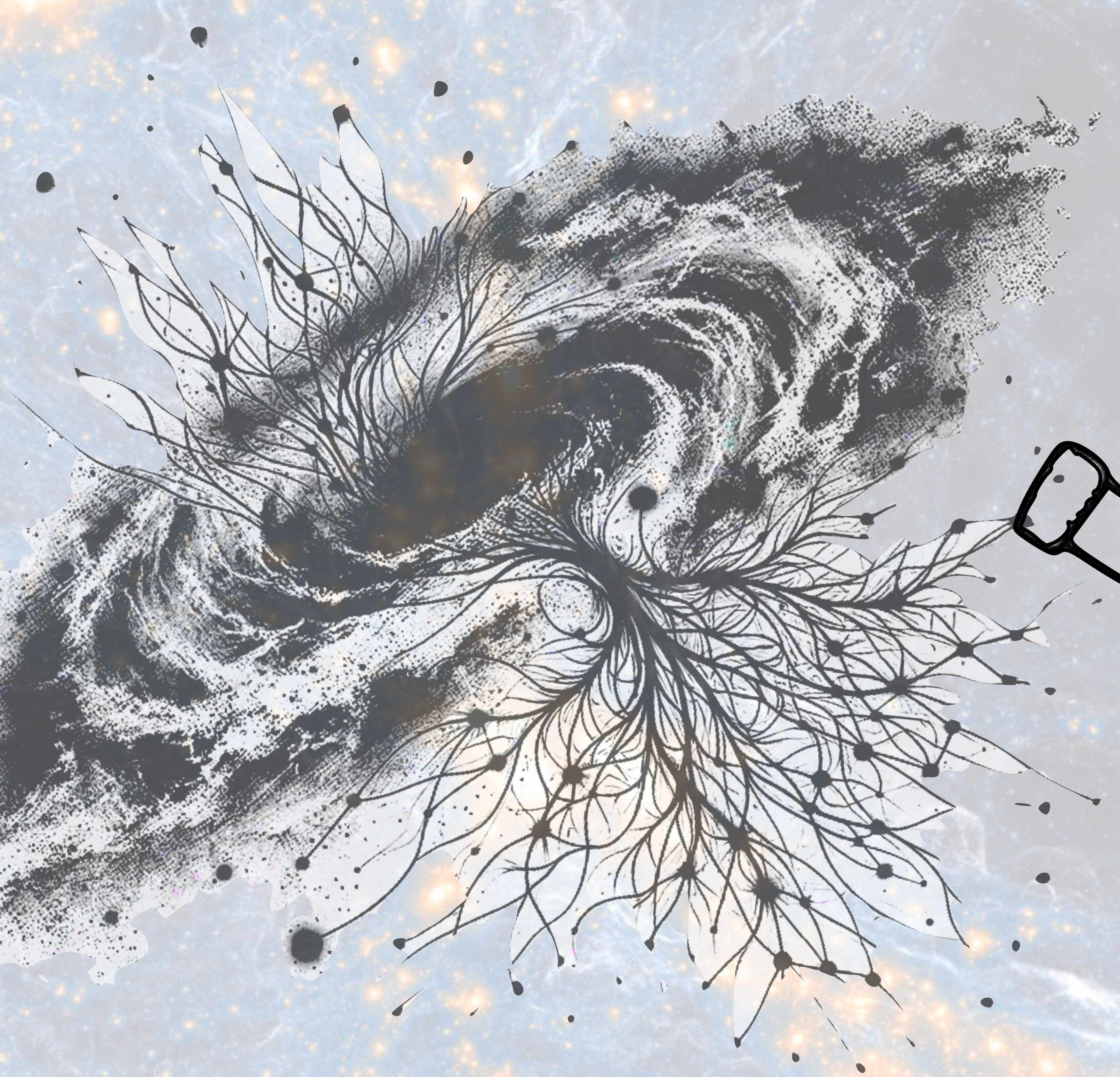
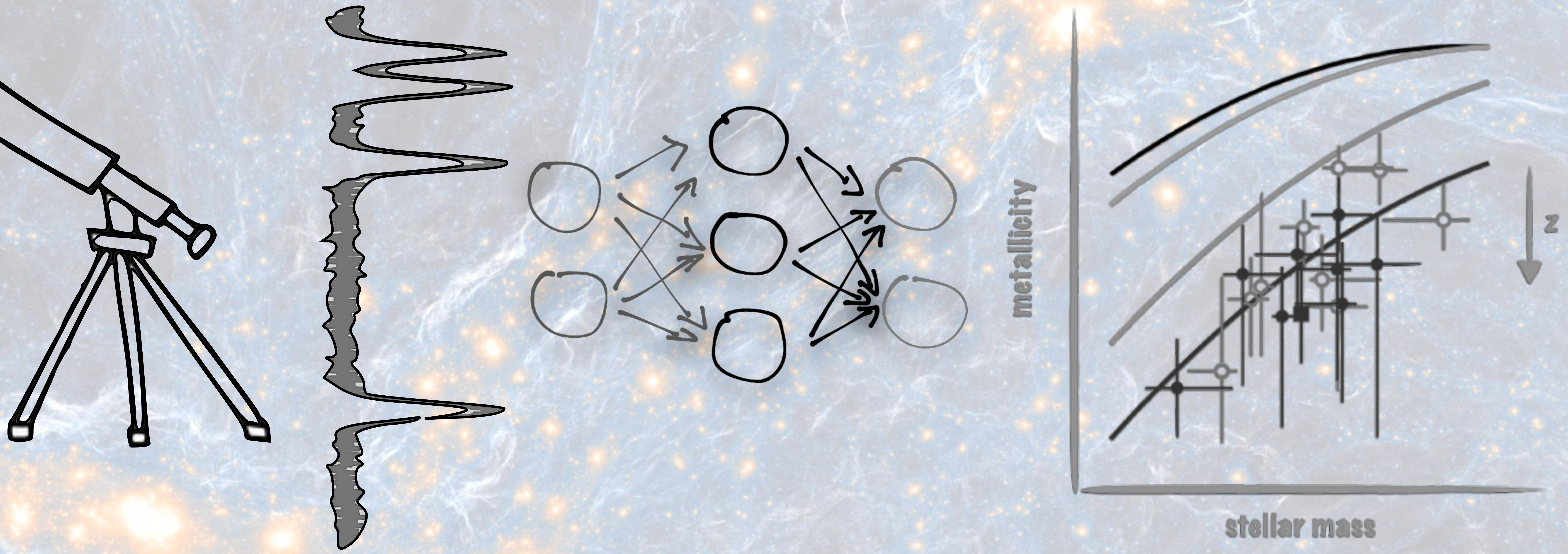
# A Multi-Task Neural Net with Monte Carlo Dropout for Spectral Analysis of Galaxies


Michele Ginolfi - UniFi

+ F. Mannucci, A. Marconi, F. Belfiore, G. Cresci et al.





