USING MACHINE LEARNING TO IDENTIFY OUTLIERS IN THE FUNDAMENTAL METALLICITY RELATION

Speaker:
Francesco Pistis[1,2,3]
In collaboration with:
Margherita Grespan[2]
Agnieszka Pollo[2,4]

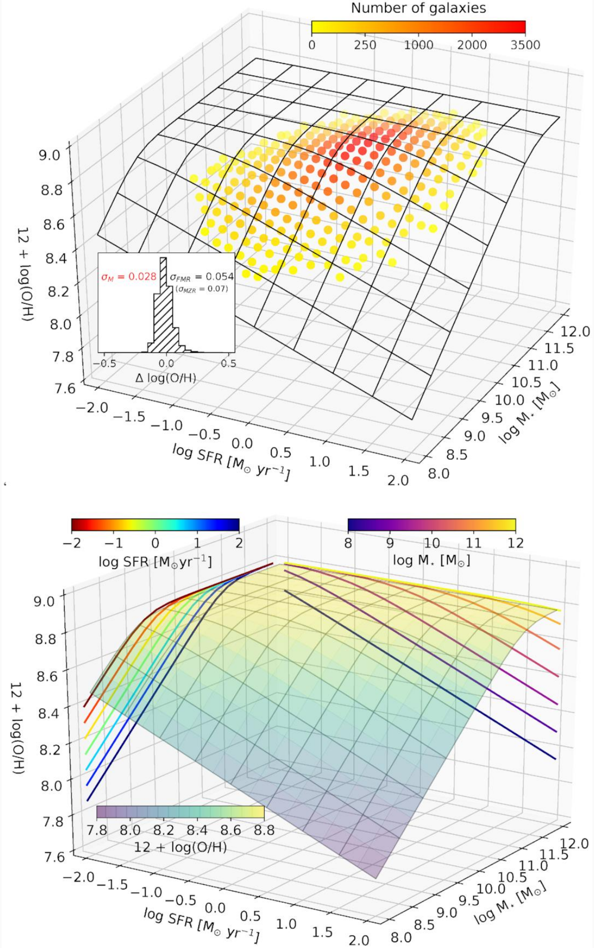
Affiliation(s):
[1] Università Milano-Bicocca
[2] National Centre for Nuclear Research
[3] INAF-OAS
[4] Jagiellonian University

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
Funded by the European Union
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Introduction

The fundamental metallicity relation

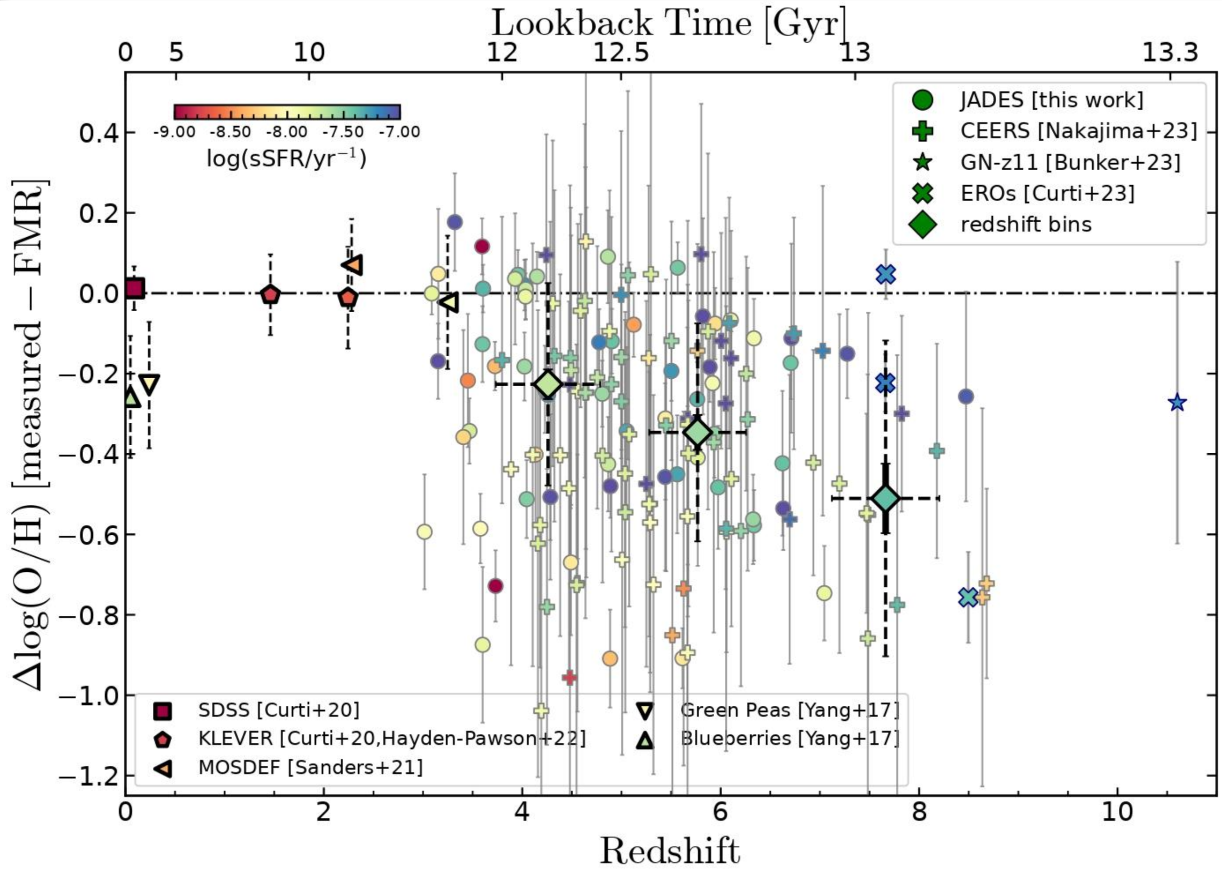
- Defined for star-forming galaxies
- Relation between stellar mass, SFR, and gas-phase metallicity of the ISM
- Galaxies lies on a well-defined surface
- Shaped by the inflows and outflows of gas



Mon Not R Astron Soc, Volume 491, Issue 1, January 2020, Pages 944–964,
https://doi.org/10.1093/mnras/stz2910

