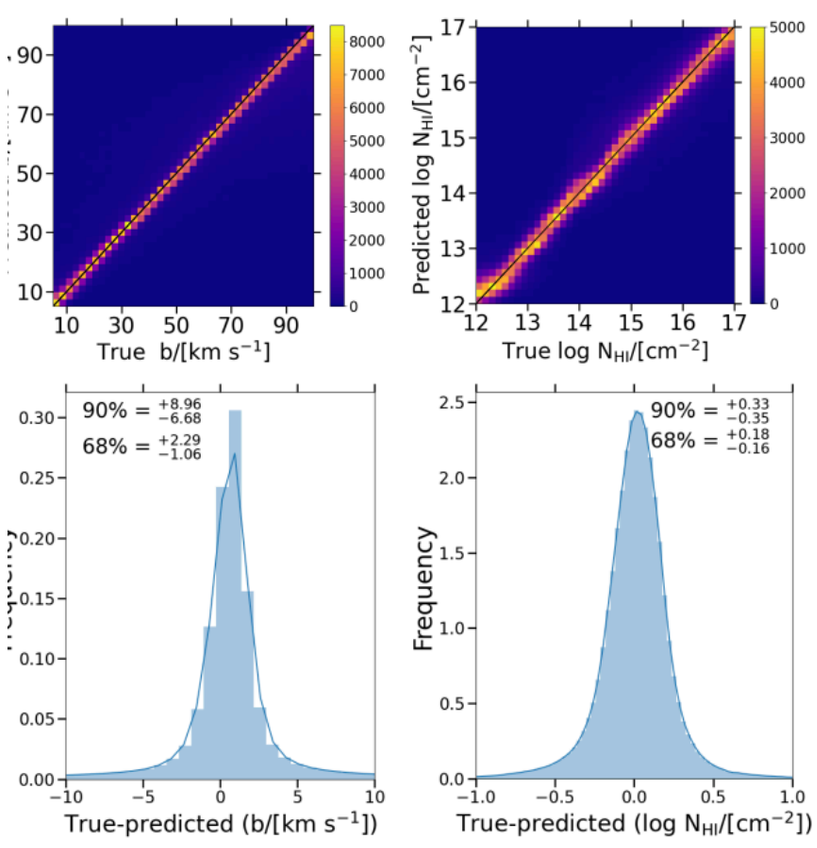
FLAME: Fitting Lyα absorption lines using machine learning

Priyanka Jalan, Vikram Khaire , M. Vivek , and Prakash Gaikwad, 202 4,A&A, 688, 126