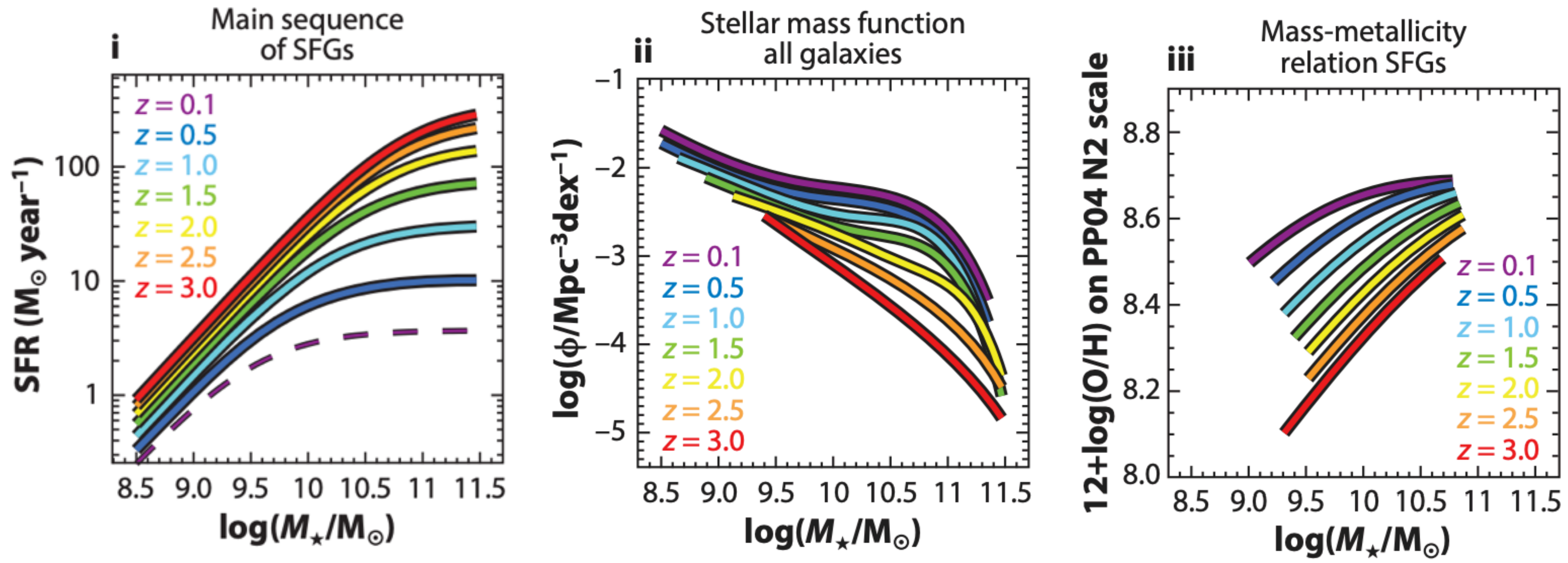
UNIVERSITÀ DEGLI STUDI FIRENZE

INAF ISTITUTO NAZIONALE DI ASTROFISICA

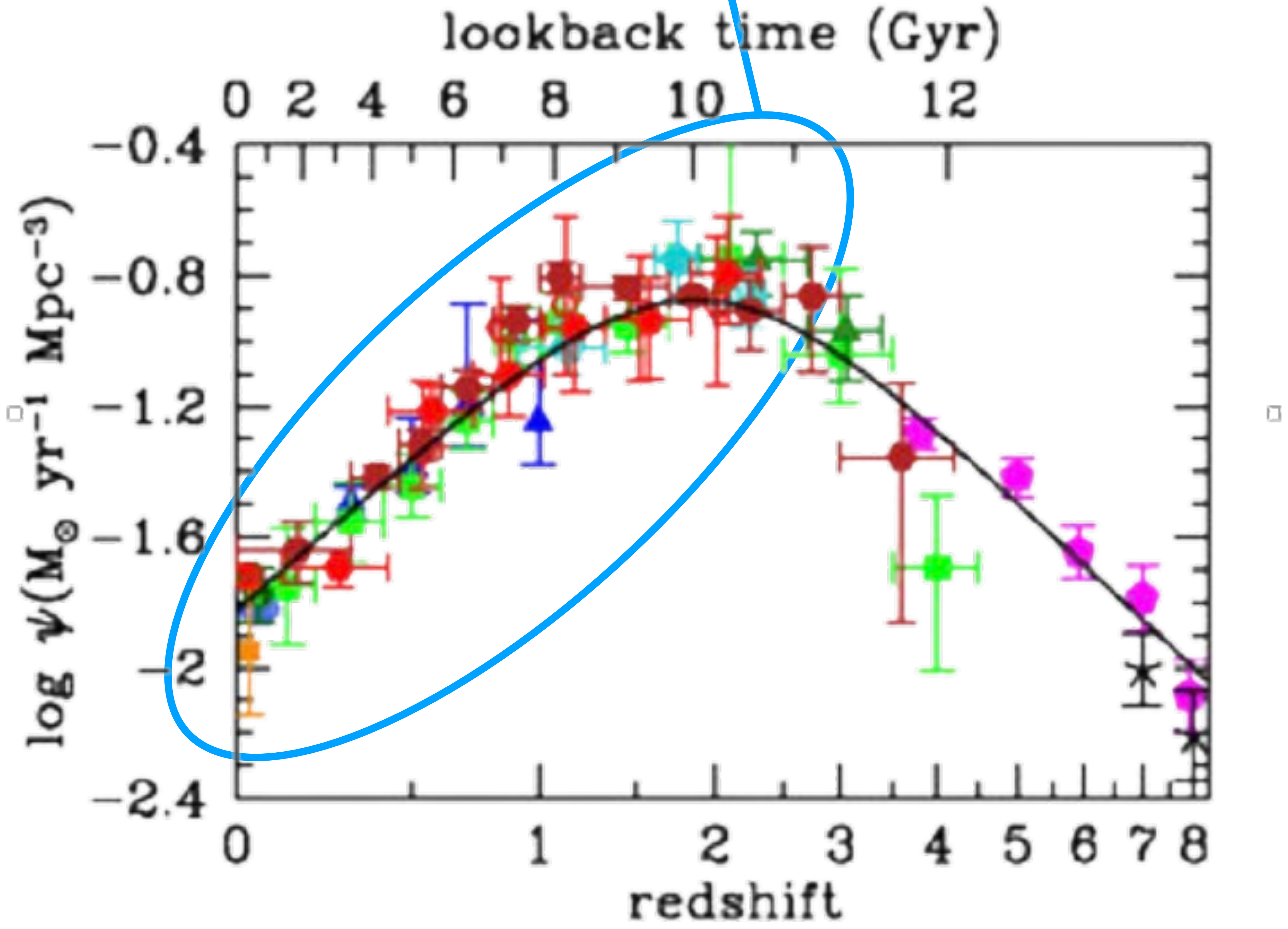

MACHINE LEARNING FOR ASTROPHYSICS

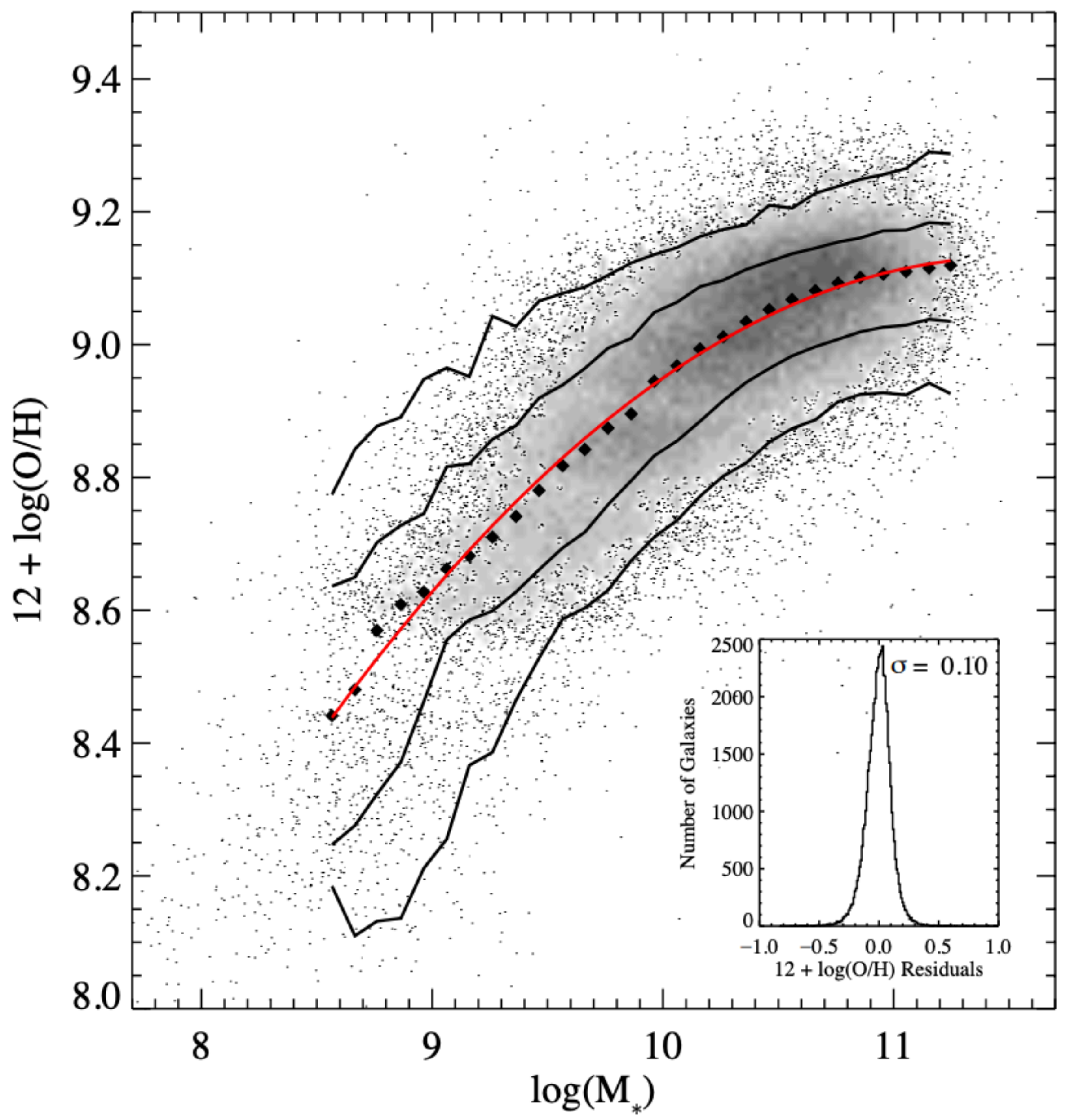
2ND EDITION

CATANIA, 8-12 JULY, 2024




Madau+14; Förster Schreiber+20



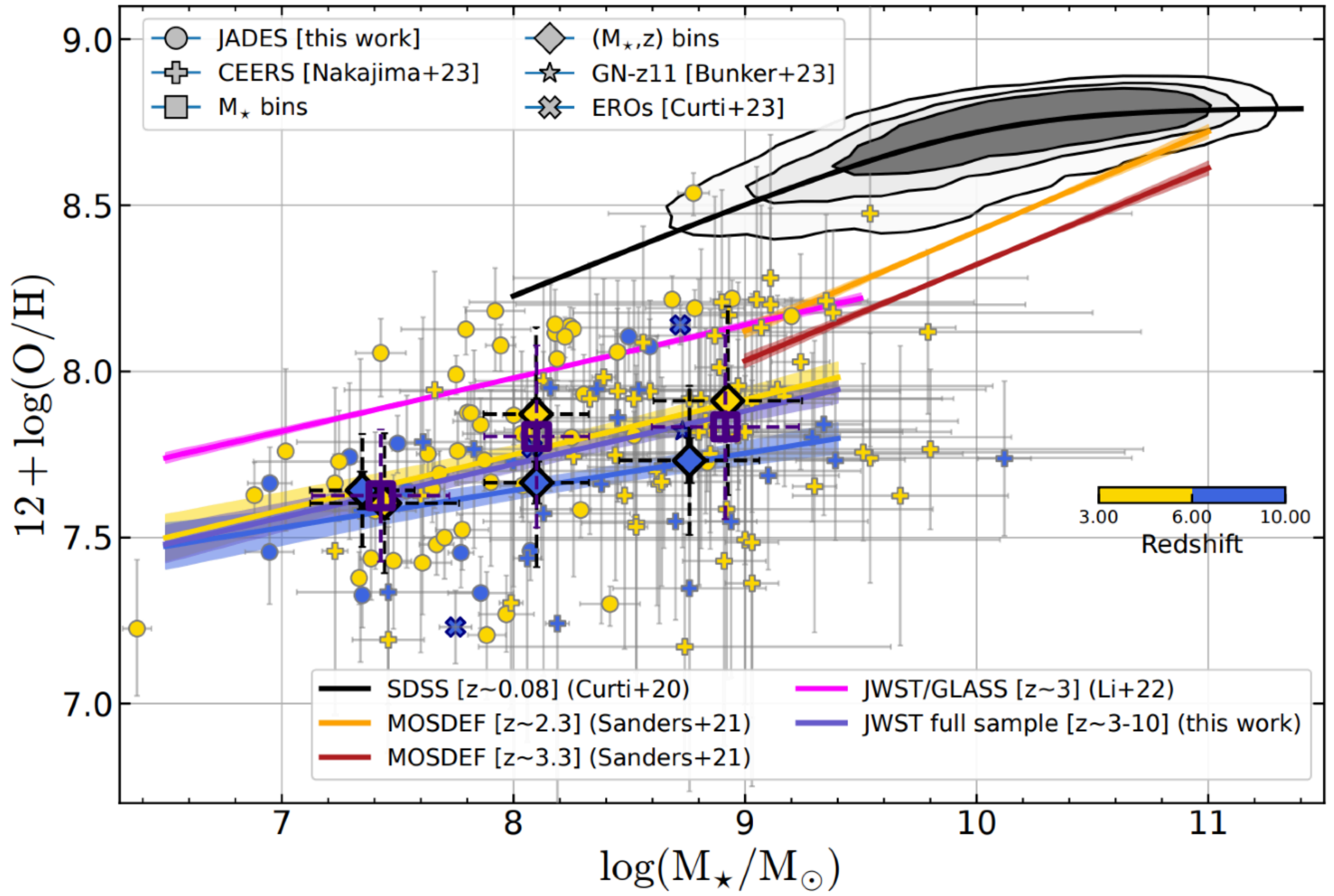




Tremonti+04

$z \sim 0$

> 50.000 galaxies from SDSS (~2.000.000 galaxy spectra)

~20 years (VLT, Keck… …JWST !)




Curti+24

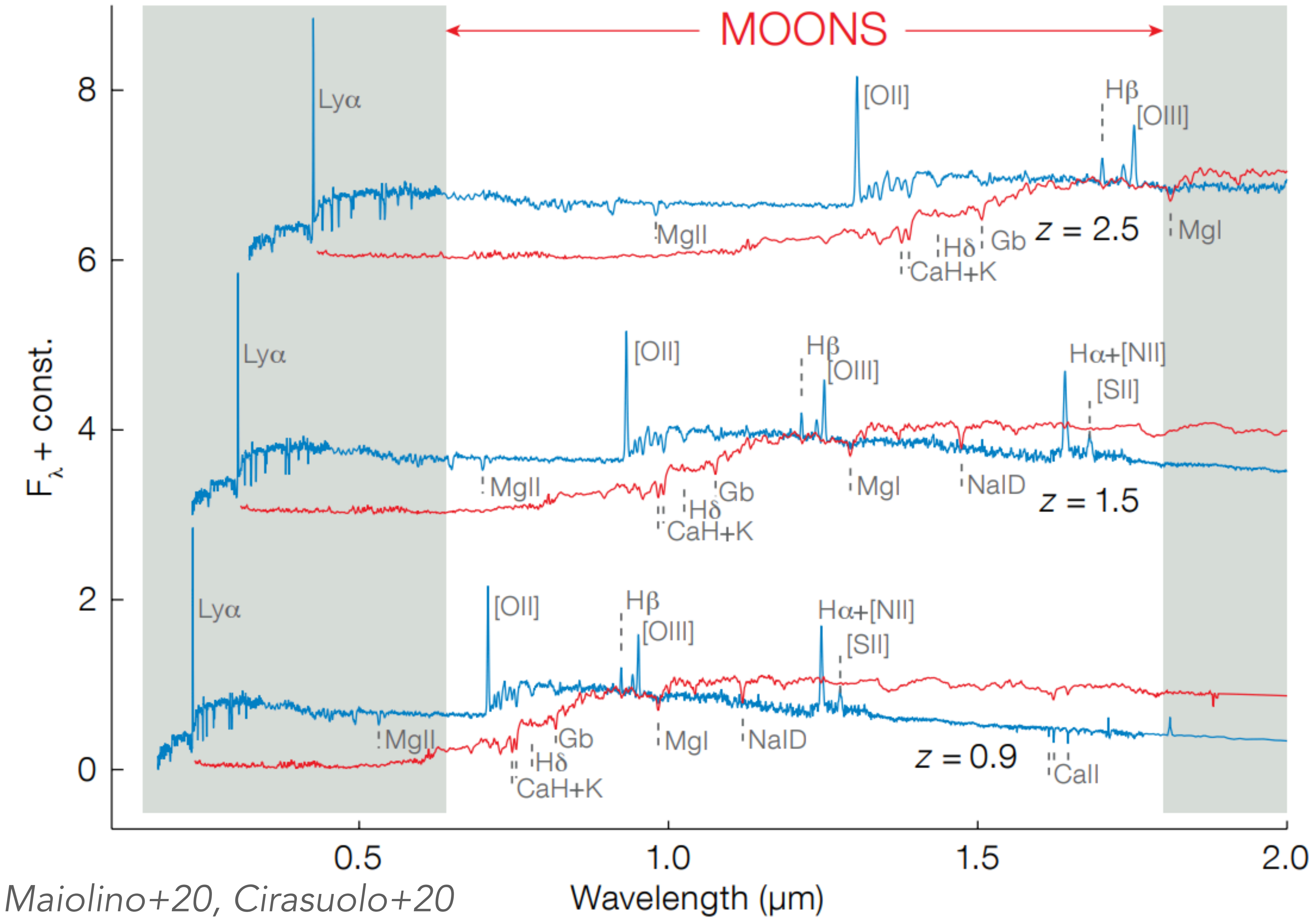
$z \sim 0 - 10$

- Galaxy evolution across cosmic time is a key topic of modern astro
- How it works? We observe galaxies at all epochs, measure their physical properties and use the relations among them to tune physical models.

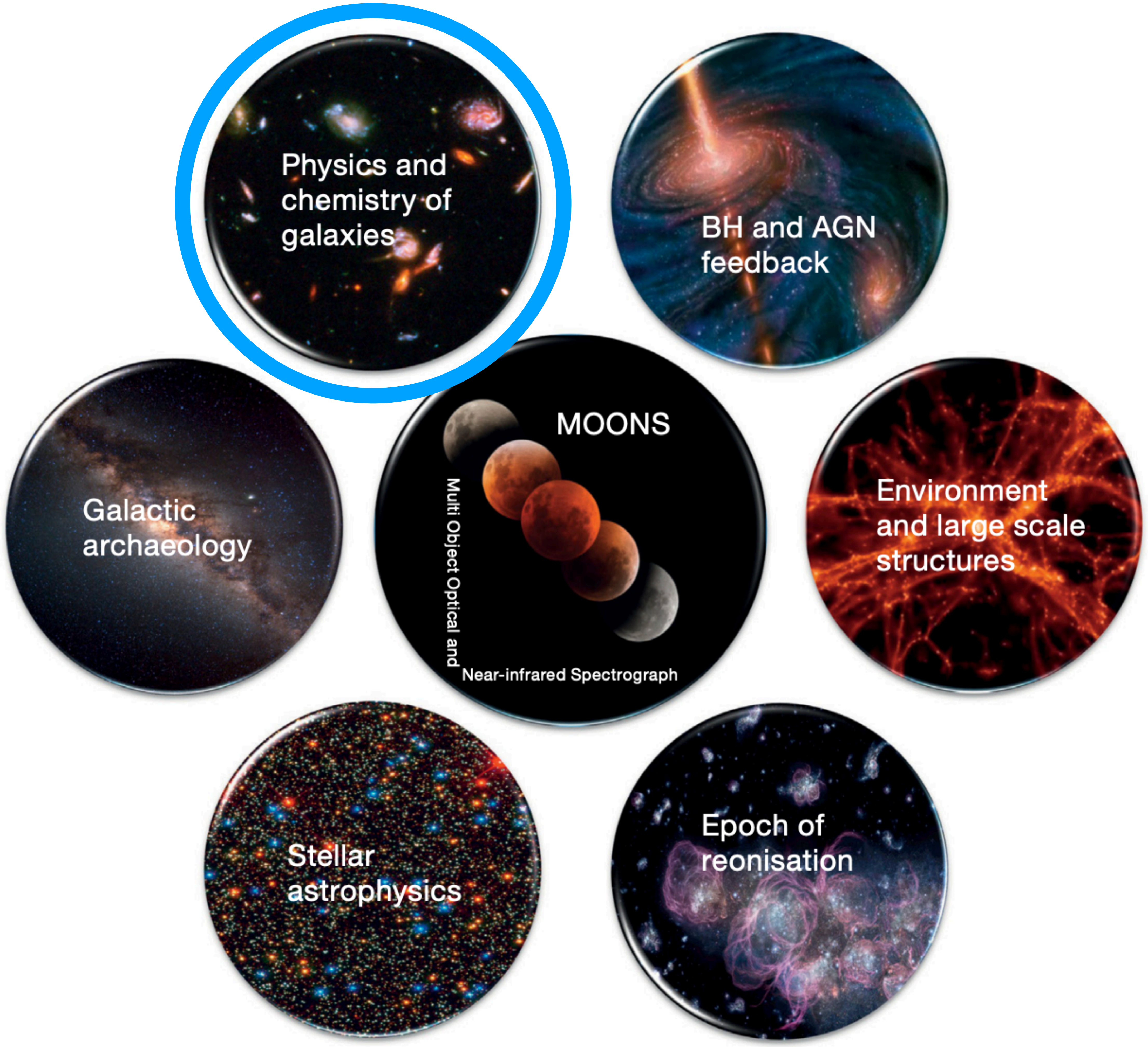
Upcoming all-sky spectroscopic survey (DESI, 4MOST, MOONS): **hundreds of millions of spectra** will be acquired in the next half-decade.

**<u>MOONS</u>** is the new **Multi-Object Optical and Near-infrared Spectrograph**, soon to be operated @VLT, ESO (8-m telescope)

- 1000 fibres, over a field of view of $\sim 500^2$ arcmin;
- low- (R~4000–7000) / high-resolution (~19000 in *H*);
- 0.64 – 1.8 µm wavelength range.