Evolution of the fundamental metallicity relation

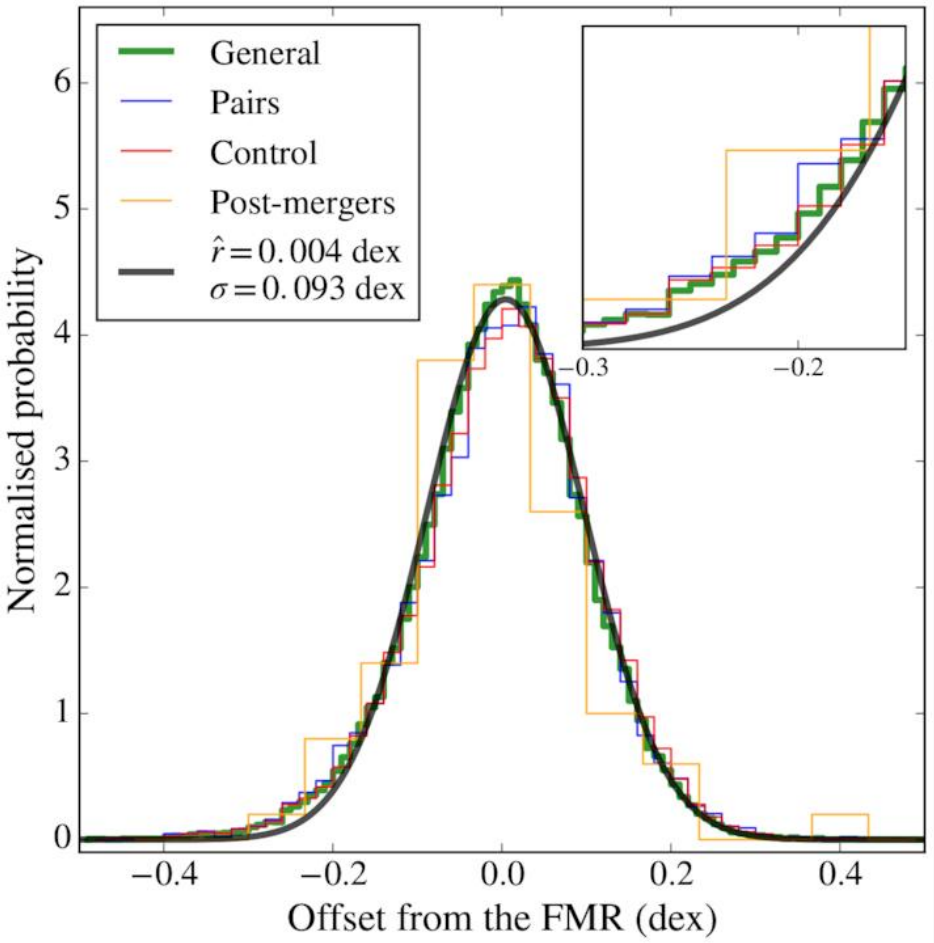
- No evolution observed up to z~2.5 — fundamental relation
- The metallicity decreases with redshift



A&A, 684, A75 (2024), https://doi.org/10.1051/0004-6361/202346698

Pairs/mergers on the fundamental metallicity relation

- Galaxies interact also between each other and not only with the surrounding gas
- Where do interacting galaxies take place over the surface?
- Excess of residuals observed on the negative tail

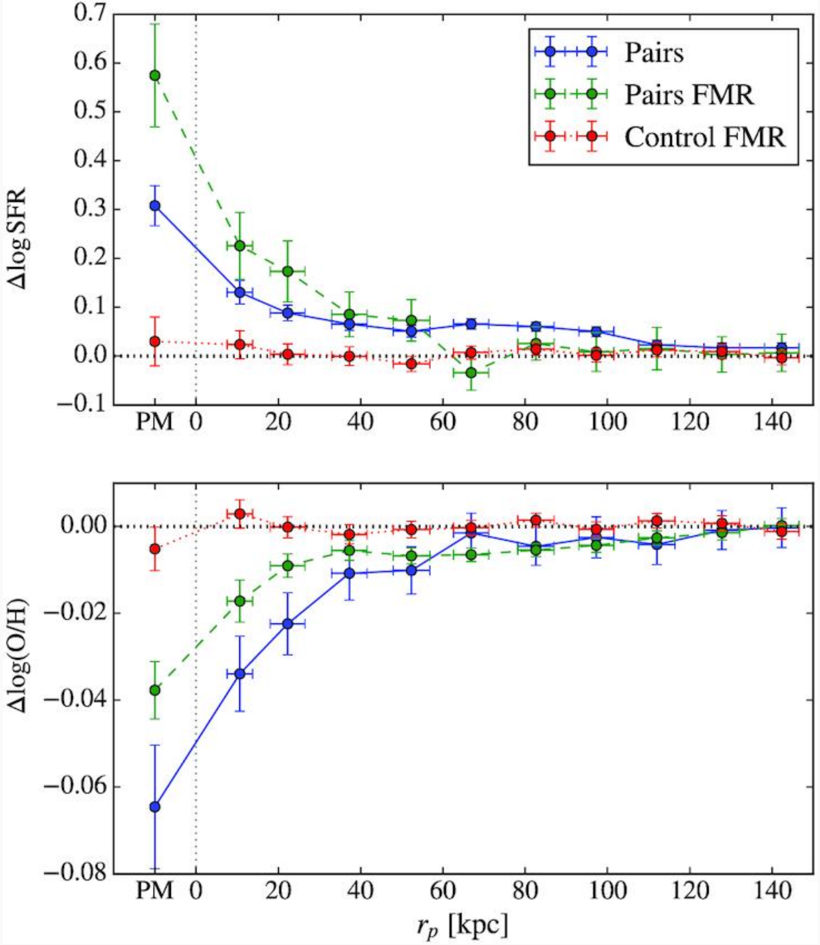


Mon Not R Astron Soc, Volume 494, Issue 3, May 2020, Pages 3469–3480, https://doi.org/10.1093/mnras/staa1025

Pairs/mergers on the fundamental metallicity relation

- Interacting galaxies have enhanced star formation rate and reduced metallicity
- Pairs/mergers can explain the excess of residuals from the fundamental metallicity relation

- Can we detect mergers/pairs as outliers in the fundamental metallicity relation?

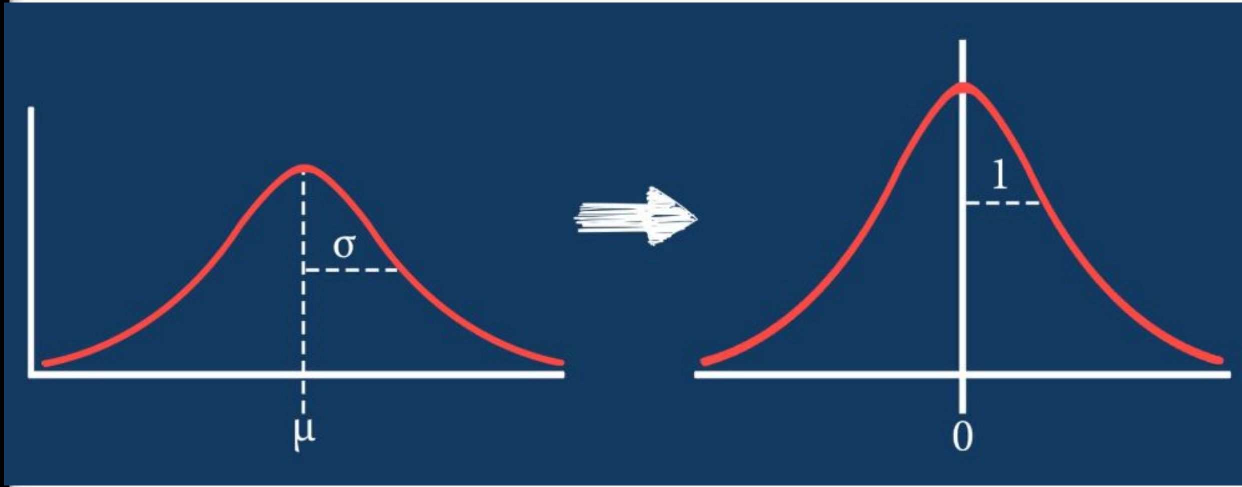


Mon Not R Astron Soc, Volume 494, Issue 3, May 2020, Pages 3469–3480, https://doi.org/10.1093/mnras/staa1025