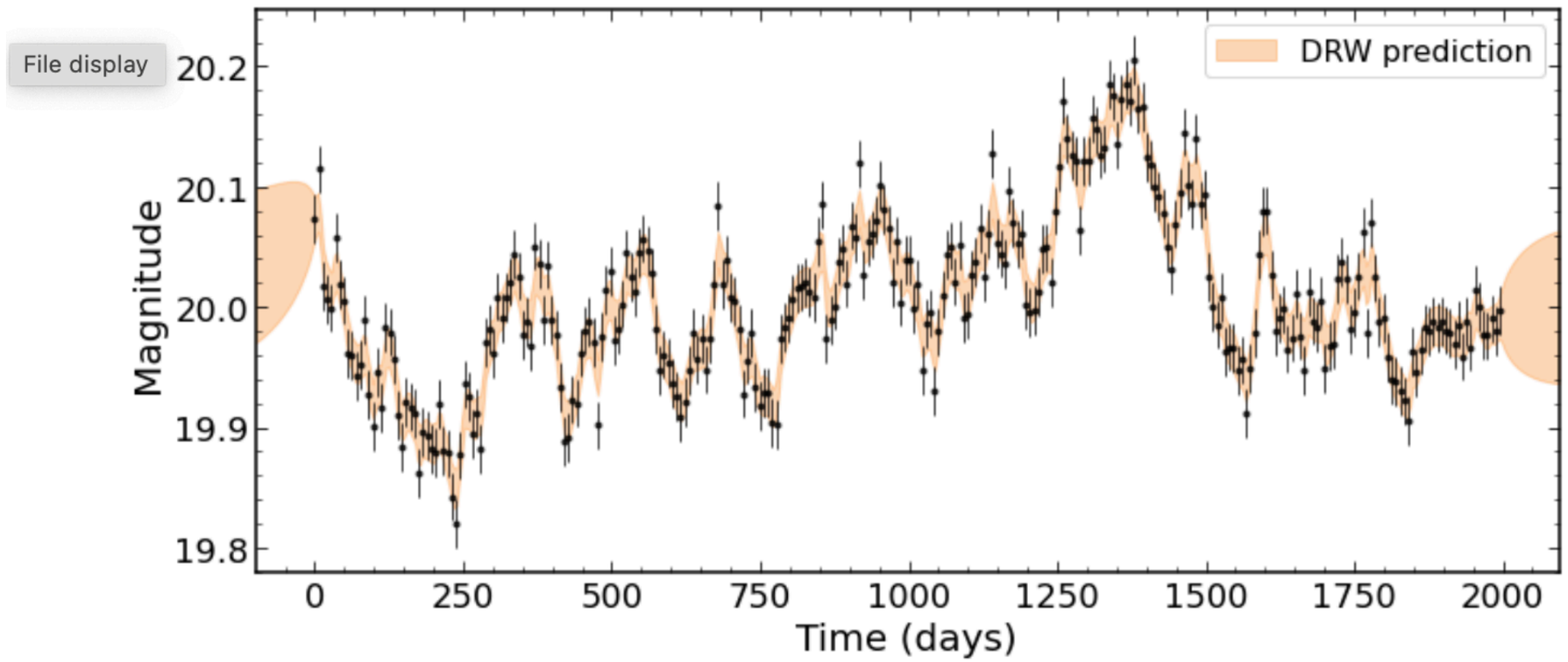


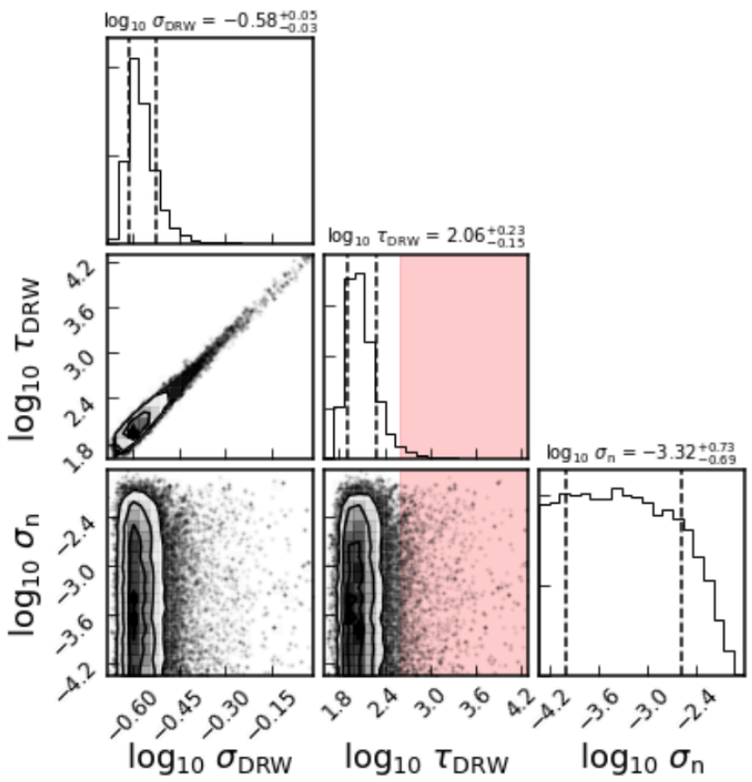


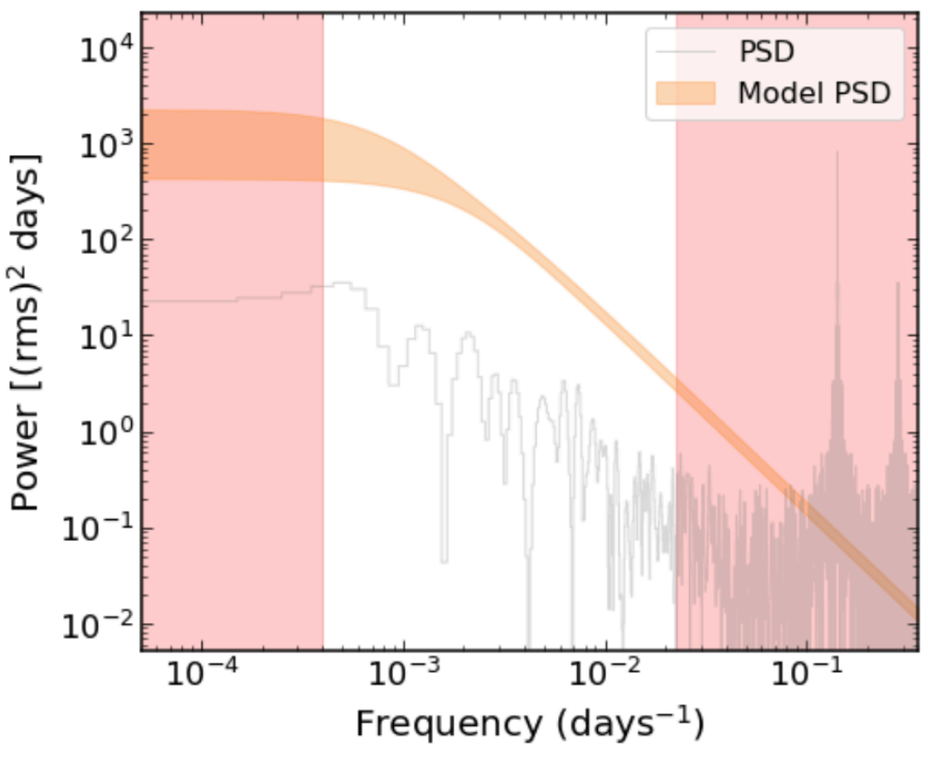