*Maiolino+20, Cirasuolo+20*

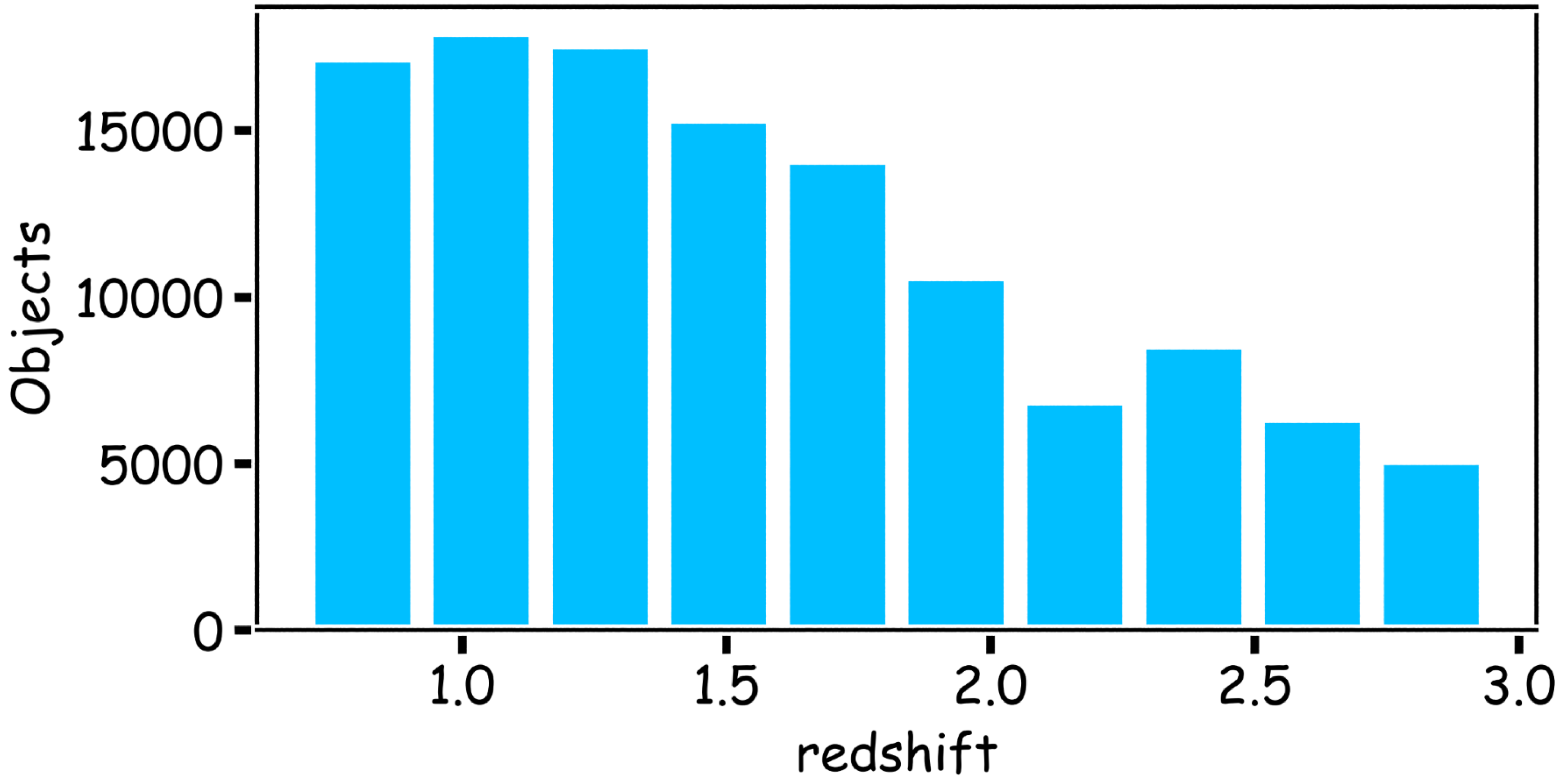


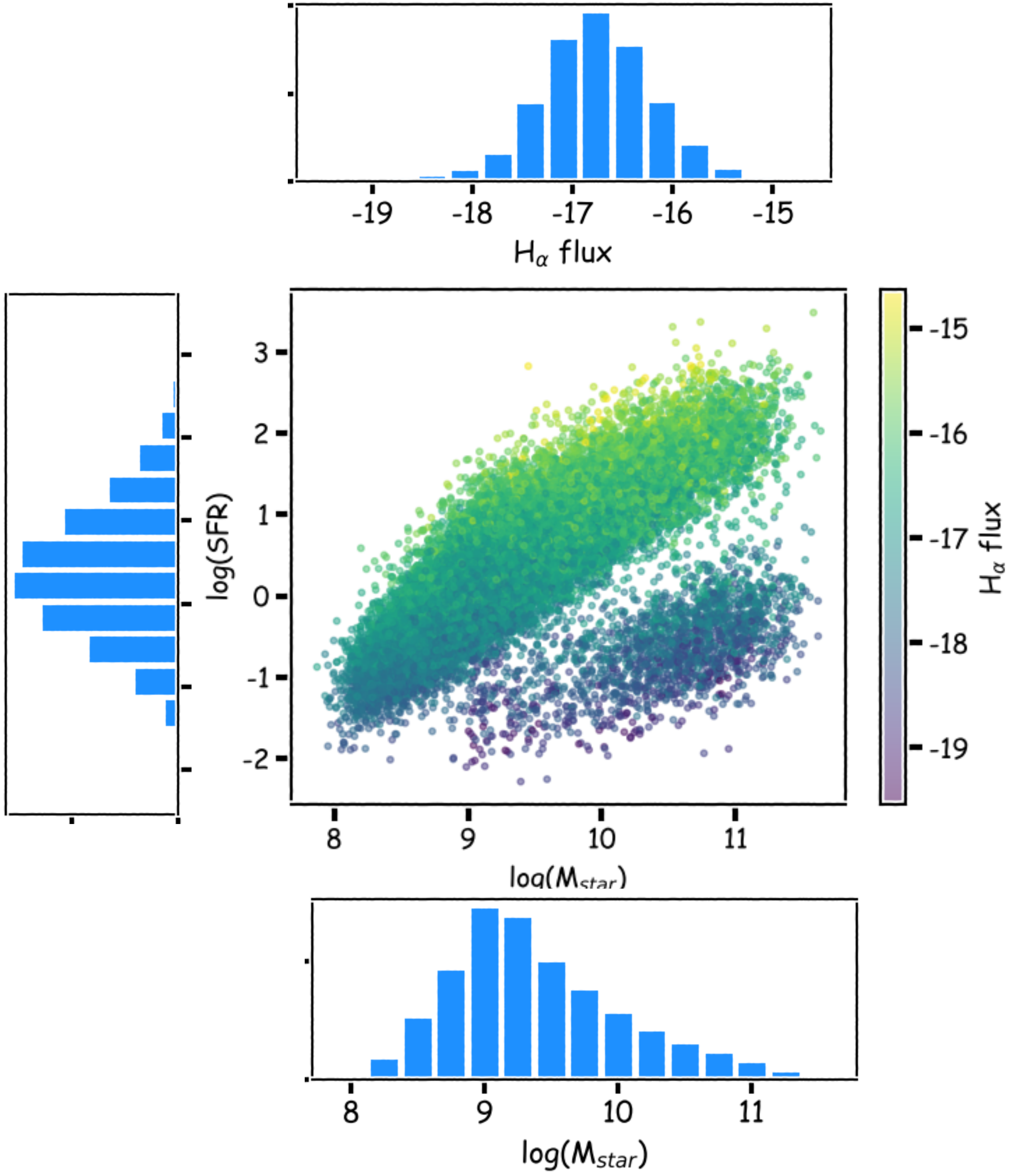

<u>Any data challenge?</u>

- up to about **half a million galaxies** at $0.9 < z < 2.6$
- >12000 elements per spectrum in low-resolution!
- Standard fitting methods are slow and fail in weak-signal regimes

# Simulated dataset

- **~120.000** spectra
- generated by running **MAMBO** templates through **moons1d**
- moons1d ran with low resolution mode for all 3 channels (RI, YJ, H), **0.64-1.8 μm**
- a seeing of 0.8'' and airmass of 1.2
- $t_{exp} = 2 - 4 - 8\ h$



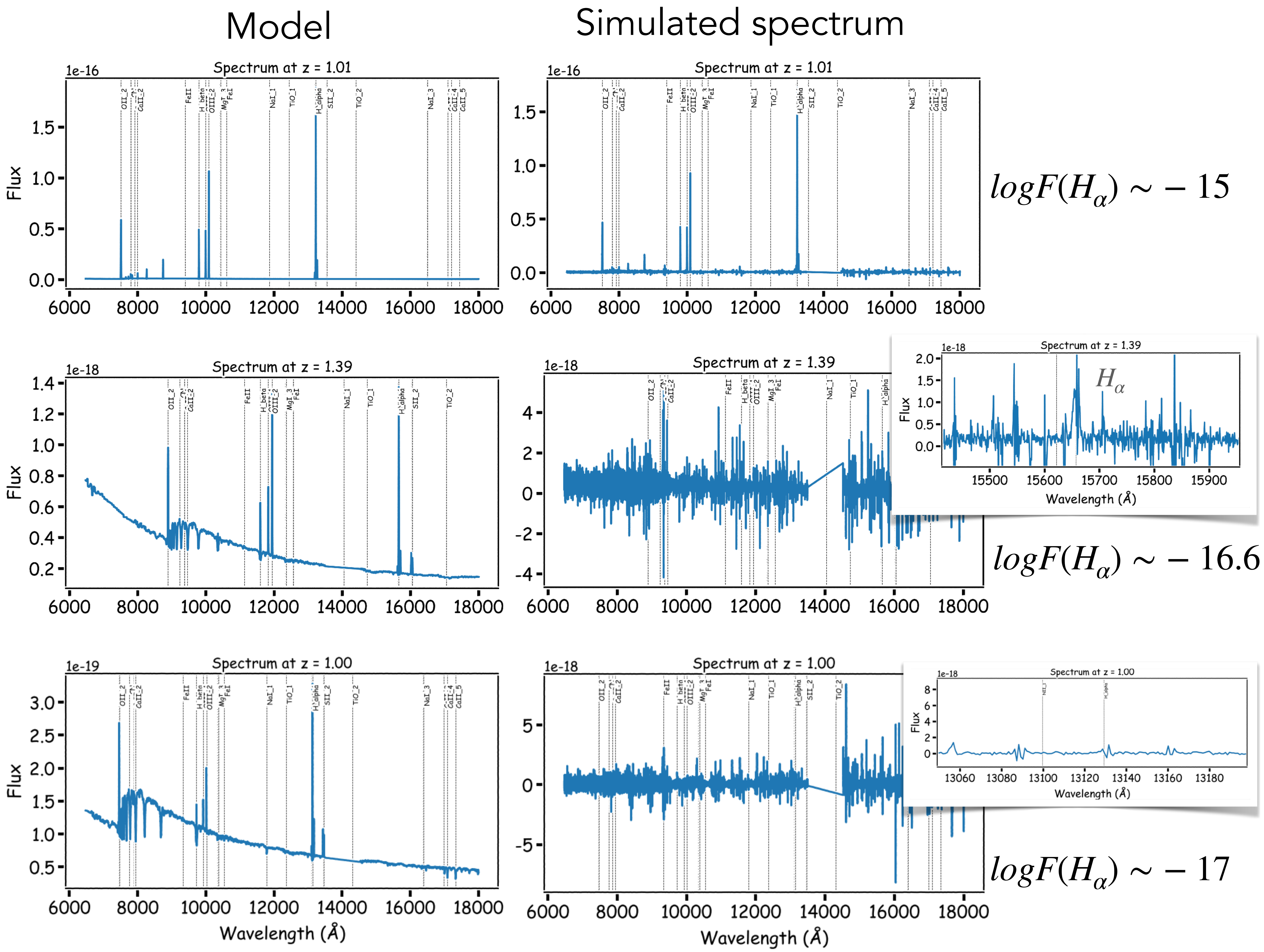



# Dataset

- $t_{exp} = 2, 4, 8\ h$
- 0.64 – 1.8 μm
- 12.217 channels

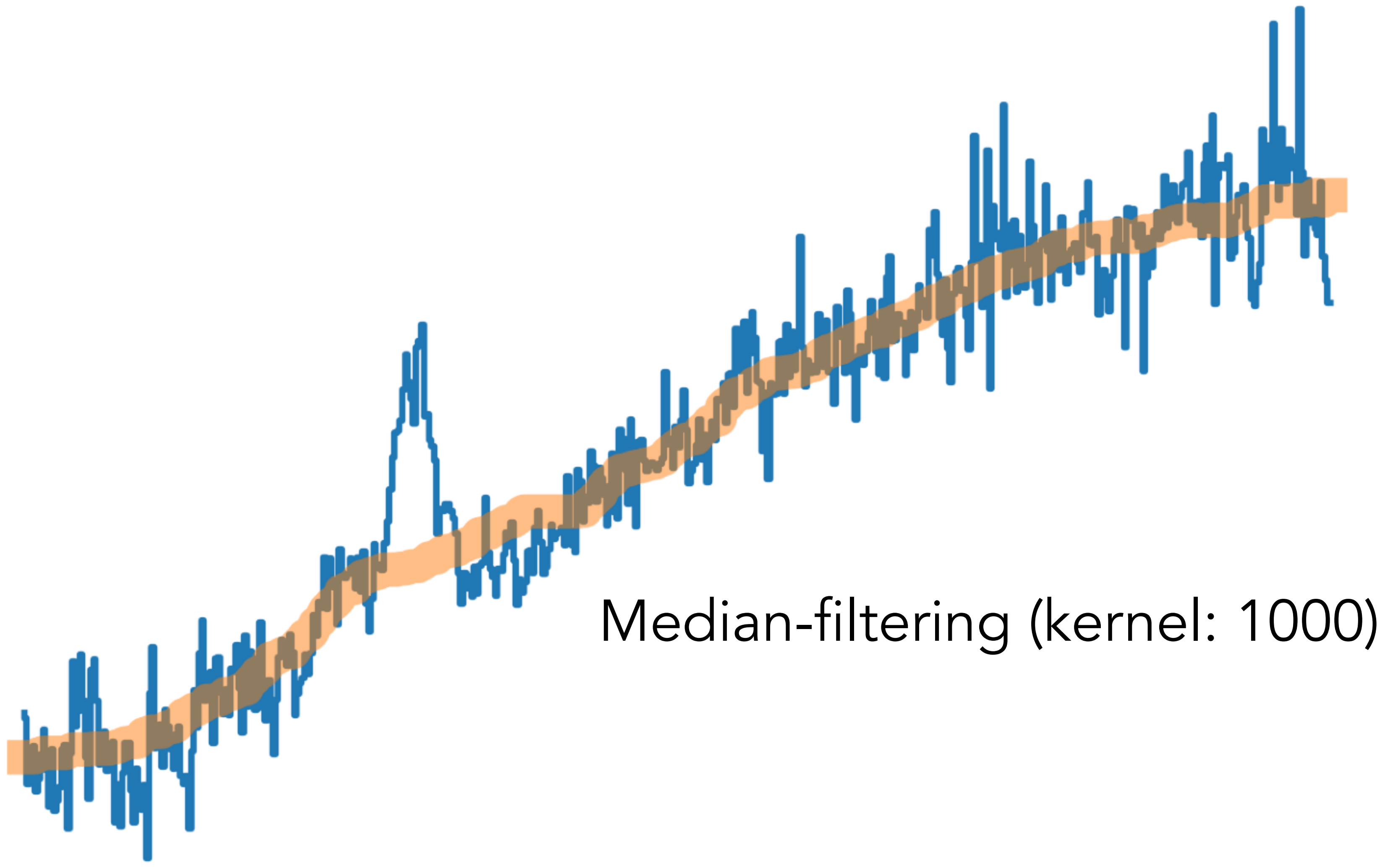
## Target physics

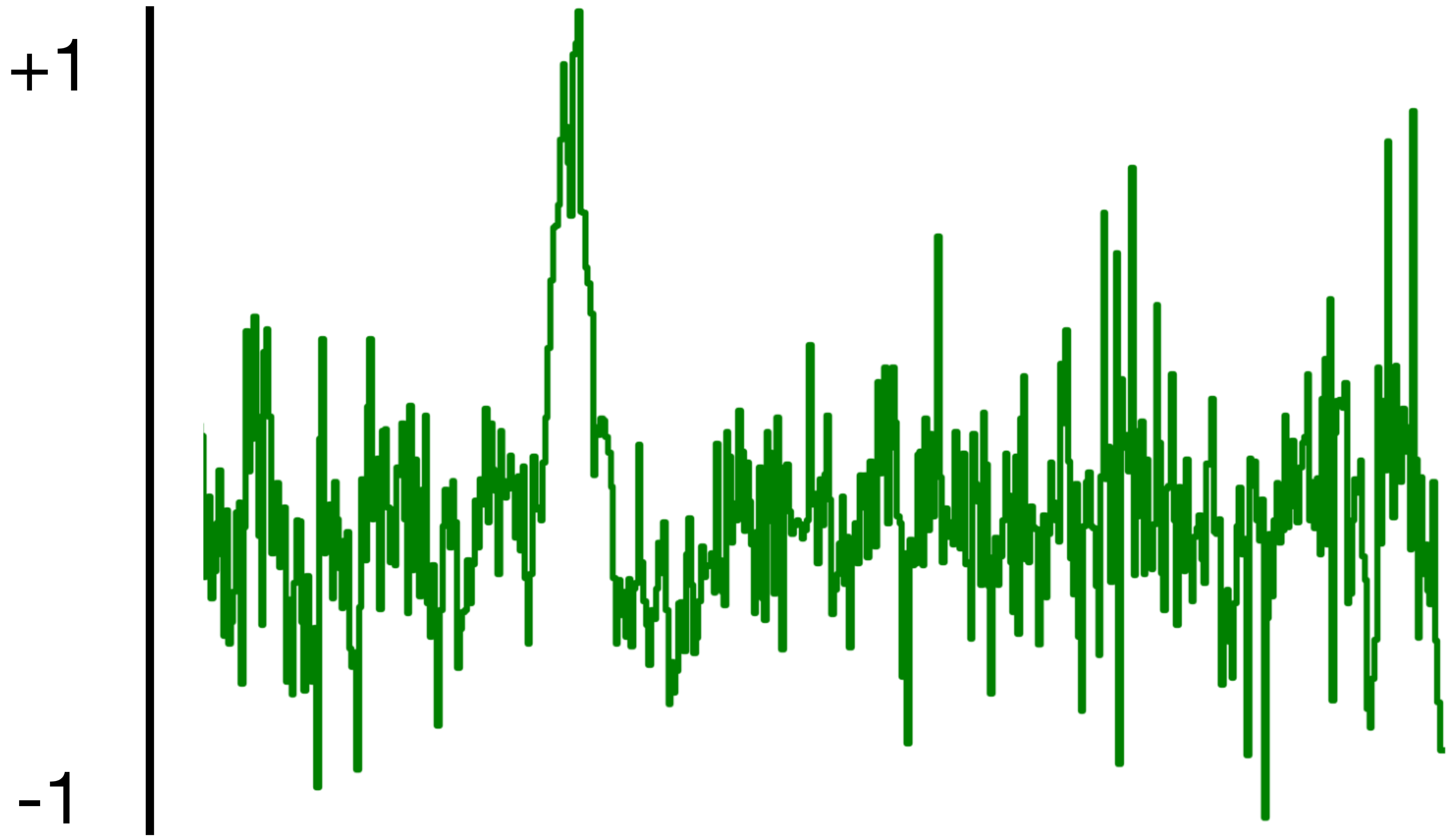
- redshift, $z$
- stellar mass, $M_{\rm star}$
- star formation rate, $SFR$