Methodology

Preprocessing

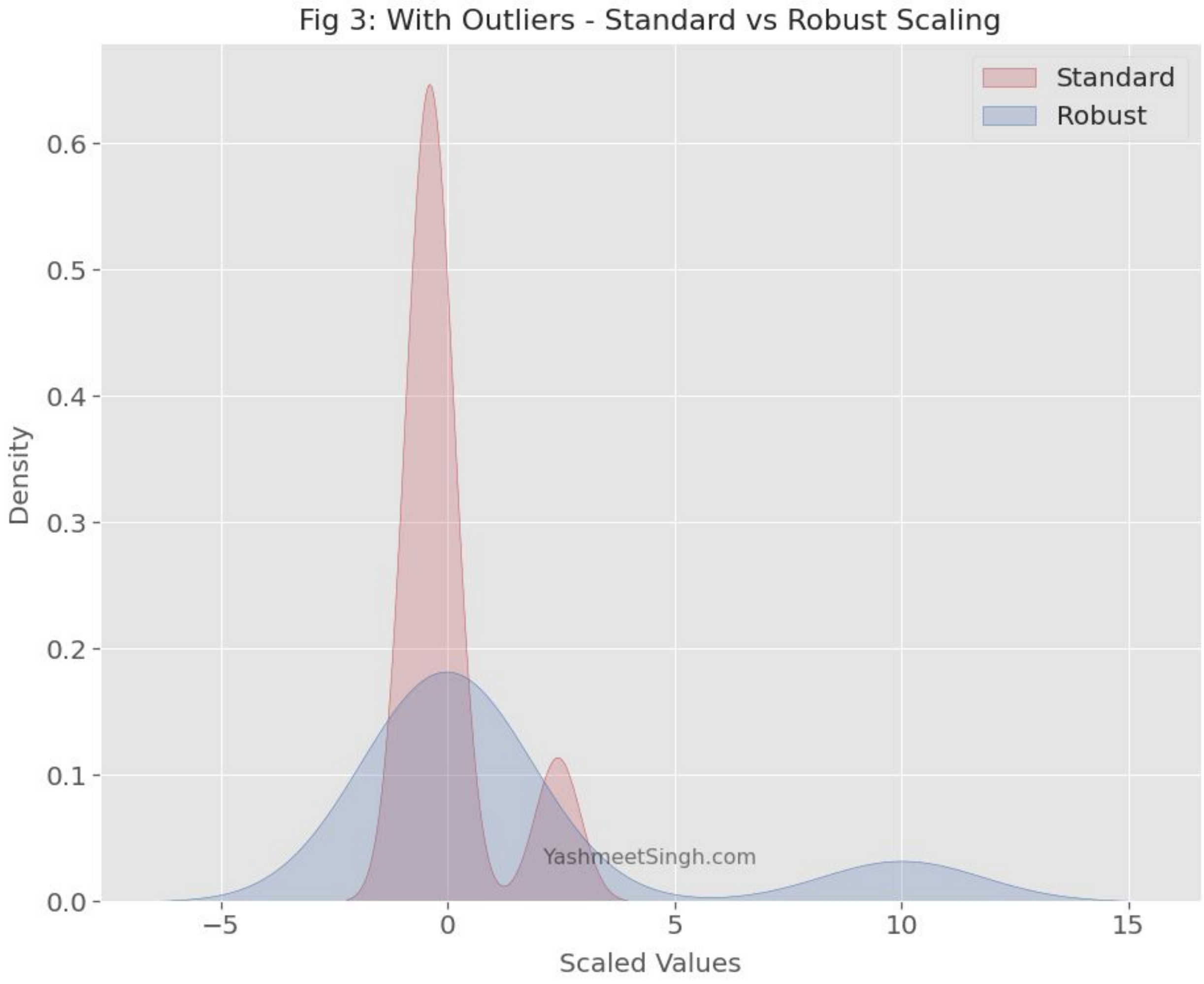
- Important to reduce the impact of different order of magnitude between features
- Standard scaling: center shifted at zero and unit σ
  - Sensitive to outliers



Credits: *[Normalization vs Standardization — Quantitative analysis]*

Preprocessing

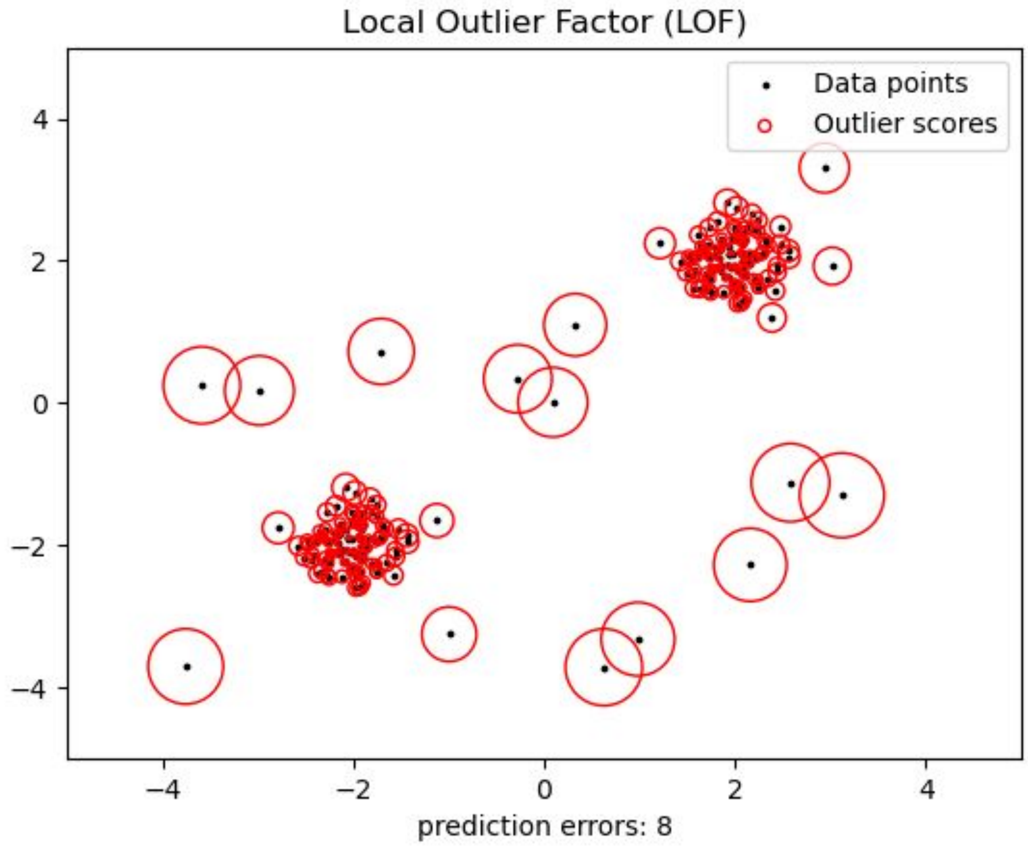
- Important to reduce the impact of different order of magnitude between features
- Standard scaling: center shifted at zero and unit σ
  - Sensitive to outliers
- Robust scaling: calculate median and standard deviation within the interquartile range (25th and 75th)
  - Reduced impact of outliers