Gaussian Process Regression for damping timescales

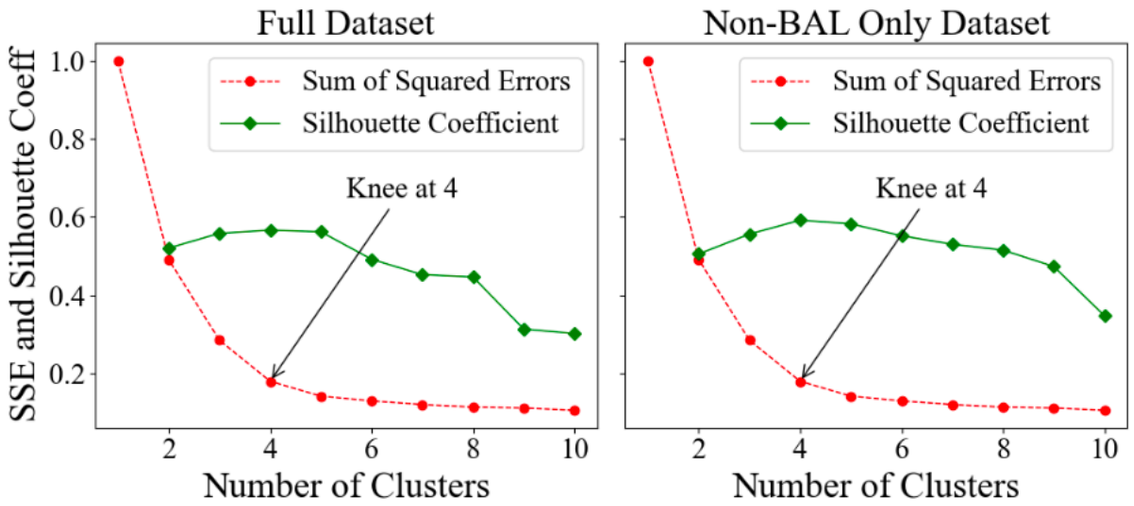


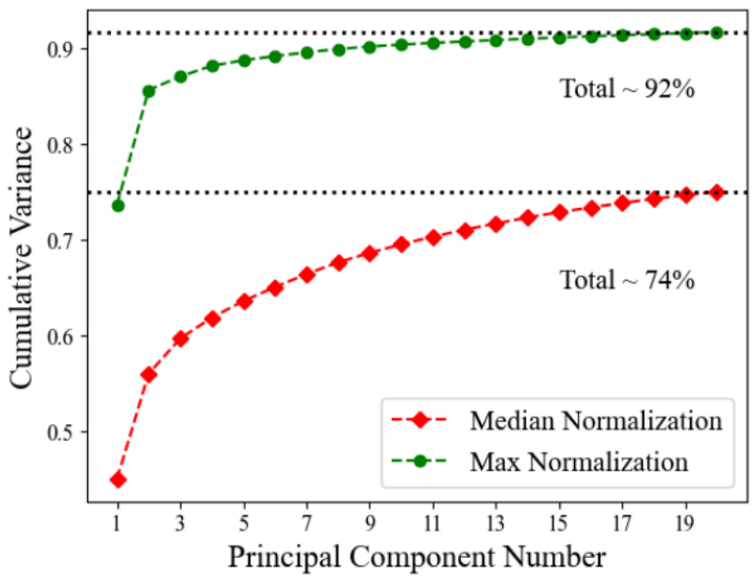




"The machine learning algorithm wants to know if we'd like a dozen wireless mice to feed the Python book we just bought."