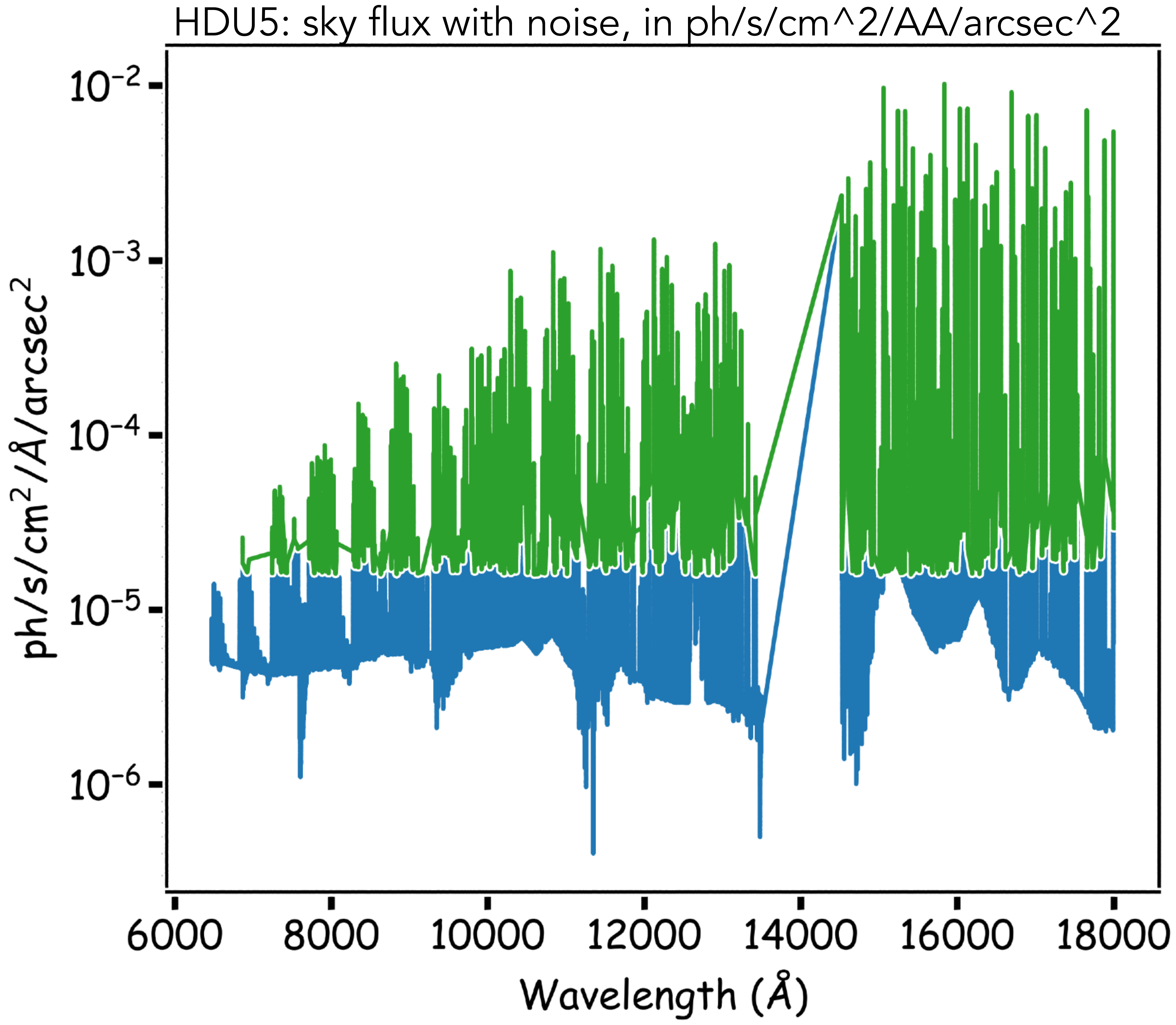
# Pre-processing

## Continuum subtraction & Min-max normalisation




## Sky-masking

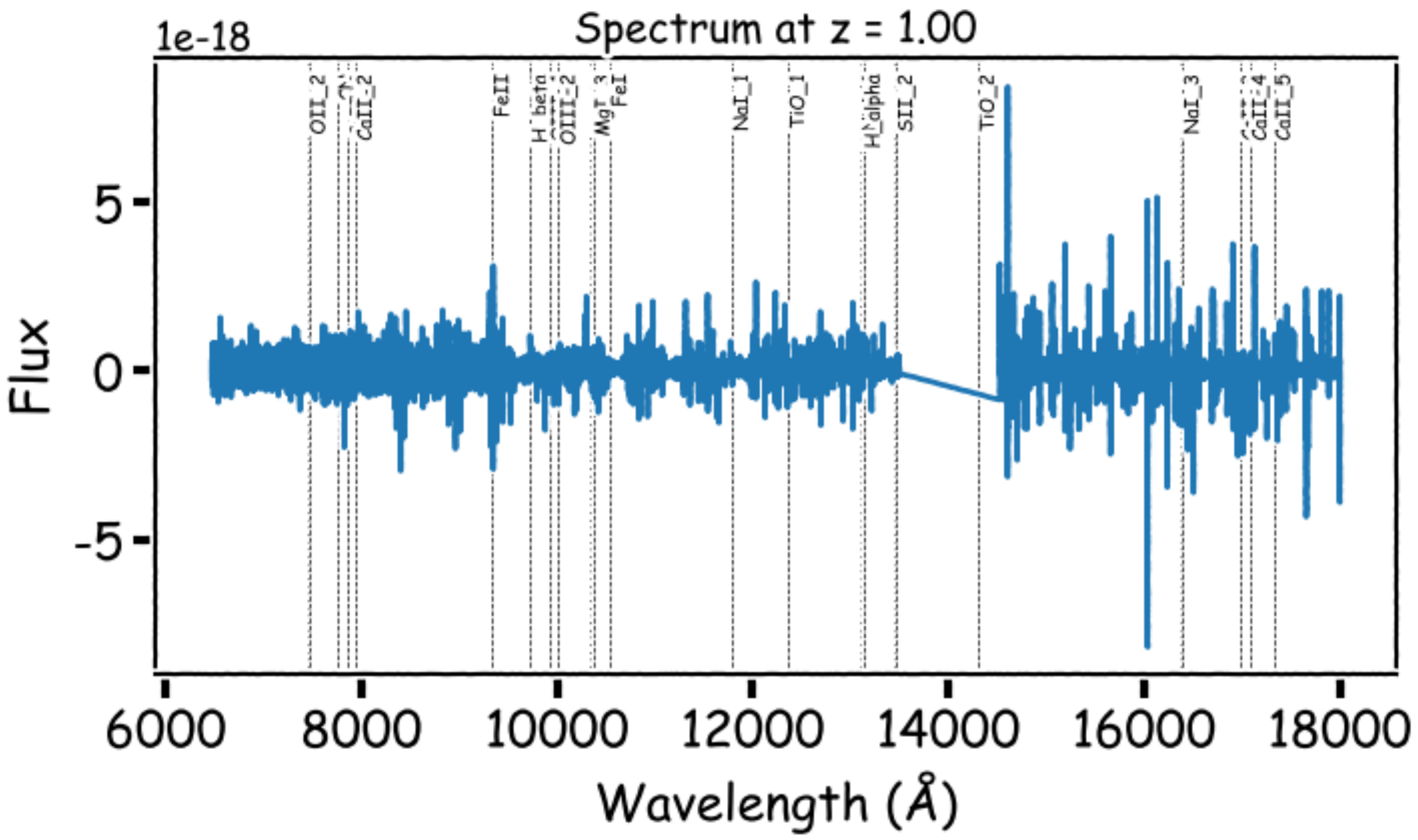


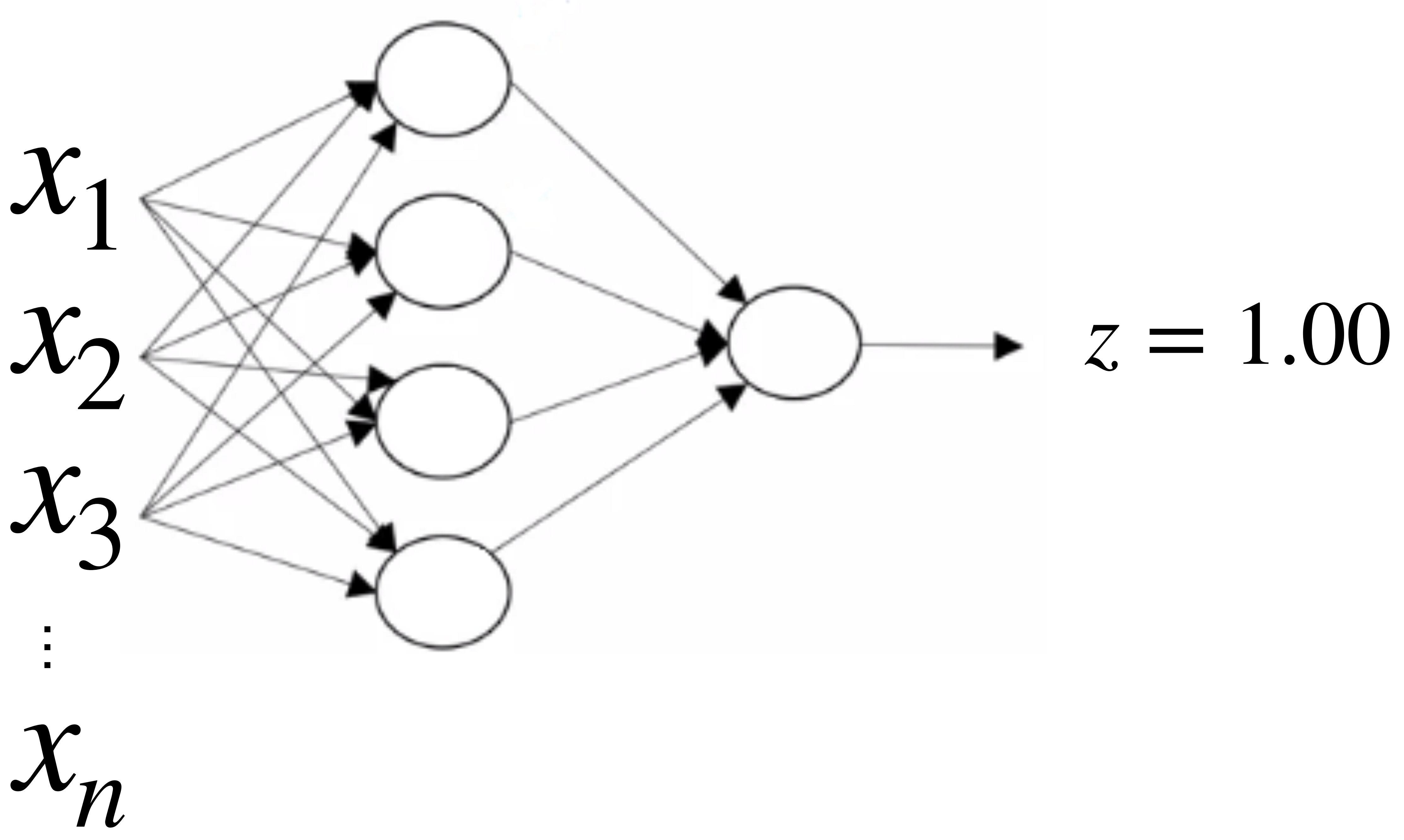

Masked about ~ 15% of spectral channels

# Deep learning MOONS spectra

## The case of redshift

*Classical scheme: a regression problem*

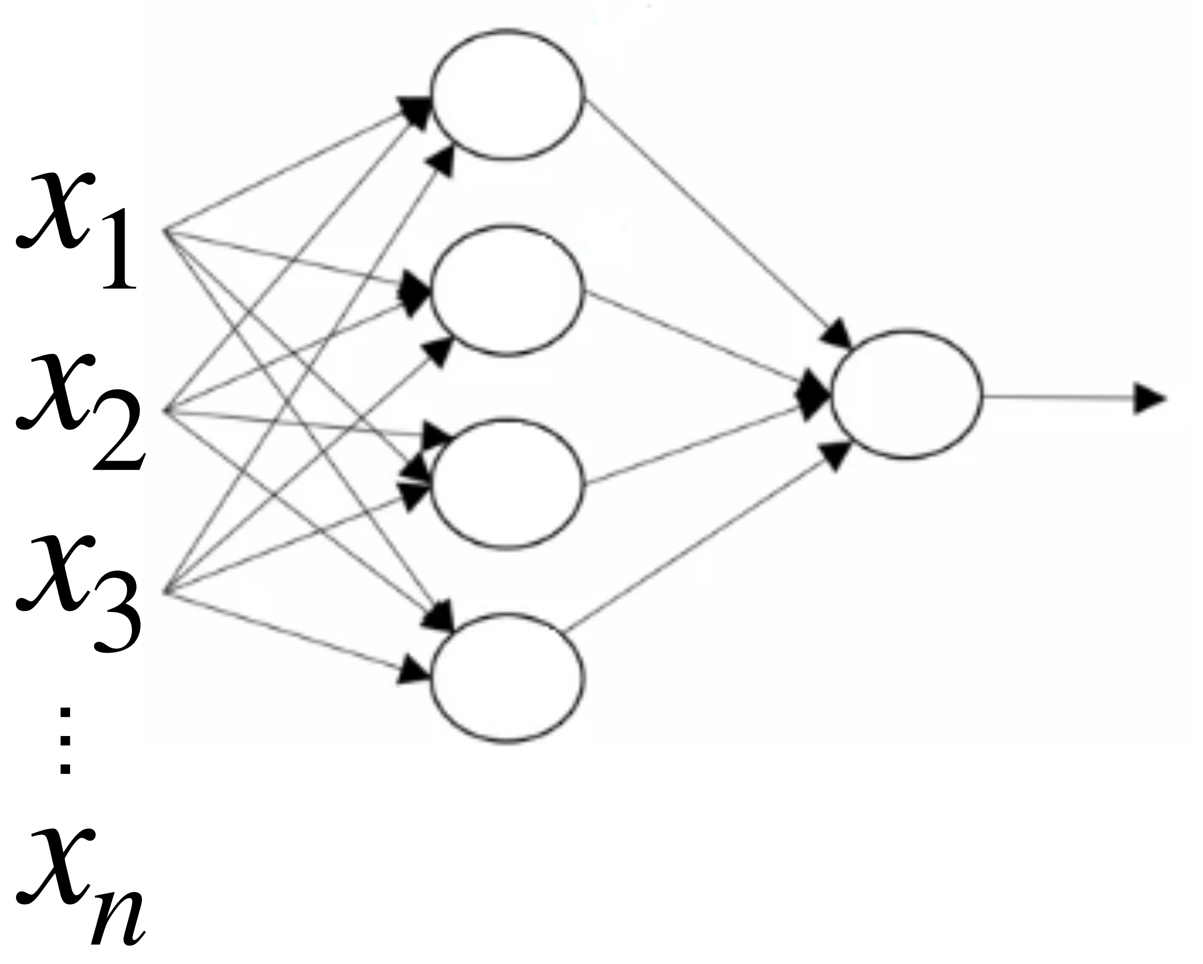


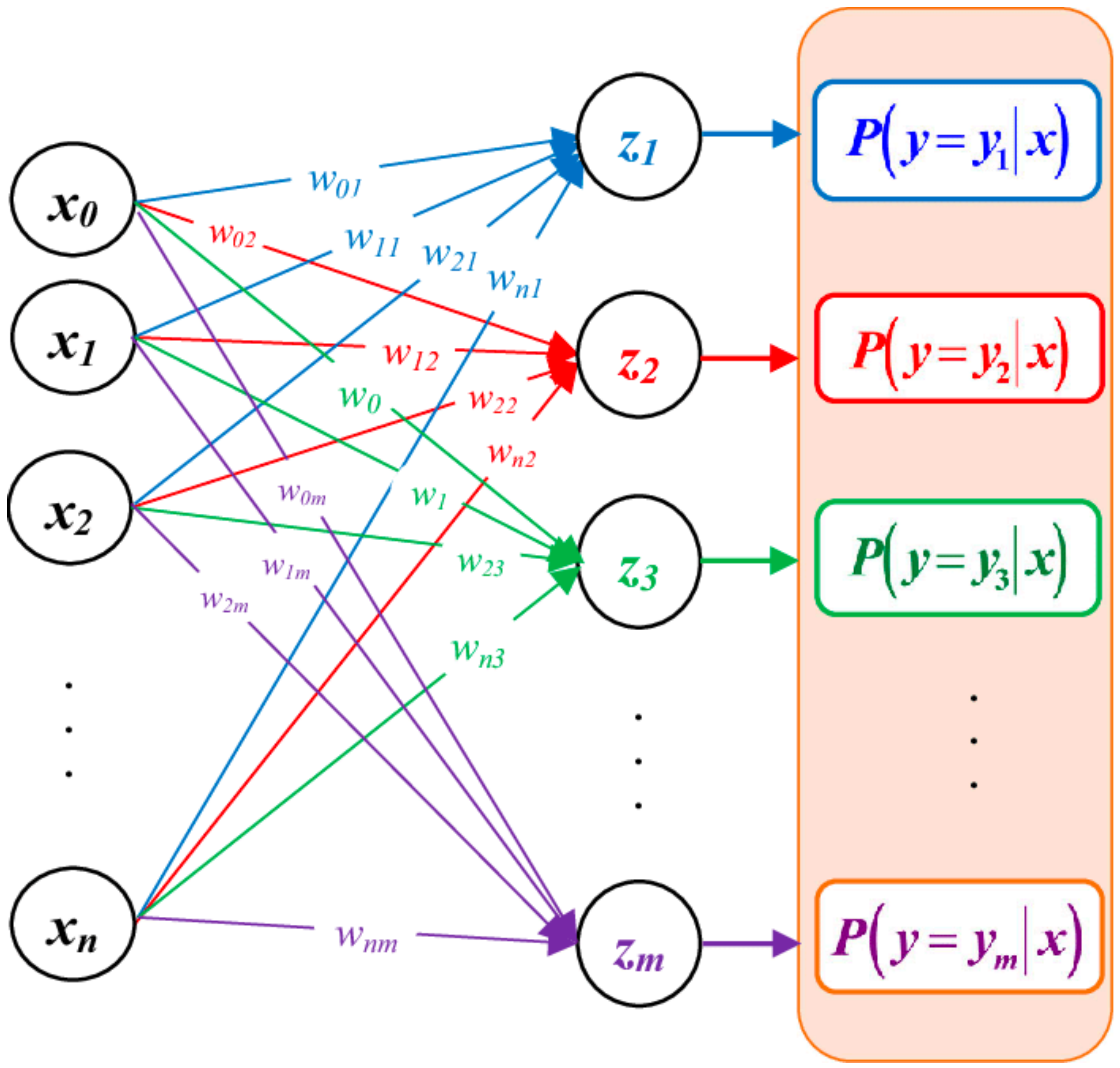
# Switch to a classification problem

… and let the “softmax + cross-entropy” team work for us (better than the “linear + mse” team)   :-)