Credits: *[Robust Scaling: Why and How to Use It to Handle Outliers]*

Local outlier factor

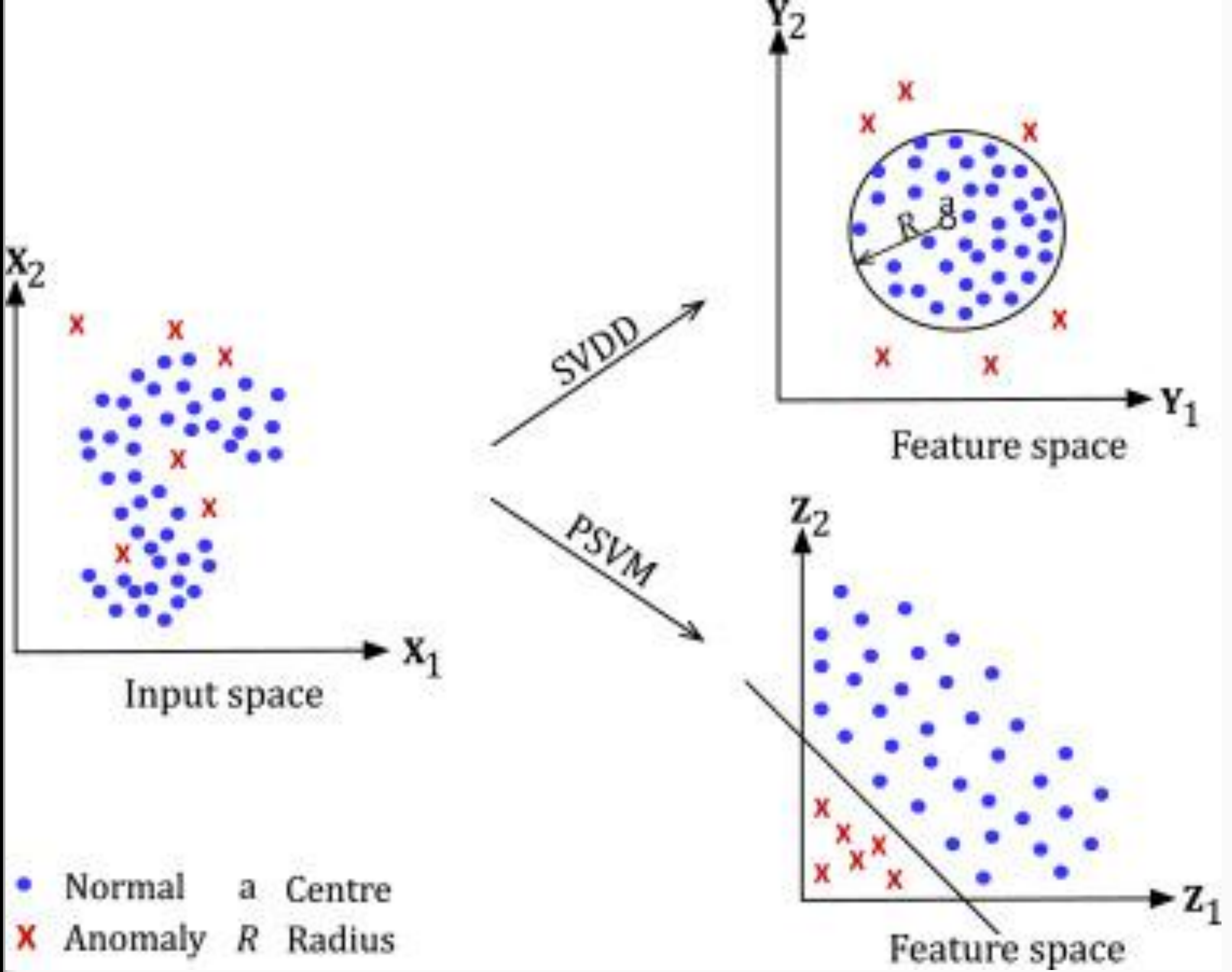
- Method 1: local outlier factor
  - Assign to each point a score according the local density deviation with respect to its neighbors



Credits: *[Outlier detection with Local Outlier Factor (LOF) --- Sklearn]*

One-class support vector machine

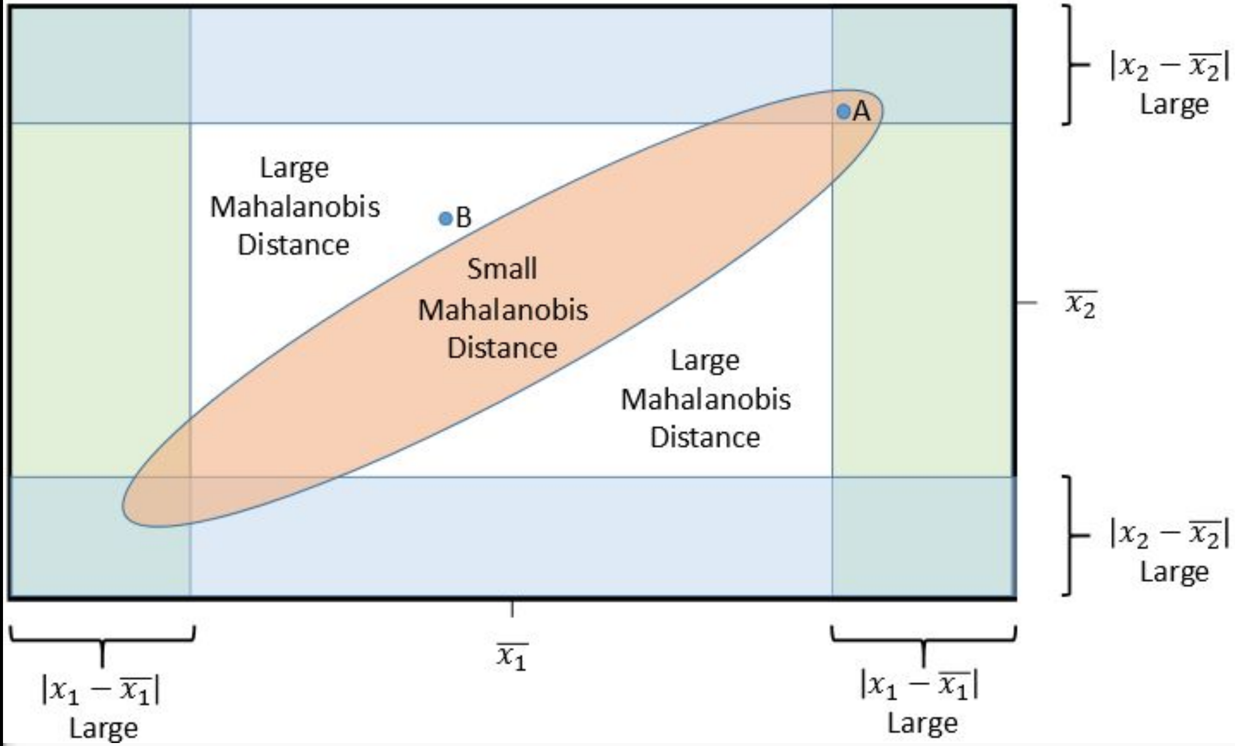
- Method 1: local outlier factor
  - Assign to each point a score according the local density deviation with respect to its neighbors
- Method 2: one-class support vector machine
  - Use of a kernel to estimate similarity of points and define a decision boundary in a higher dimensional space
- Condition 1: outliers detected by both methods



Credits: [High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning]

Mahalanobis distance

- Condition 1: outliers detected by both methods
- Condition 2: Mahalanobis distance higher than 99th percentile



Credits: [The geometry of multivariate versus univariate outliers]()

Input parameter space

Run 1

- Stellar mass
- Star formation rate
- Gas-phase metallicity

Run 2

- Stellar mass
- Star formation rate
- Gas-phase metallicity
- Scale factor $a(t)=(z+1)^{-1}$ (Expected to not have effects in the range explored by the data)