Thank you





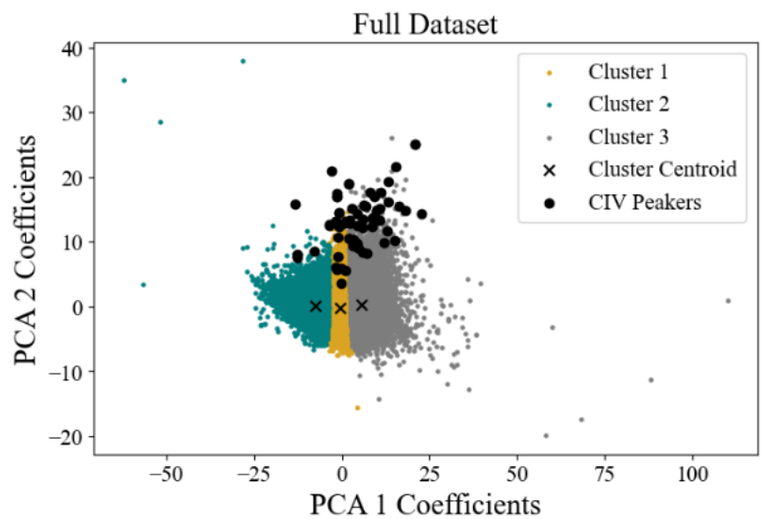
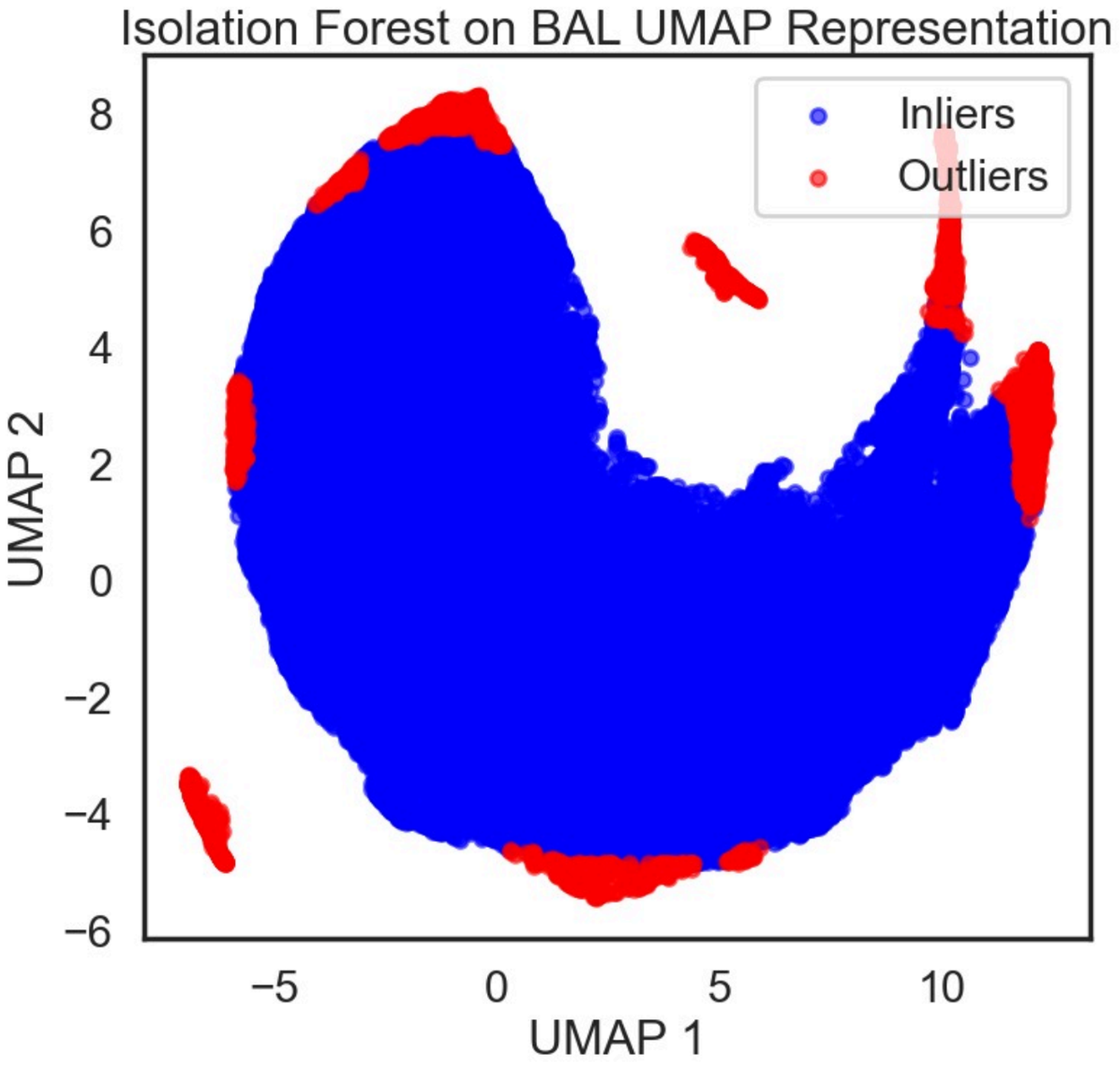
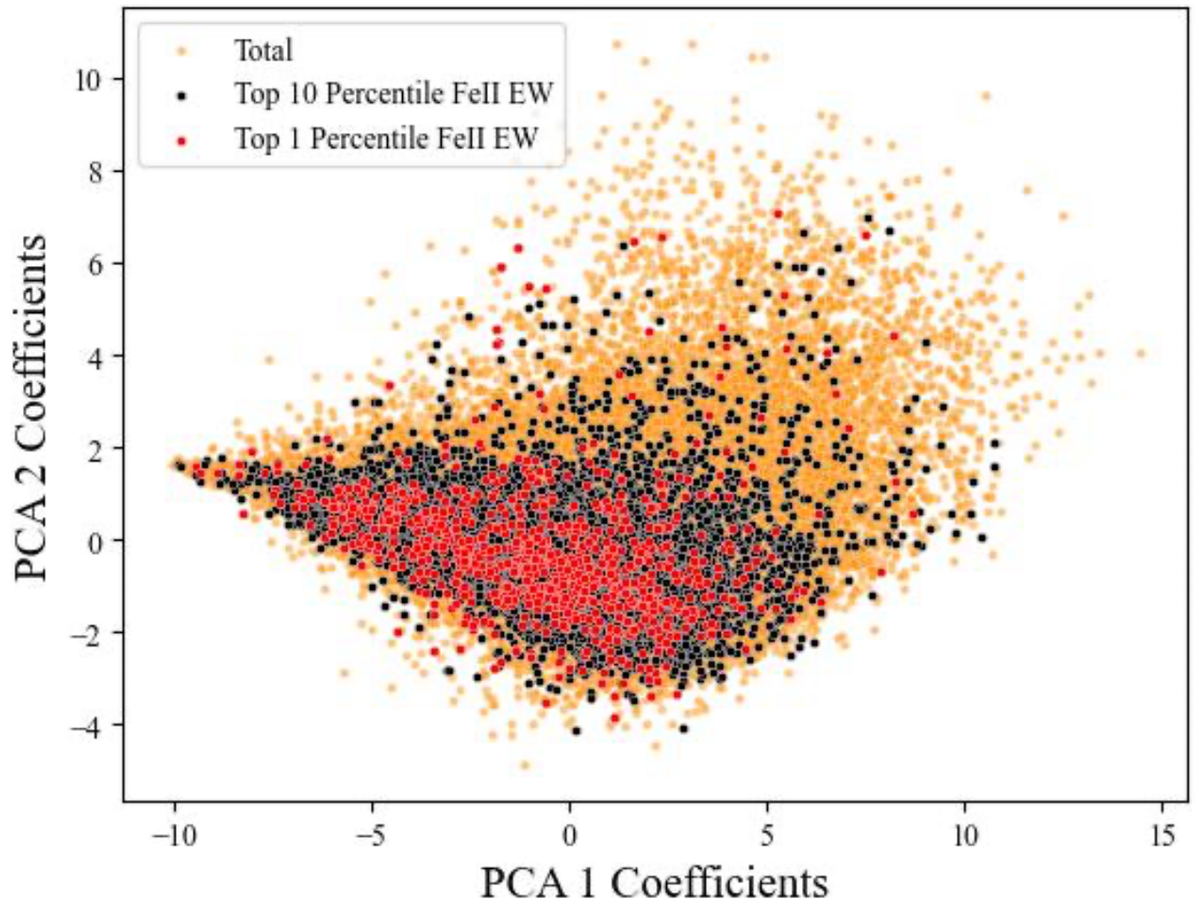


Figure 1: Left: Total explained variance by 20 PCA components for Max (current manuscript) and Median normalization (referee suggested). Right: Cluster visualization of the first two PCA components for Median normalization (referee suggested). C IV Peakers are overlayed on the spread to depict the ineffectiveness of the suggested approach.


Isolation Forest on BAL UMAP Representation
Inliers
Outliers
UMAP 2
UMAP 1
8
6
4
2
0
−2
−4
−6
−5
0
5
10


Total
Top 10 Percentile FeII EW
Top 1 Percentile FeII EW
PCA 2 Coefficients
PCA 1 Coefficients
10
8
6
4
2
0
−2
−4
−10
−5
0
5
10
15