We adopt $dz = 0.003$




$z = 1.00$




$z = 0.997 \quad (P = 0.1)$

$z = 1.000 \quad (P = 0.95)$

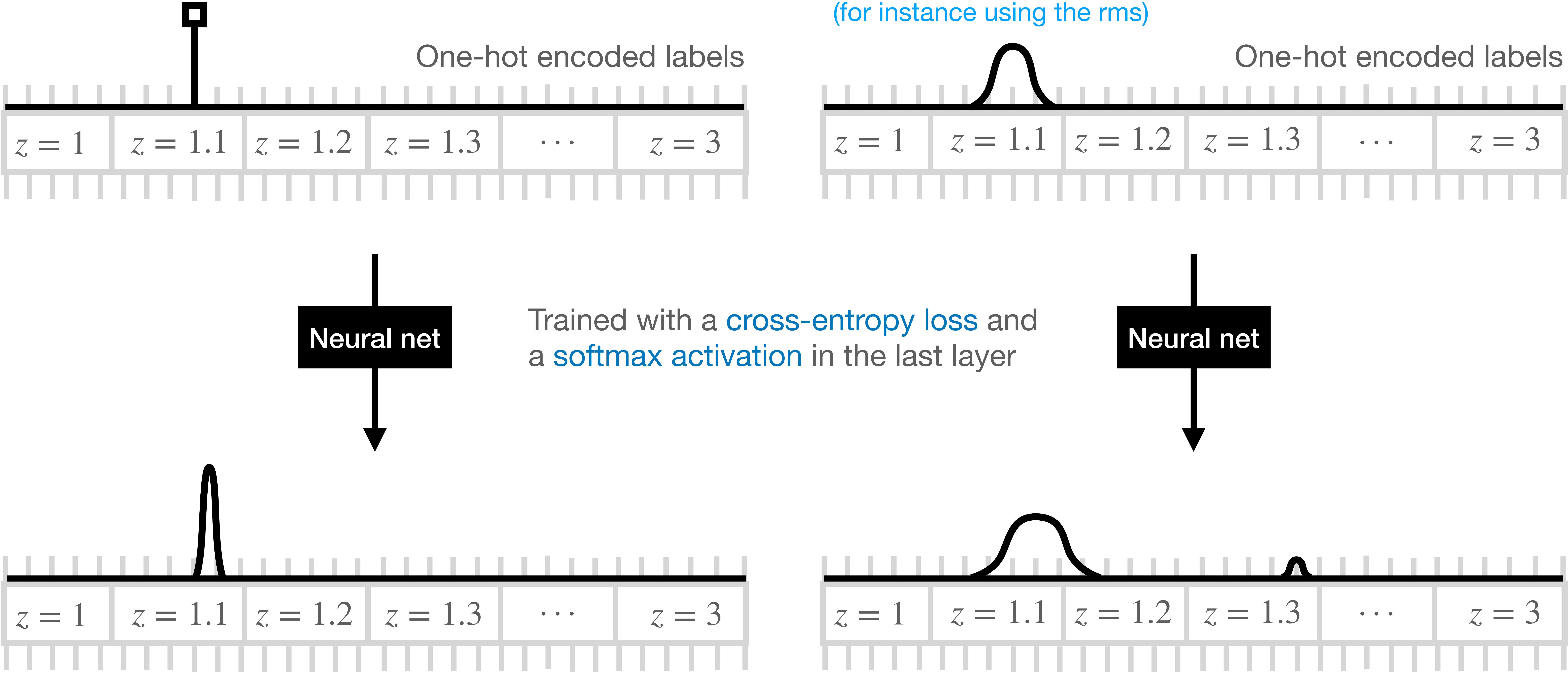
$z = 1.003 \quad (P = 0.05)$

$z = 2 \quad (P = 0)$

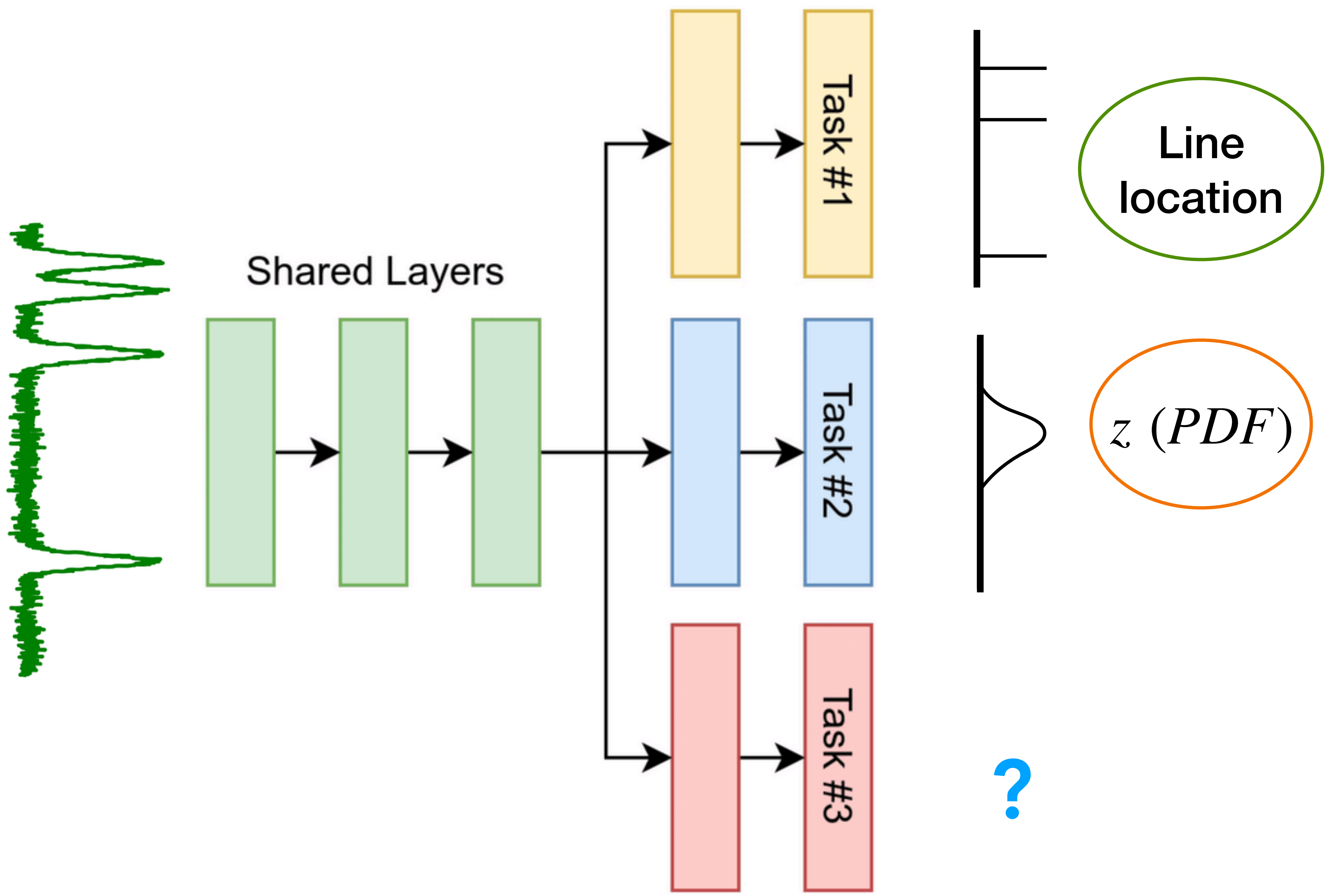
# Switch to a classification problem

We can also account for the *quality* of spectra

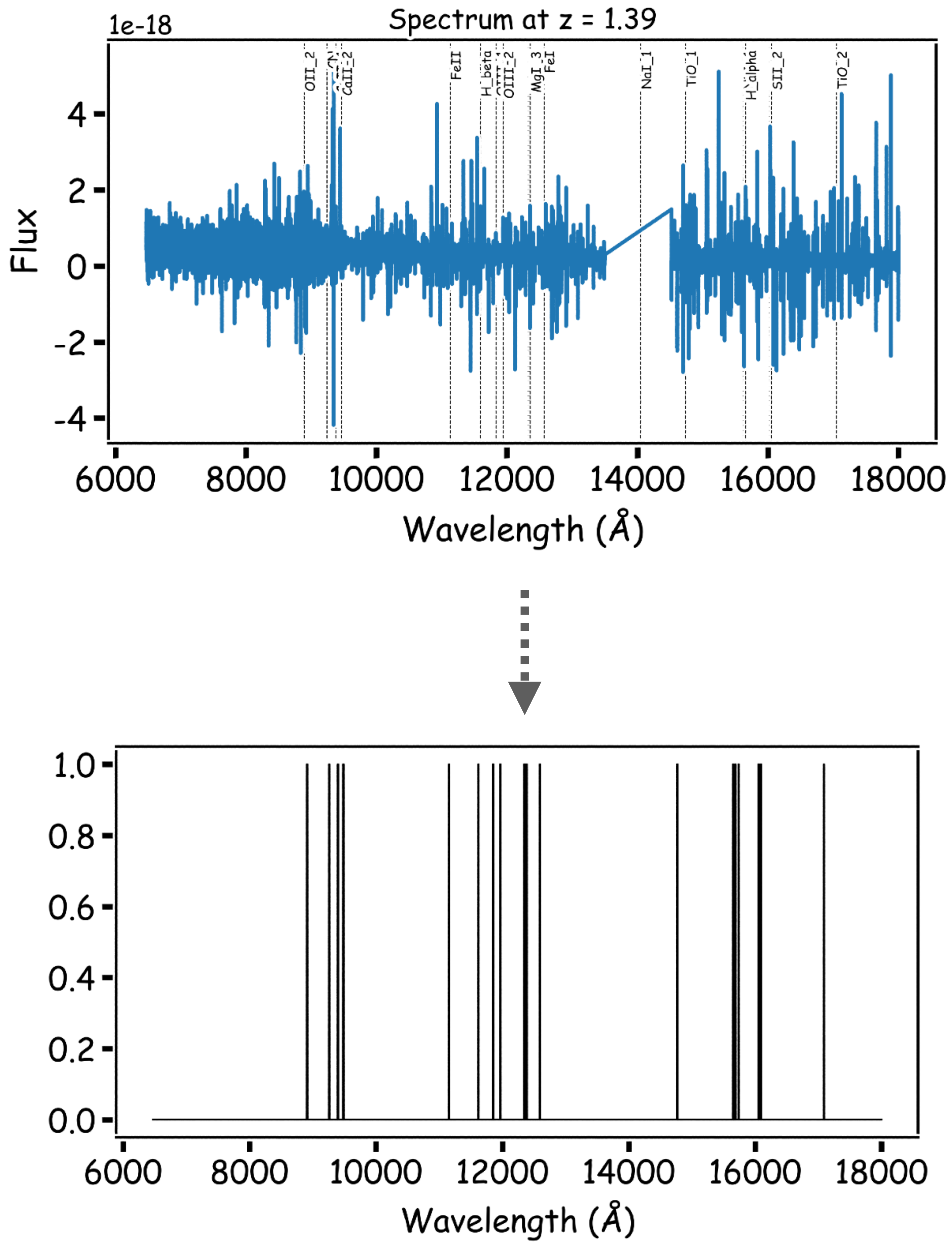
(for instance using the rms)



# Learning through multi-task training




## Line-Location Task



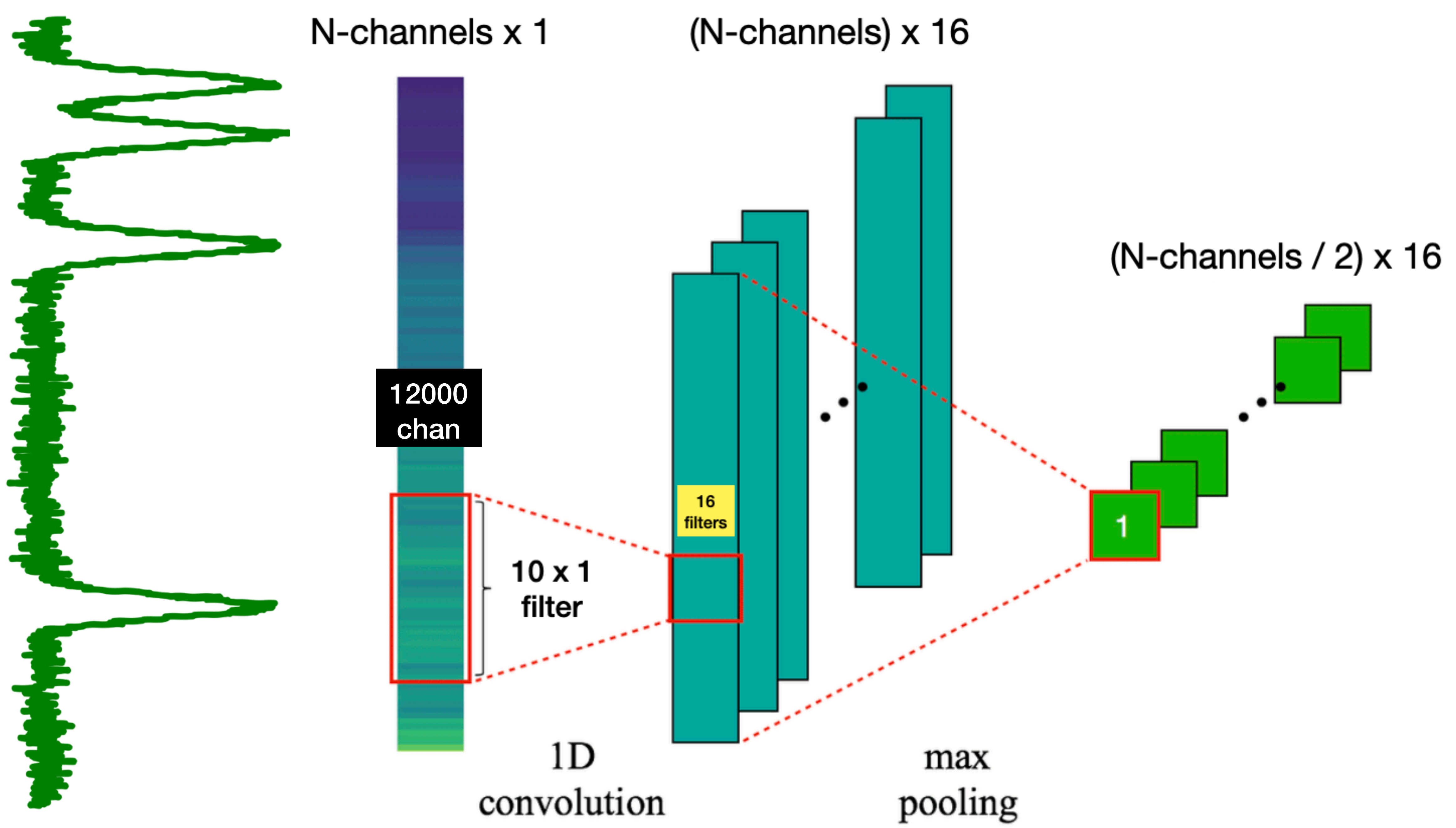

```
emission_lines_rest = {
   'OII_1': 3727.1,          # Oxygen II
   'OII_2': 3729.9,          # Oxygen II
   'H_beta': 4862.7,          # Hydrogen-beta
   'OIII_1': 4960.3,         # Oxygen III
   'OIII_2': 5008.2,         # Oxygen III
   'NII_1': 6549.8,          # Nitrogen II
   'H_alpha': 6564.6,         # Hydrogen-alpha
   'NII_2': 6585.3,          # Nitrogen II
   'SII_1': 6718.3,          # Sulfur II
   'SII_2': 6732.7,          # Sulfur II
   'SIII_1': 9070,          # Sulfur III
   'SIII_2': 9532,          # Sulfur III
  'NV': 1240,              # Nitrogen V
   'SiII_1': 1260,          # Silicon II
   'OI': 1303,              # Oxygen I
   'CII': 1334,             # Carbon II
   'SiIV_1': 1393,           # Silicon IV
   'SiIV_2': 1402,           # Silicon IV
   'SiII_2': 1526,          # Silicon II
   'CIV_1': 1548,            # Carbon IV
   'HeII_1': 1640,           # Helium II
   'OII_3': 1660,            # Oxygen II
   'OII_4': 1666,            # Oxygen II
   'CIII': 1909,            # Carbon III
   'CN': 3875,               # Cyanide radical
   'CaII_1': 3933,           # Calcium II
   'CaII_2': 3969,           # Calcium II
   'FeII': 4668,            # Iron II
   'MgI_1': 5167,             # Magnesium I
   'MgI_2': 5172,             # Magnesium I
   'MgI_3': 5183,             # Magnesium I
   'FeI': 5270,             # Iron I
   'NaI_1': 5892,            # Sodium I
  'NaI_2': 8183,            # Sodium I
   'NaI_3': 8195,            # Sodium I
   'CaII_3': 8489,           # Calcium II
   'CaII_4': 8542,           # Calcium II
   'CaII_5': 8662,           # Calcium II
   'HeII_2': 10830           # Helium II
}
```