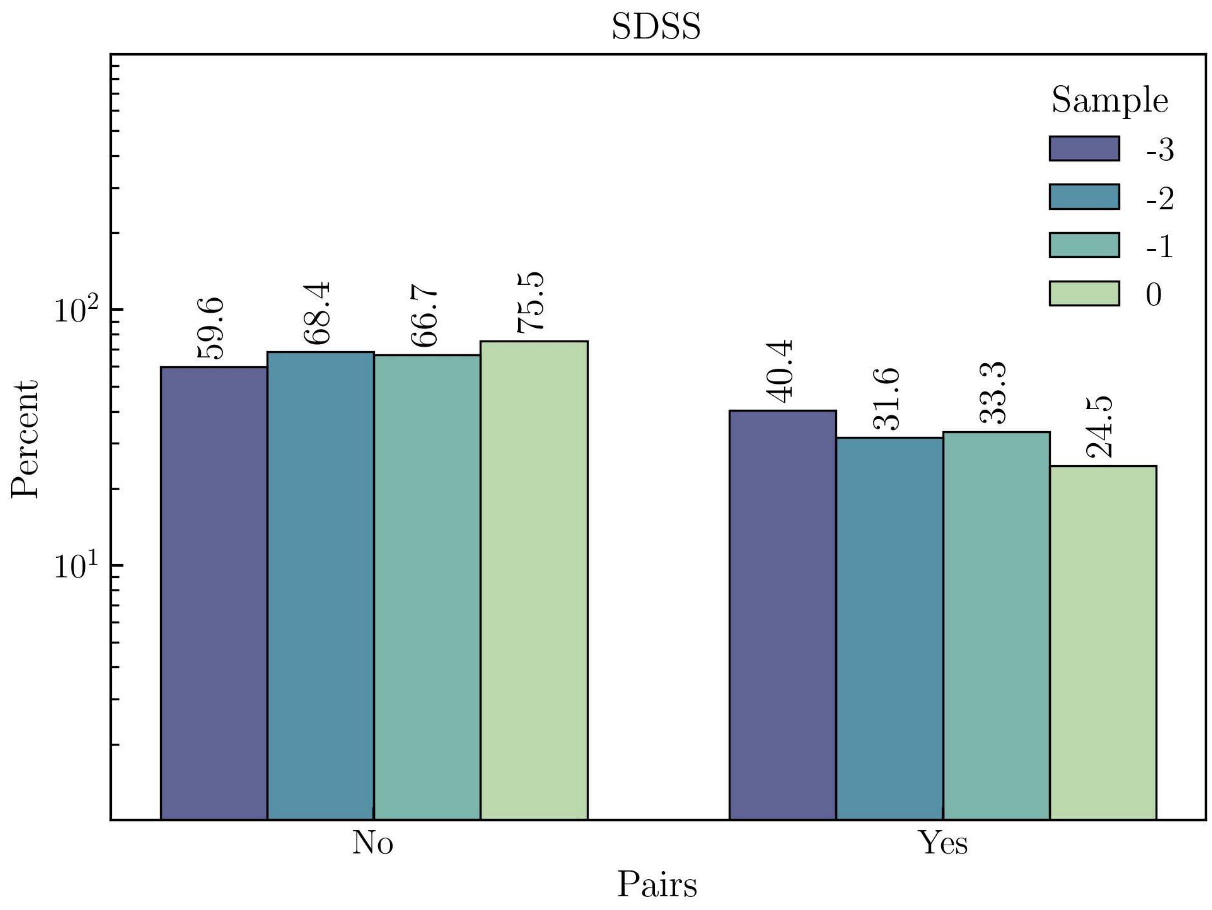
We divide the outliers in common sources between the 2 runs, and outliers detected only by one run.

Results

Galaxy pairs

- Label 0: normal data
- Label -1: common outliers between runs
- Label -2: outliers detected only on run 1 (stellar mass, star formation rate, and metallicity)
- Label -3: outliers detected only on run 2 (including scale factor)

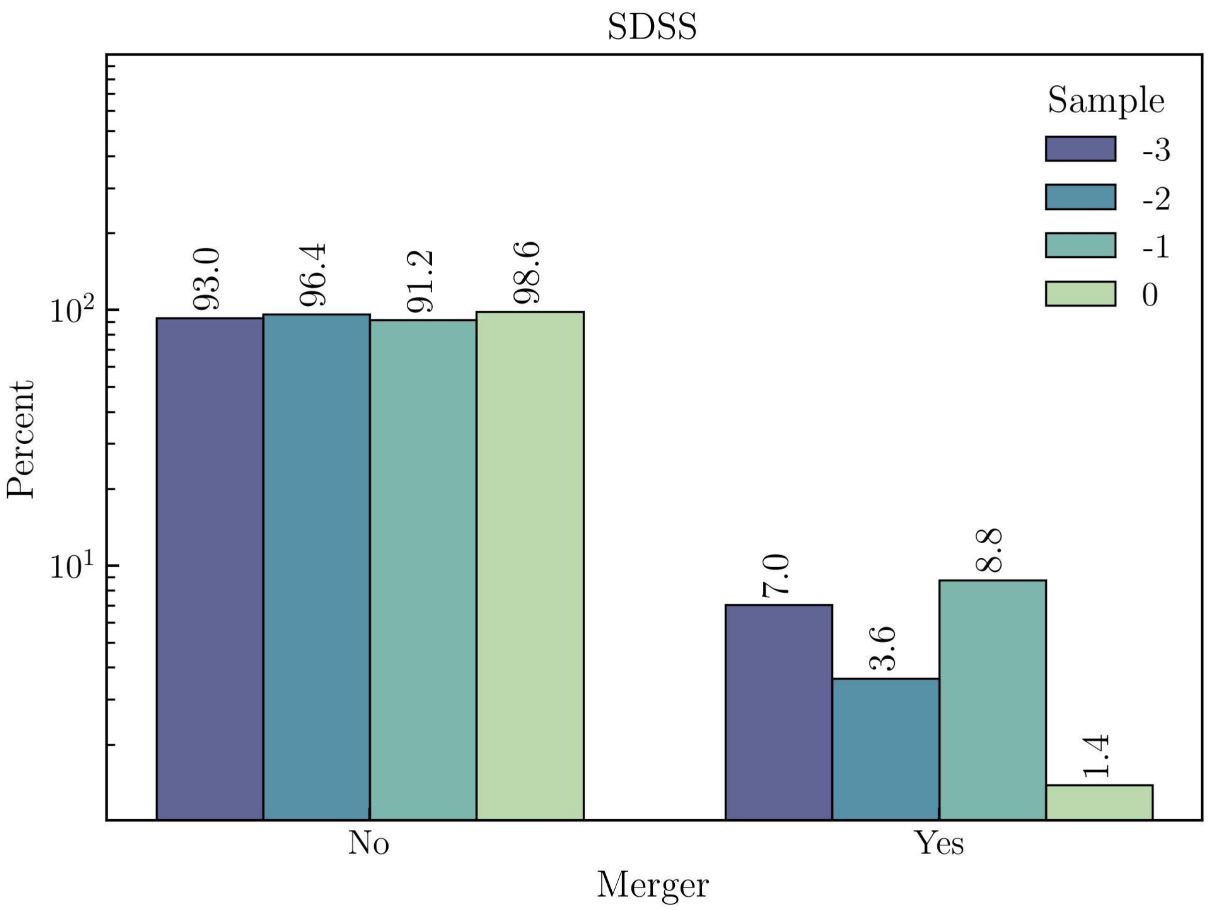
- Outliers have a slightly higher fraction of pair fraction
- Including scale factor shows the max pairs fraction



Galaxy mergers

- Label 0: normal data
- Label -1: common outliers between runs
- Label -2: outliers detected only on run 1 (stellar mass, star formation rate, and metallicity)
- Label -3: outliers detected only on run 2 (including scale factor)

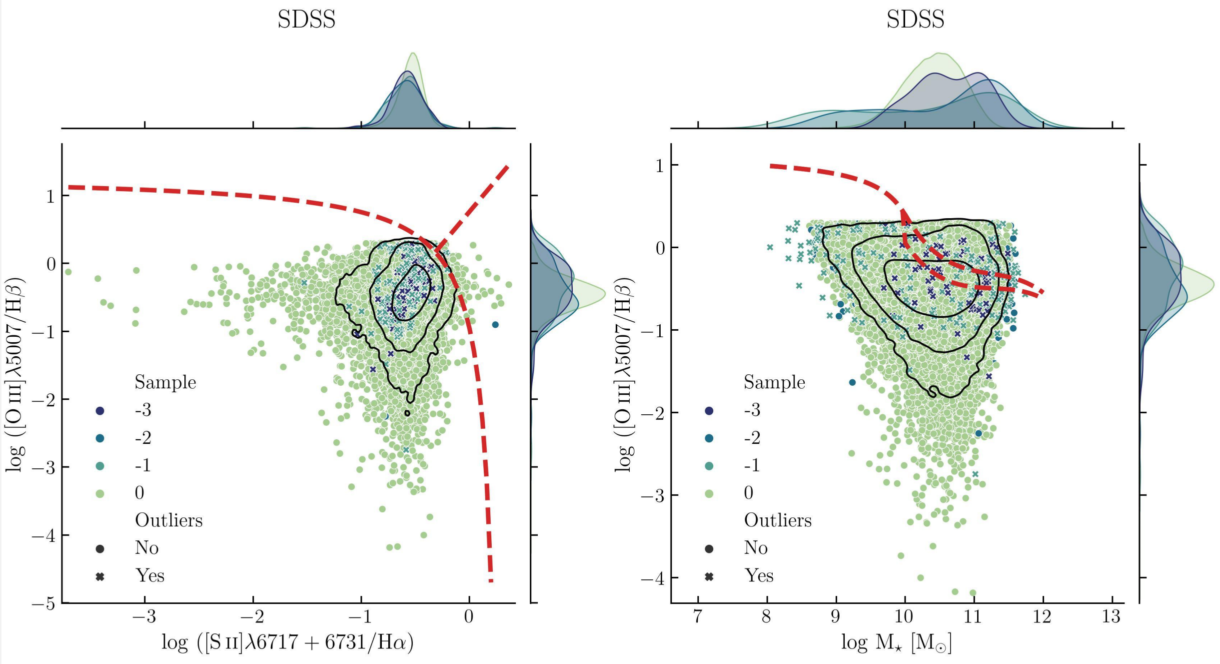
- Outliers have a fraction of merger between 2 and 6 times the normal data



Contamination from Active Galactic Nuclei

- Label 0: normal data
- Label -1: common outliers between runs
- Label -2: outliers detected only on run 1 (stellar mass, star formation rate, and metallicity)
- Label -3: outliers detected only on run 2 (including scale factor)

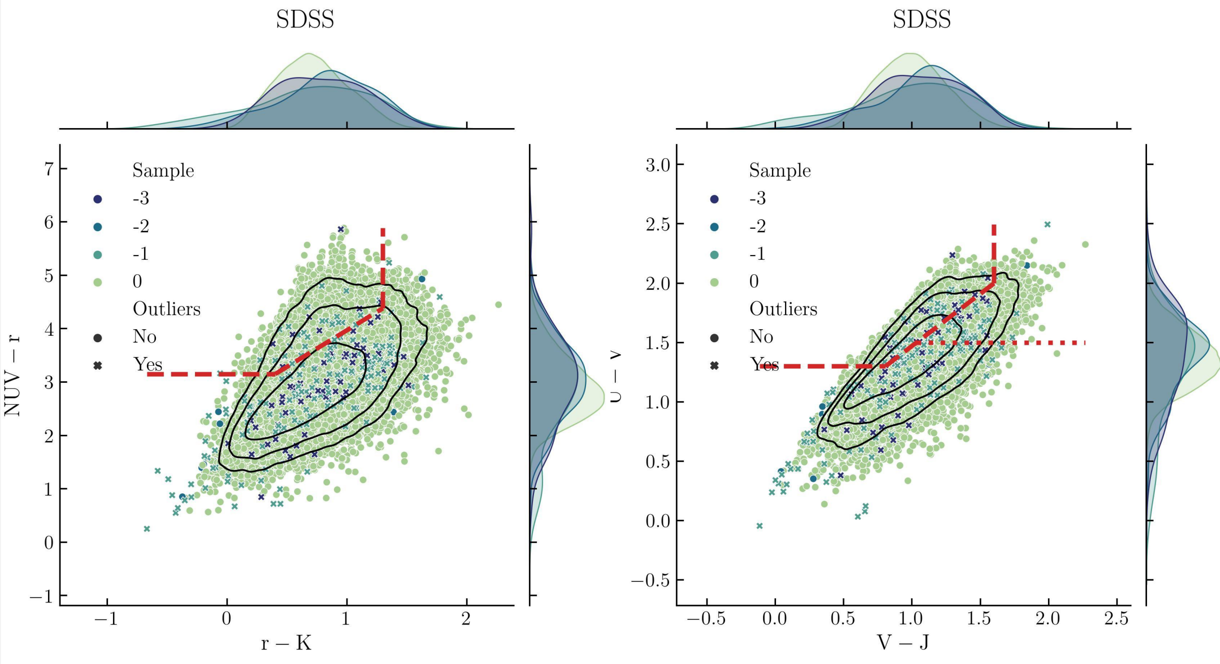
- Outliers are not well-separated in the BPT diagram…
- …or MEx diagram



Contamination from passive galaxies

- Label 0: normal data
- Label -1: common outliers between runs
- Label -2: outliers detected only on run 1 (stellar mass, star formation rate, and metallicity)
- Label -3: outliers detected only on run 2 (including scale factor)

- Outliers are not well-separated in the color-color diagrams to separate active and passive galaxies