*Ruder+17; Crawshaw+20; Hervella+24 etc*

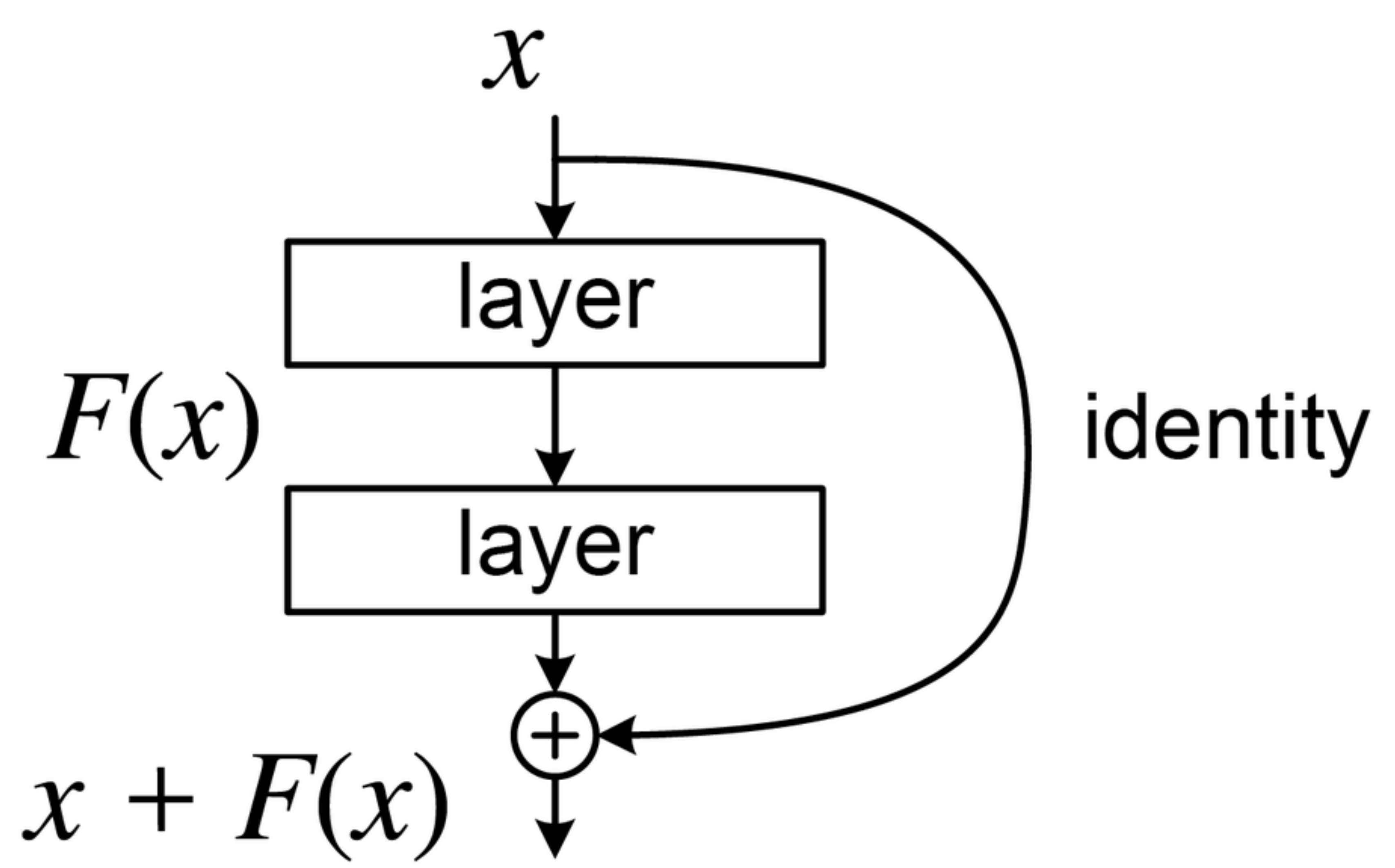
# Learning through multi-task training

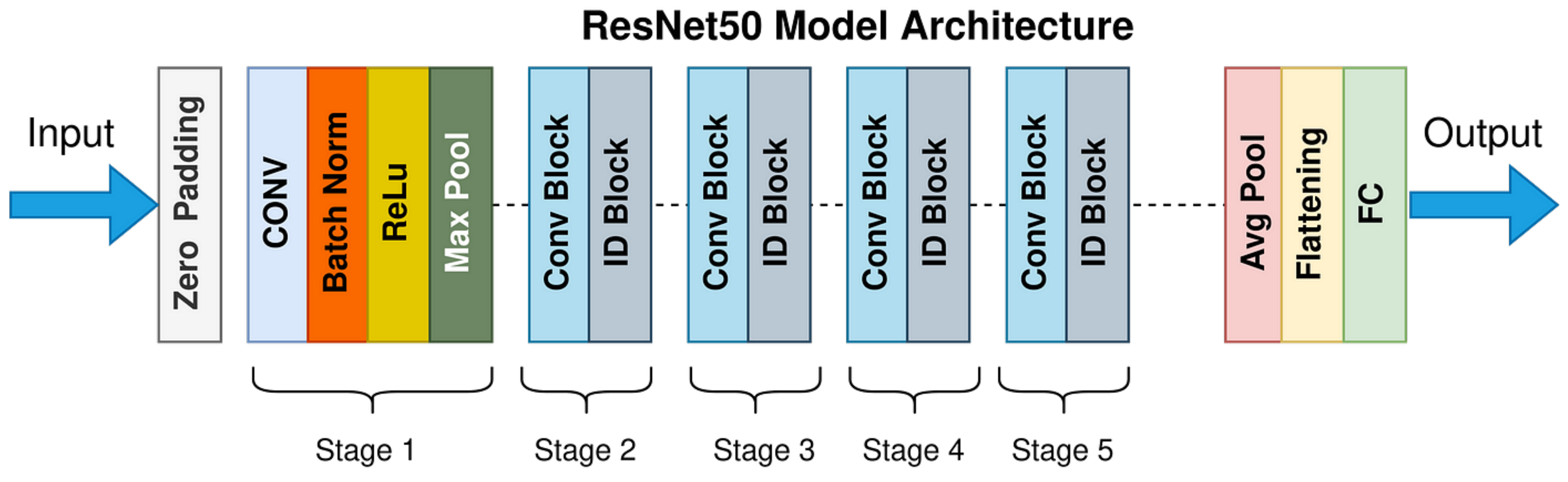
# Residual blocks with convolutional layers