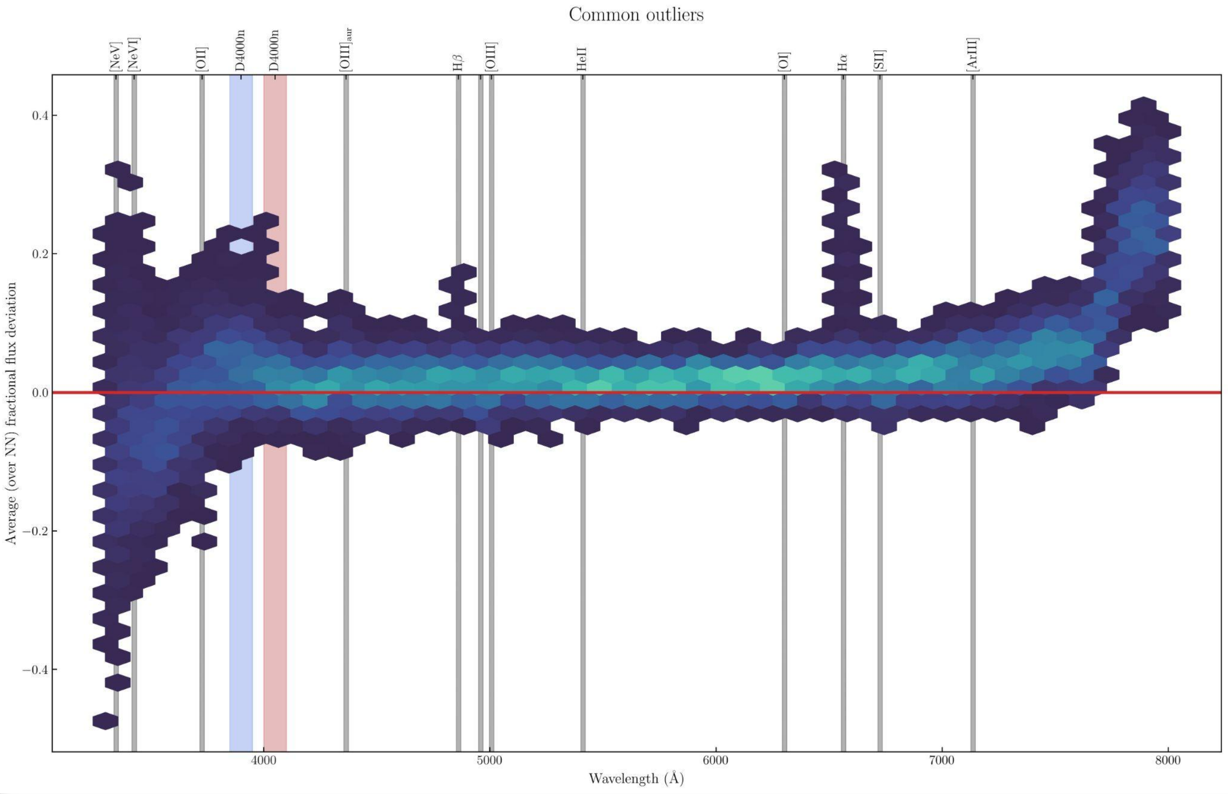


Common outliers

- Average (over 3 nearest neighbors in the stellar mass-star formation rate plane) fractional flux deviation

$$\left\langle \frac{F_{out} - F_{in}}{F_{in}} \right\rangle_{3NN}$$

- Transition around D4000n break (passing from less to more brighter than nearest neighbors)
- Large dispersion of the average deviation around lines
- Largest dispersion around [NeV] line, AGN tracer

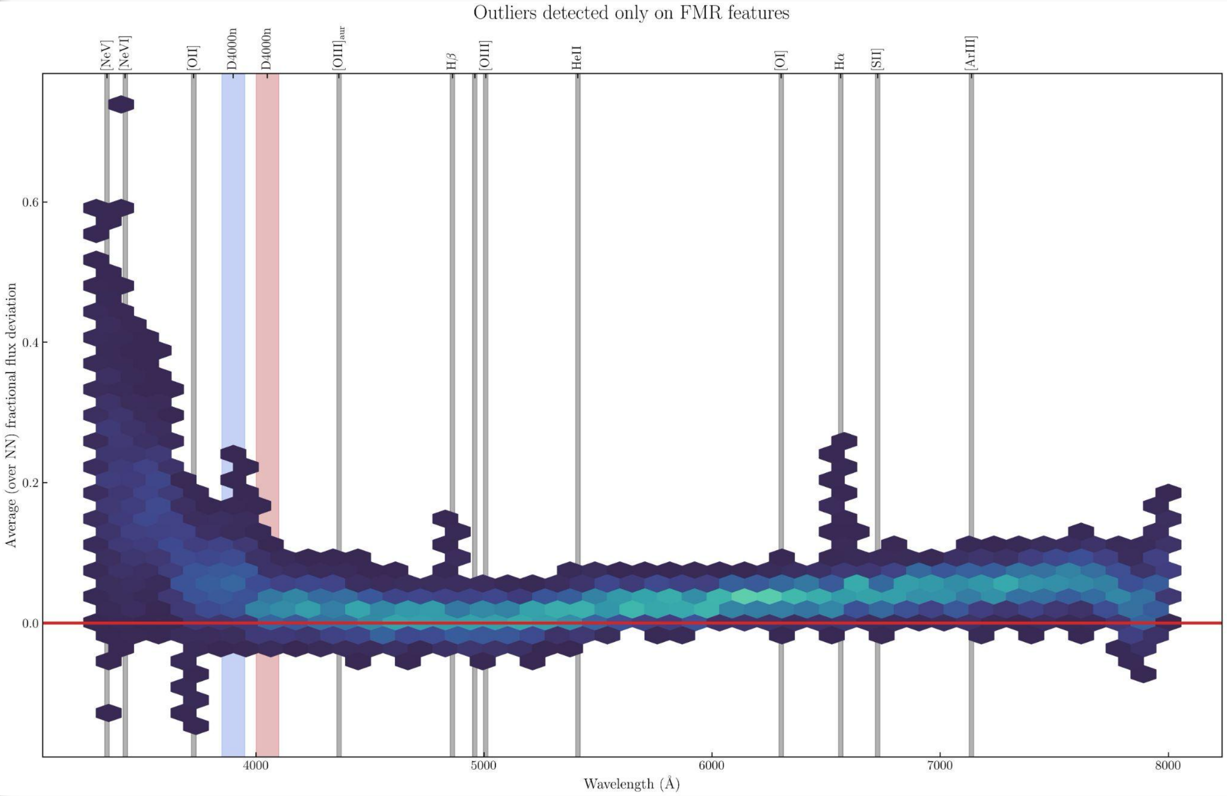


Outliers from FMR

- Average (over 3 nearest neighbors in the stellar mass-star formation rate plane) fractional flux deviation

$$\left\langle \frac{F_{out} - F_{in}}{F_{in}} \right\rangle_{3NN}$$

- Outliers having brighter spectra then nearest neighbors
- Large dispersion of the average deviation around lines
- Largest dispersion around [NeV] line, AGN tracer

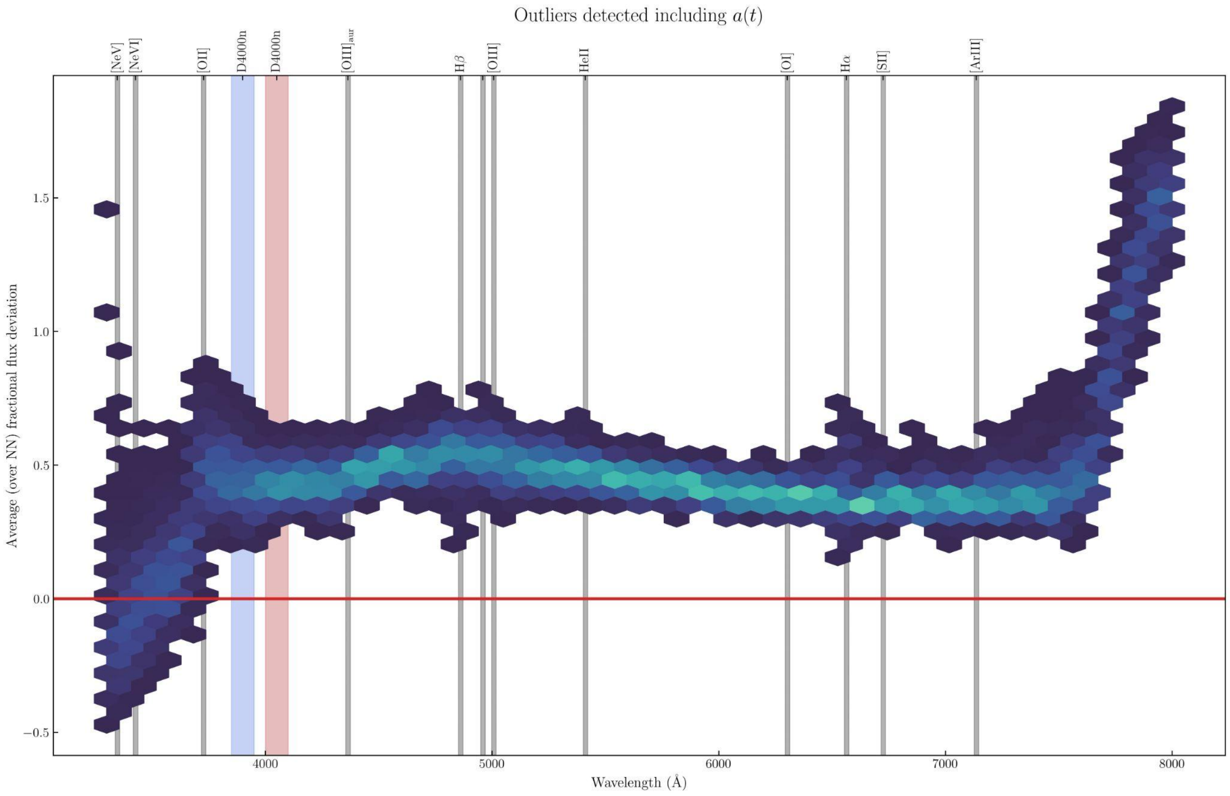


Outliers including a(z)