## 1d convolution scheme




## Residual learning

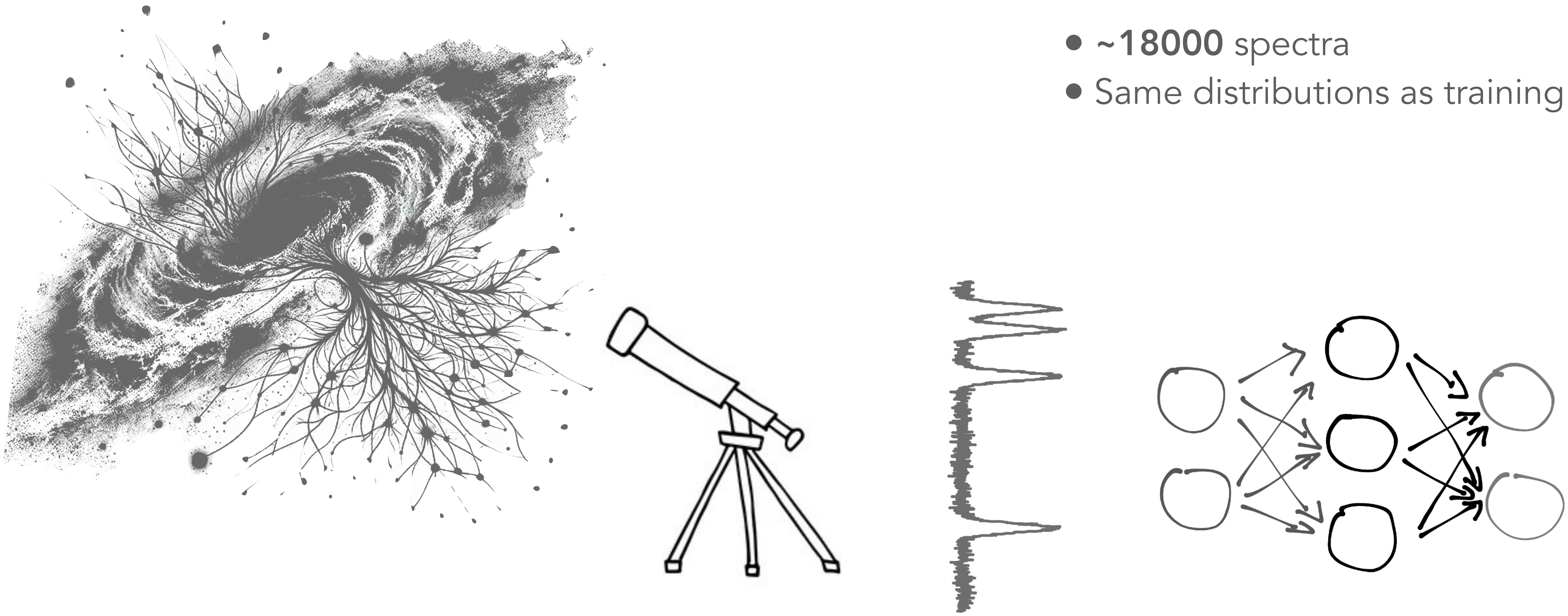


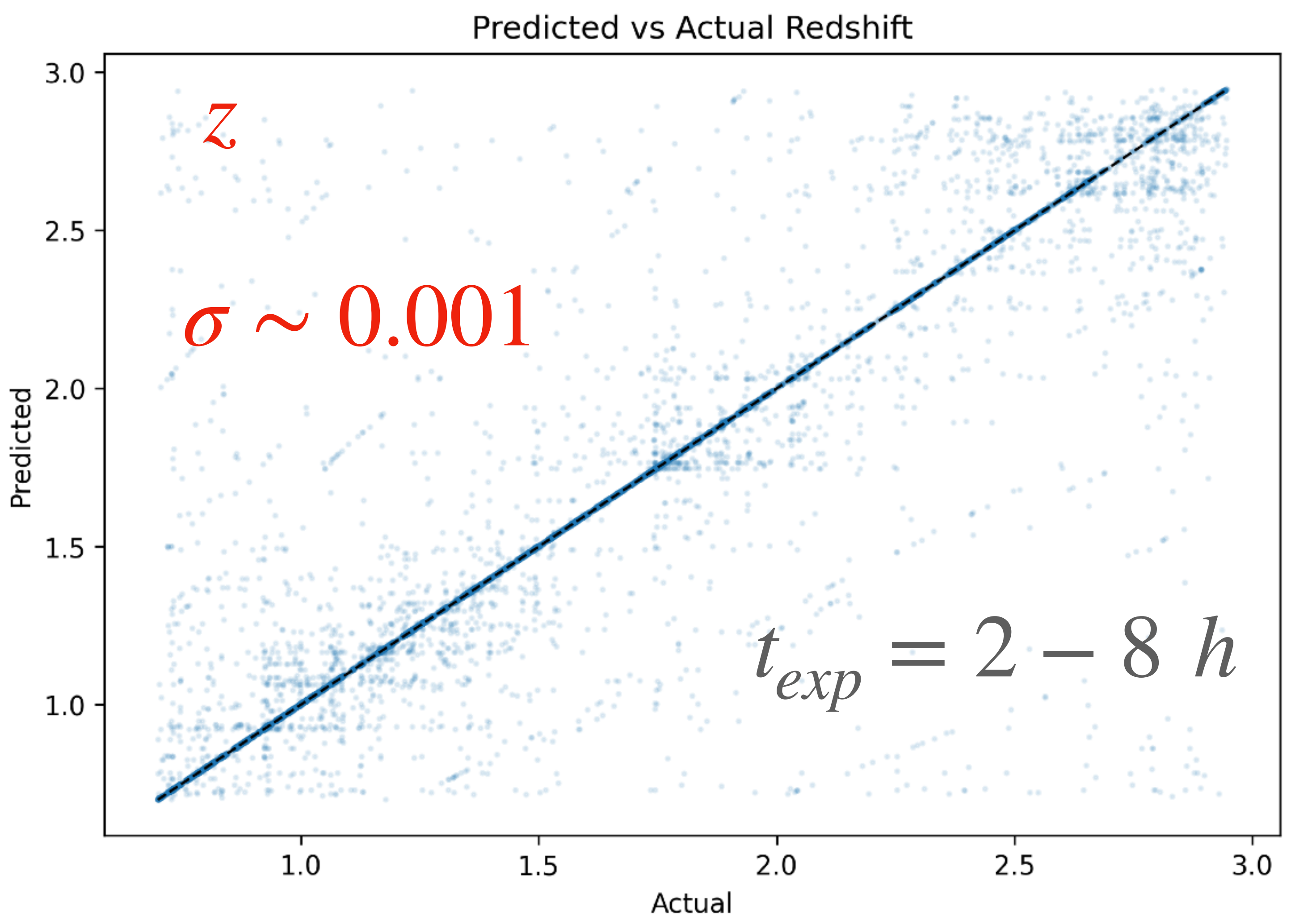

*He+15 (ResNet); Yoo+22*

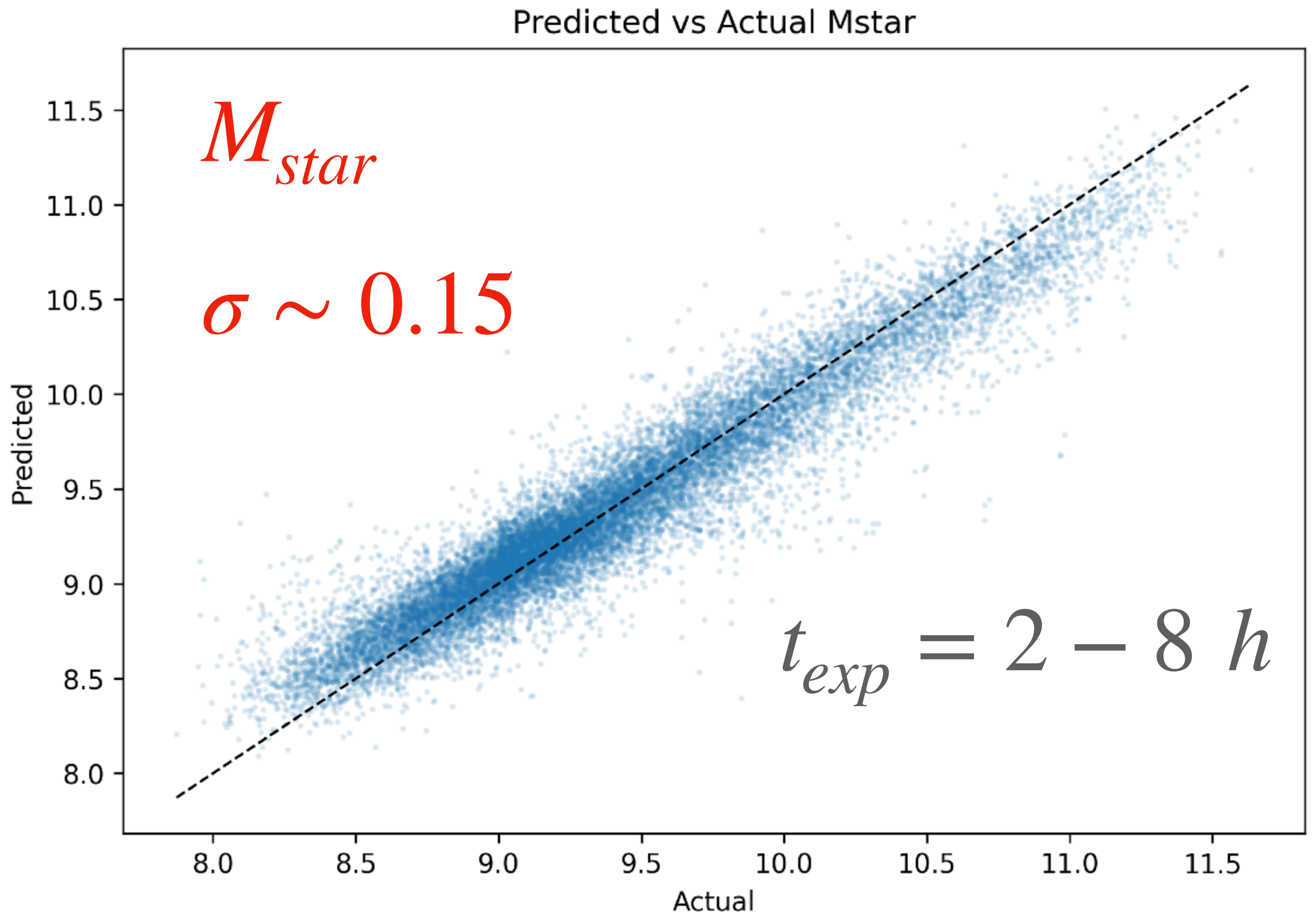
# Predictions on the a test set

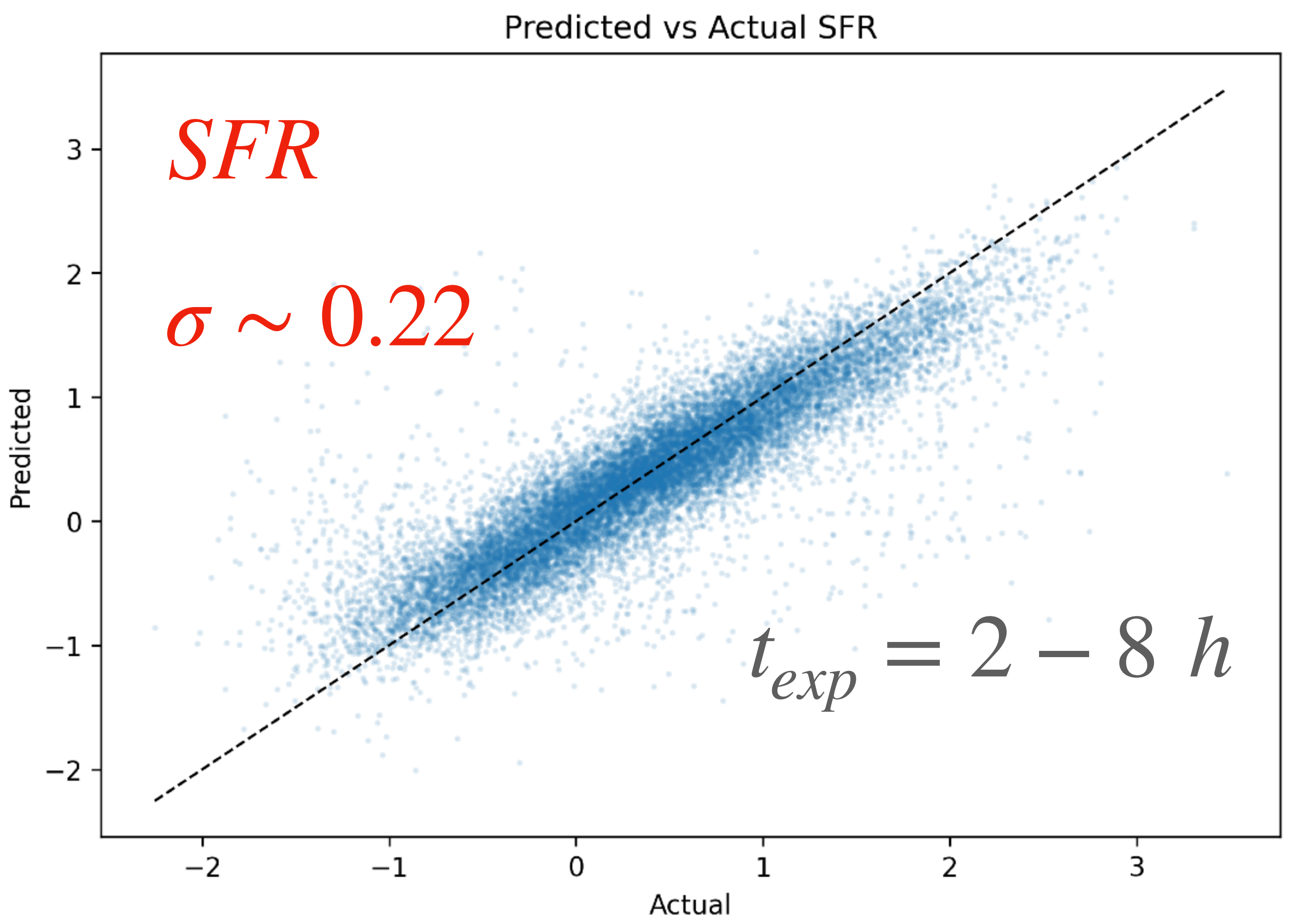
- **~18000** spectra
- Same distributions as training set