- Average (over 3 nearest neighbors in the stellar mass-star formation rate plane) fractional flux deviation

$$\left\langle \frac{F_{out} - F_{in}}{F_{in}} \right\rangle_{3NN}$$

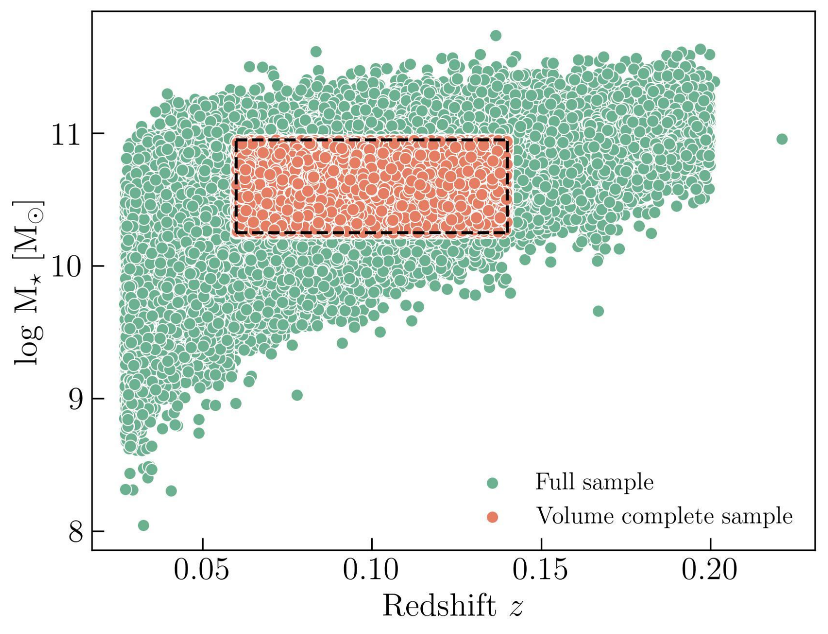
- Outliers brighter than nearest neighbors
- Larger dispersion of the average deviation around emission lines
- Largest dispersion around [NeV] line, AGN tracer



Conclusions

Conclusions

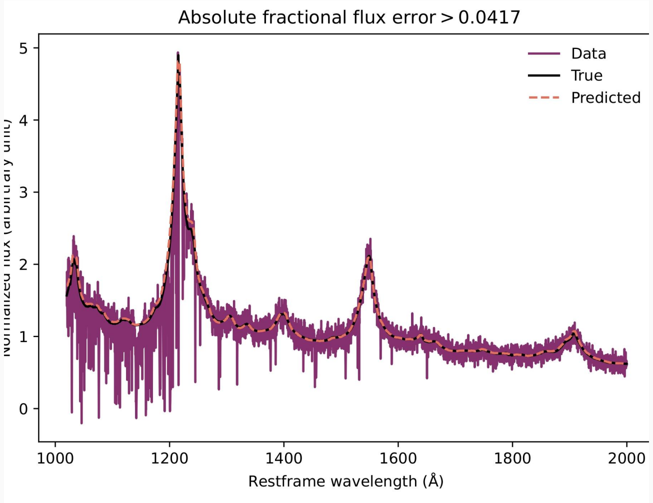
- Outliers have between 2 and 6 times higher merger fraction than normal sample
- Including the scale factor a(z) slightly increases the fraction of pairs galaxies
- Outliers are not well separated in BPT or color-color diagrams
- Outliers have on average higher fluxes than the 3 nearest neighbors, with the main emission lines having larger dispersion
- Why a(z) is so important? Survey (luminosity-limited) bias? Deeper physical explanation? WORK IN PROGRESS!



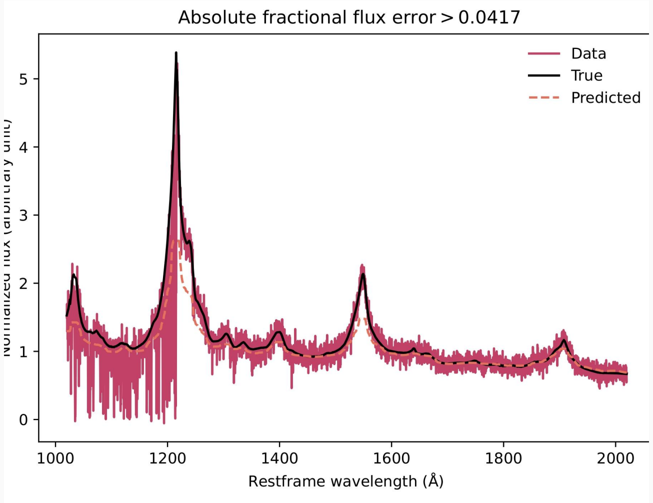
New project

New project: automated quasar and galaxy continuum estimation using neural networks

Autoencoder

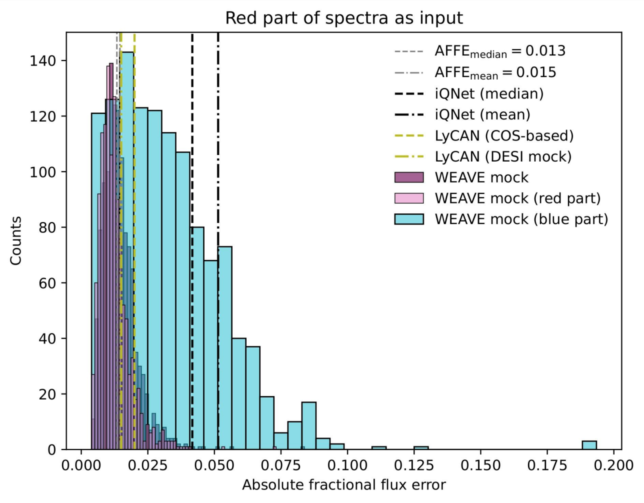


U-Net

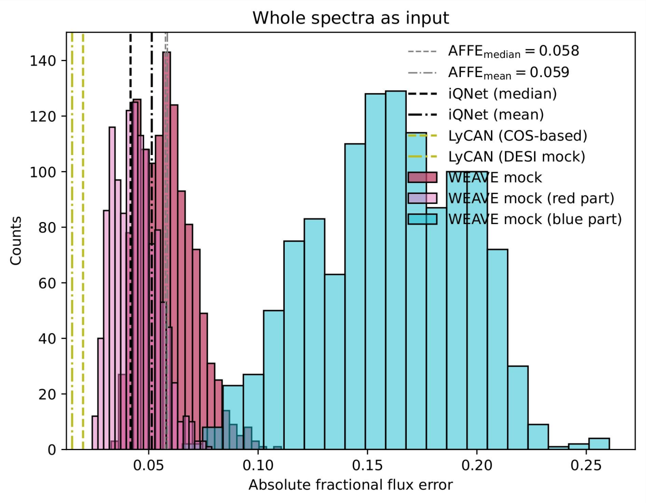


New project: automated quasar and galaxy continuum estimation using neural networks

Autoencoder



U-Net



Thank you for your attention!