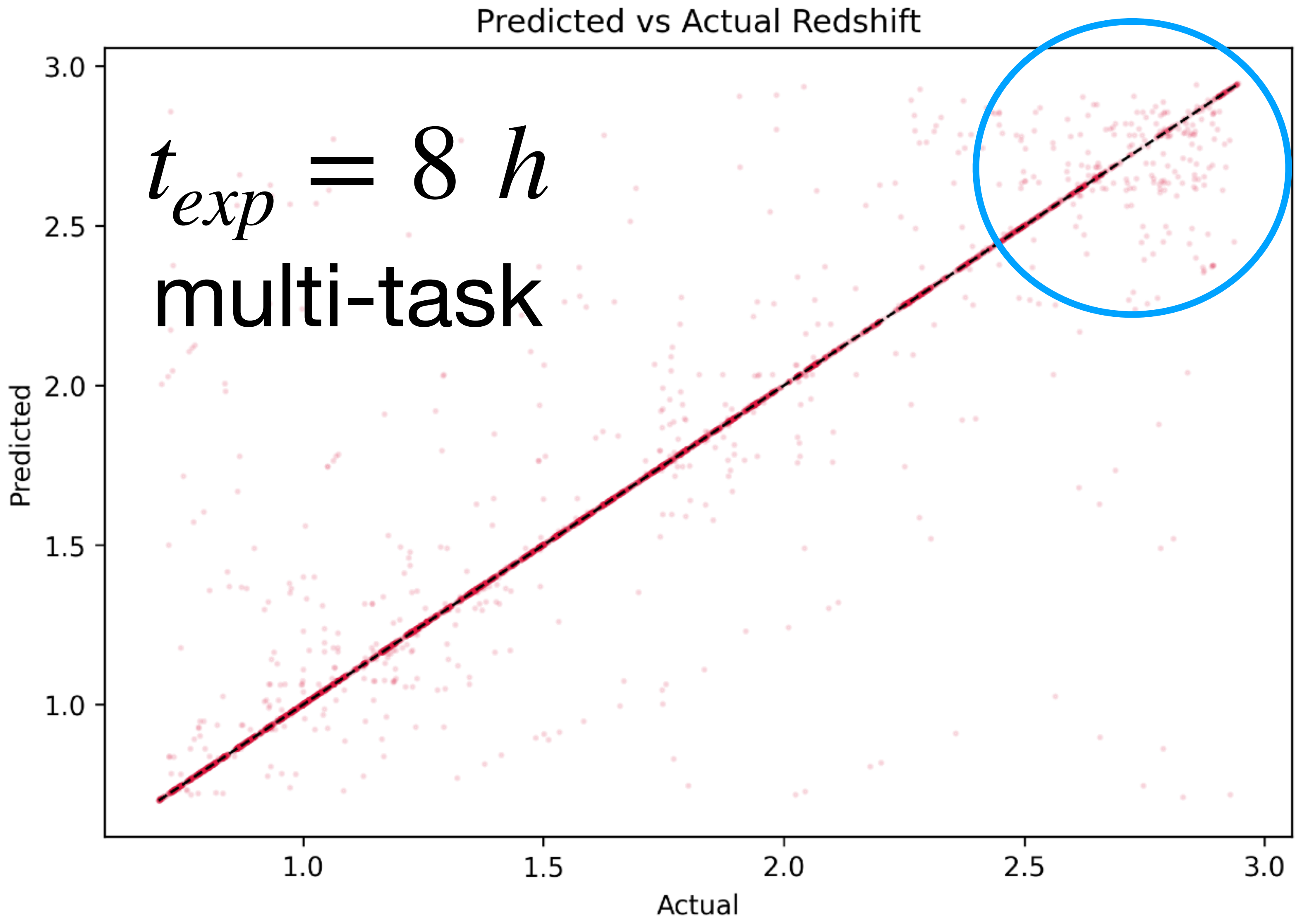


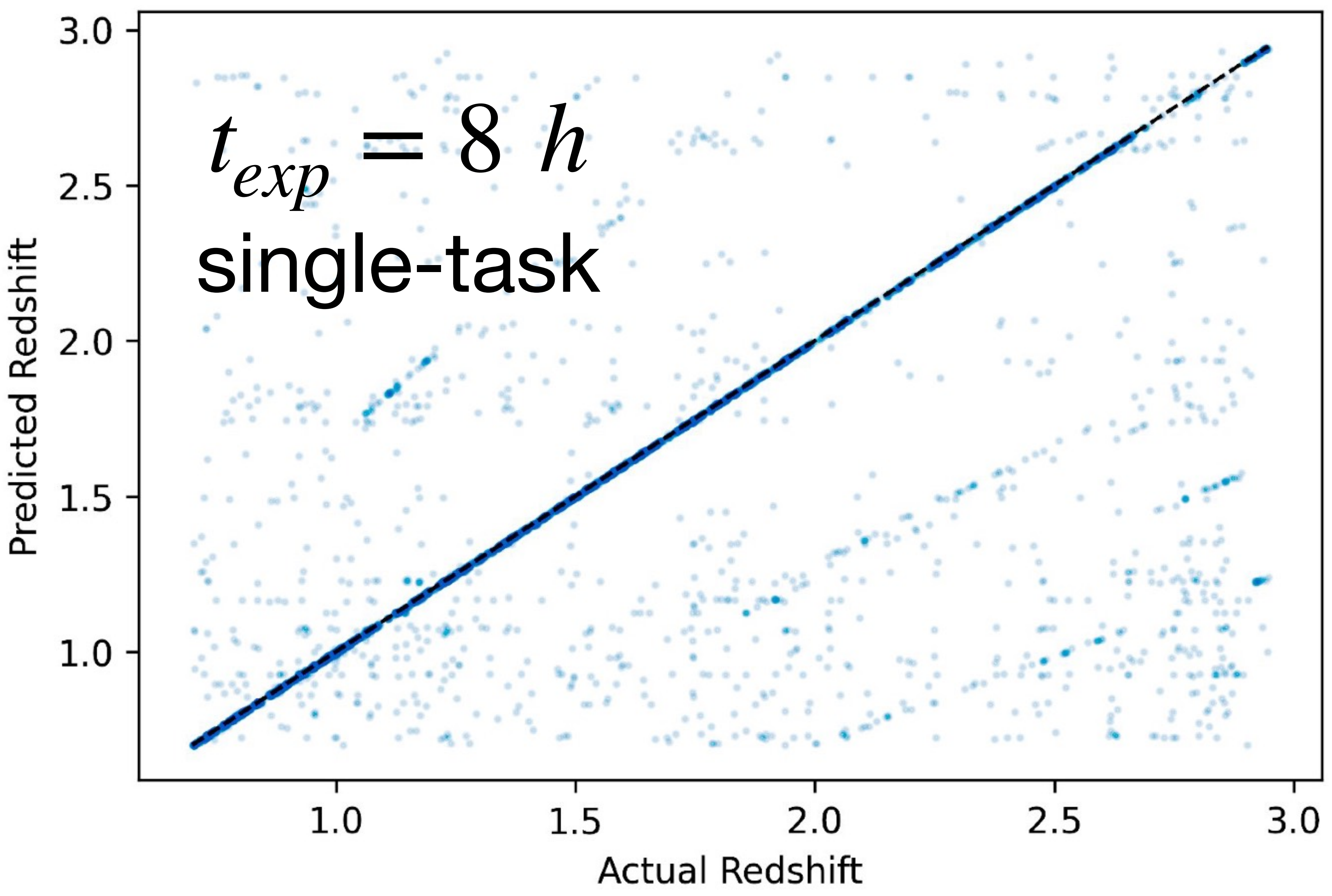
# General Performance



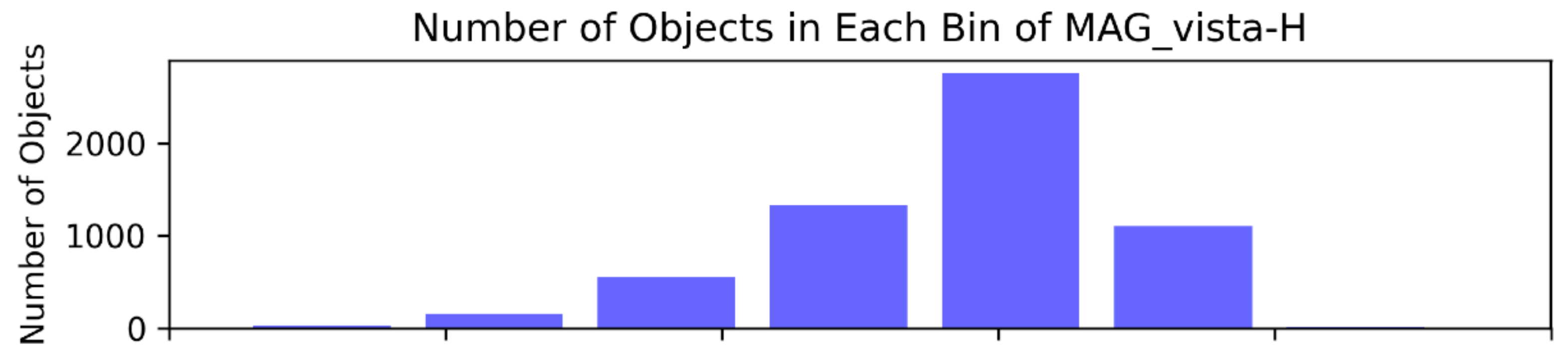



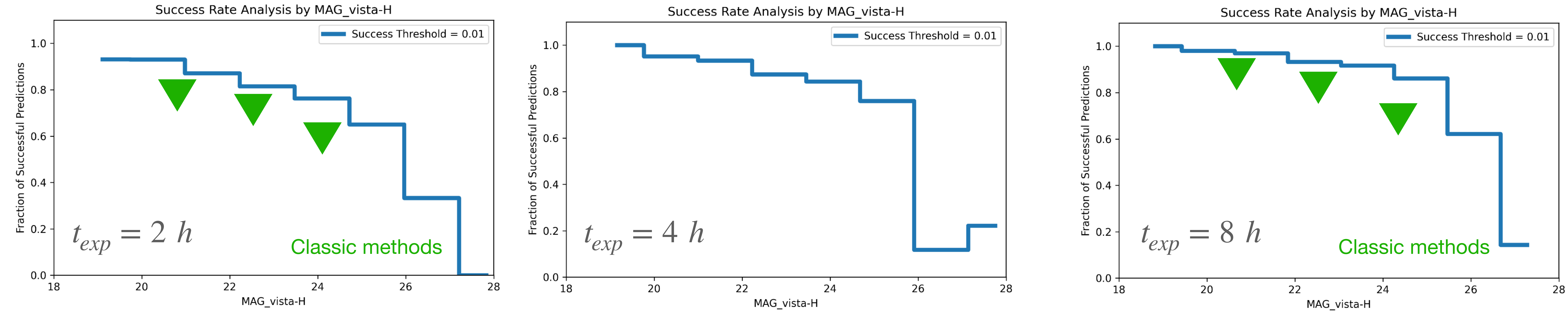



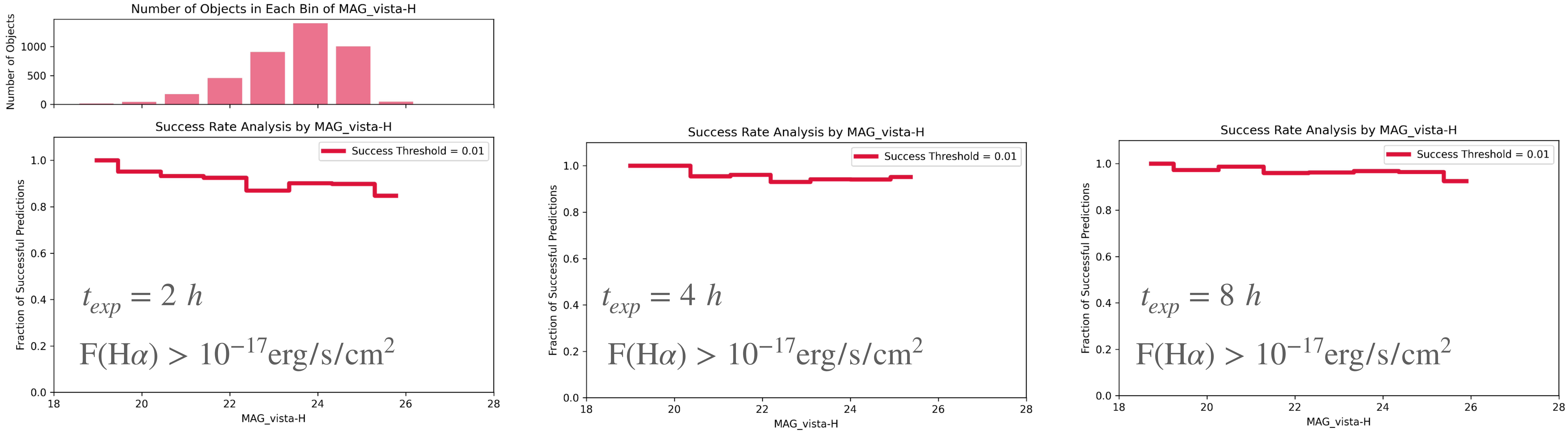



$$|\Delta z| < 0.01$$

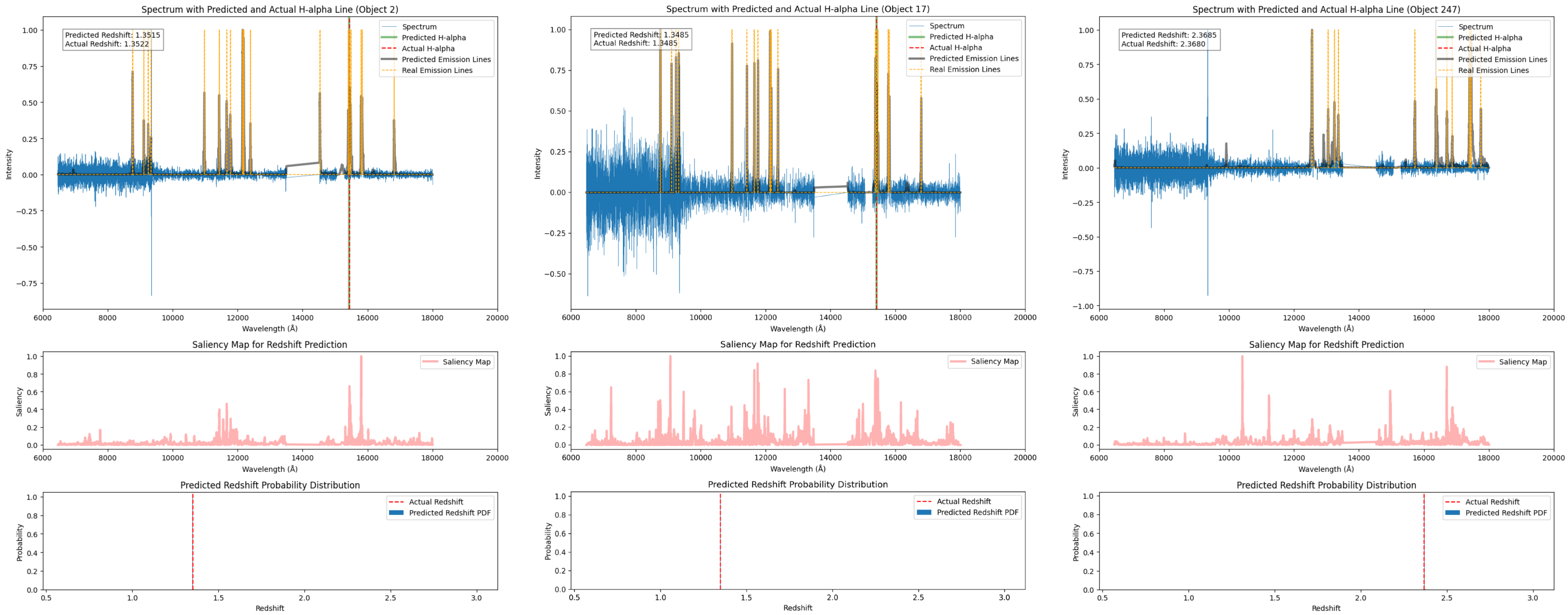

Number of Objects in Each Bin of MAG_vista-H
Number of Objects


Success Rate Analysis by MAG_vista-H
Success Threshold = 0.01
Fraction of Successful Predictions
MAG_vista-H
$t_{exp} = 2\ h$
Classic methods
$t_{exp} = 4\ h$
$t_{exp} = 8\ h$
Classic methods


Number of Objects in Each Bin of MAG_vista-H
Number of Objects
Success Rate Analysis by MAG_vista-H
Success Threshold = 0.01
Fraction of Successful Predictions
MAG_vista-H
$t_{exp} = 2\ h$
$F(H\alpha) > 10^{-17} erg/s/cm^2$
$t_{exp} = 4\ h$
$F(H\alpha) > 10^{-17} erg/s/cm^2$
$t_{exp} = 8\ h$
$F(H\alpha) > 10^{-17} erg/s/cm^2$

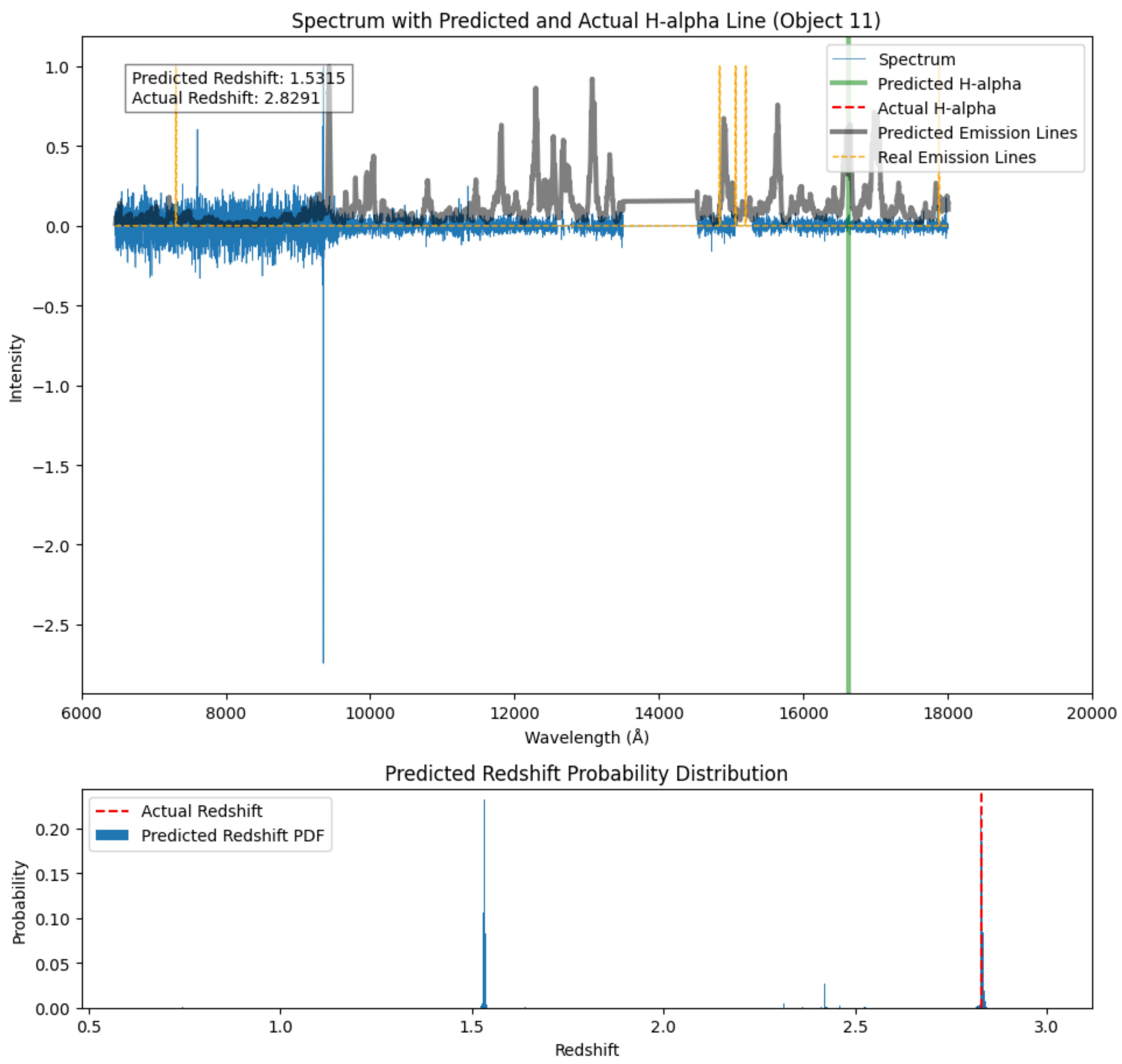
# Let's have a look at the predictions

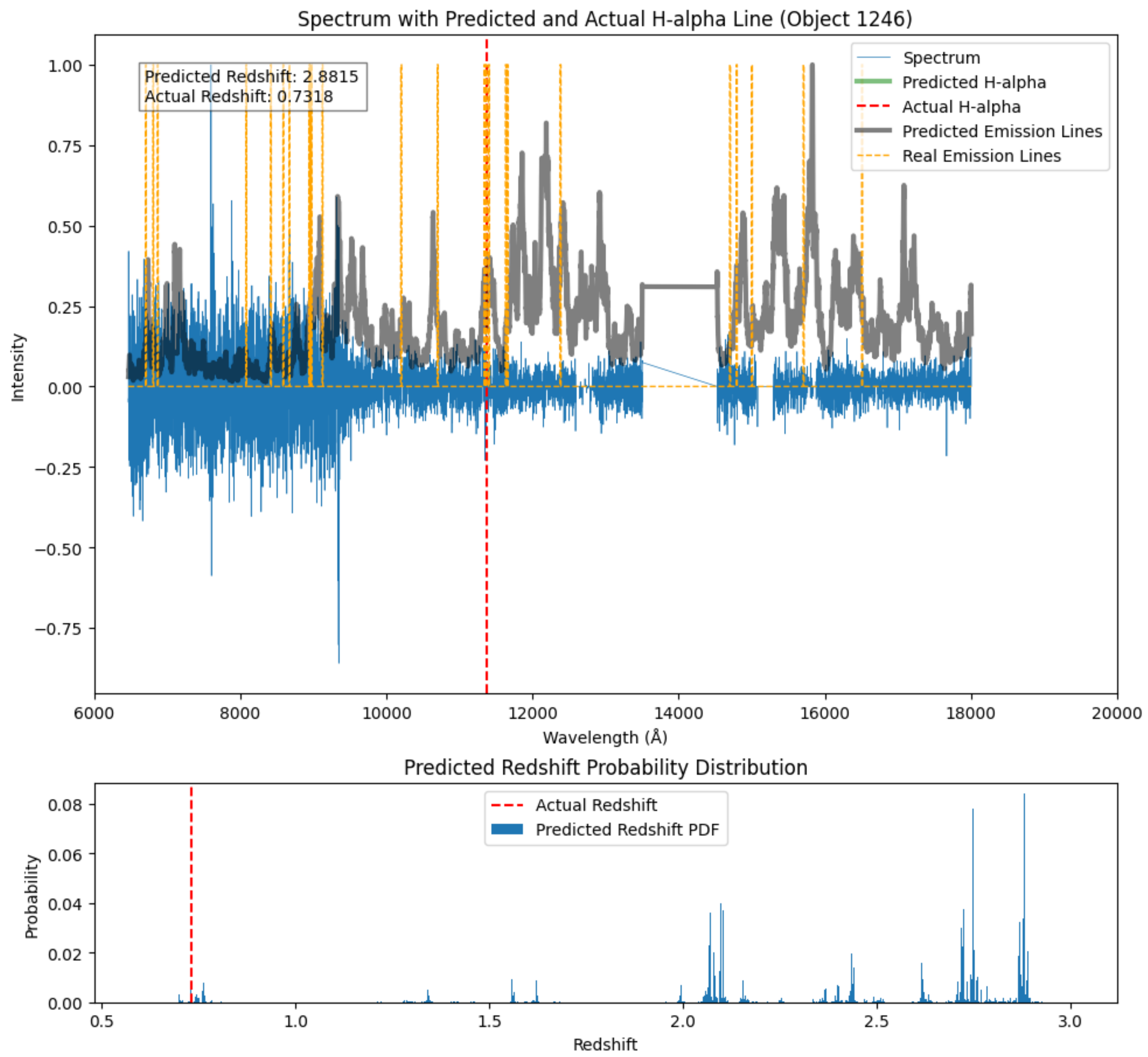
## Good predictions



# Let's have a look at the predictions

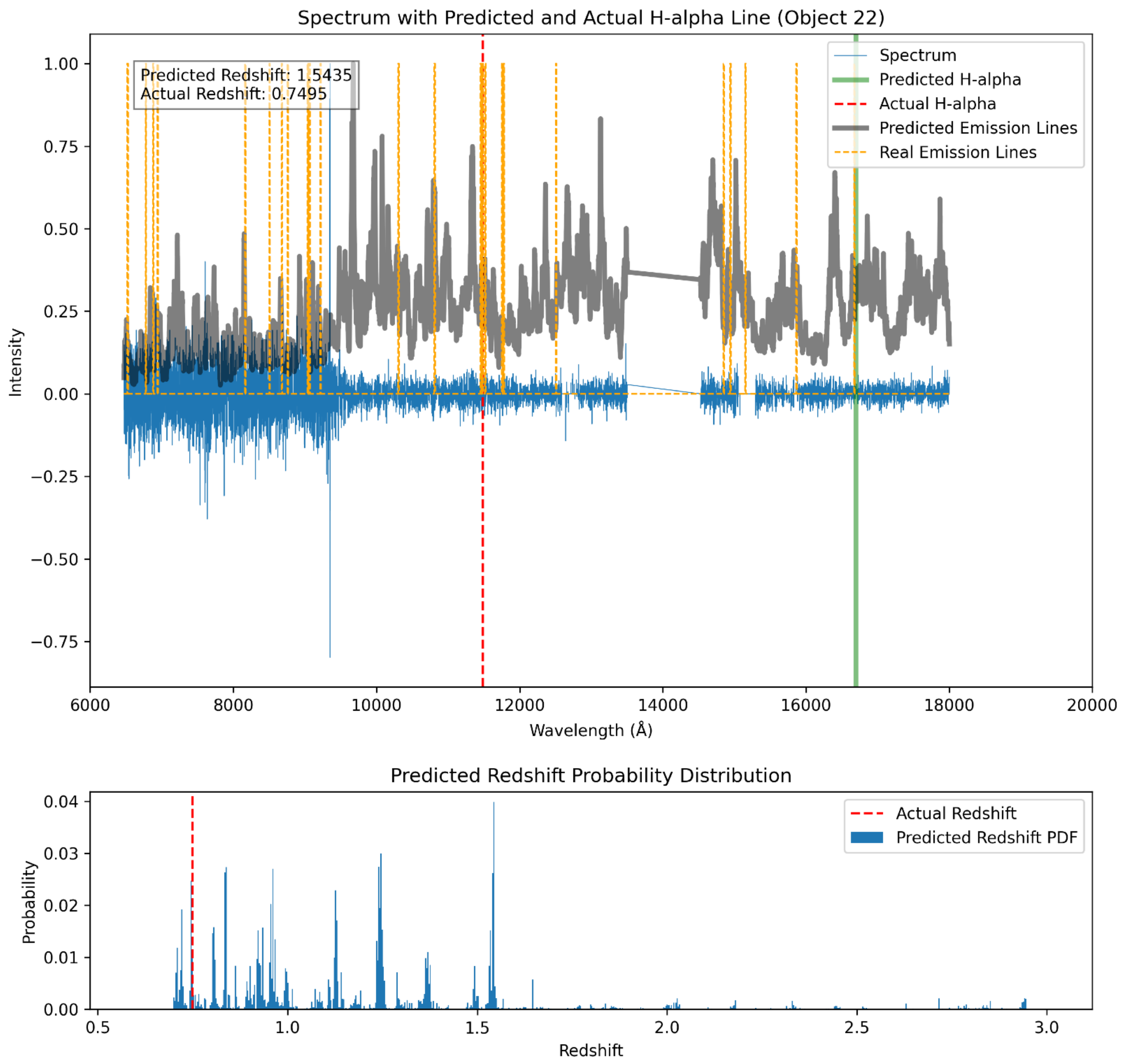
## Bad predictions



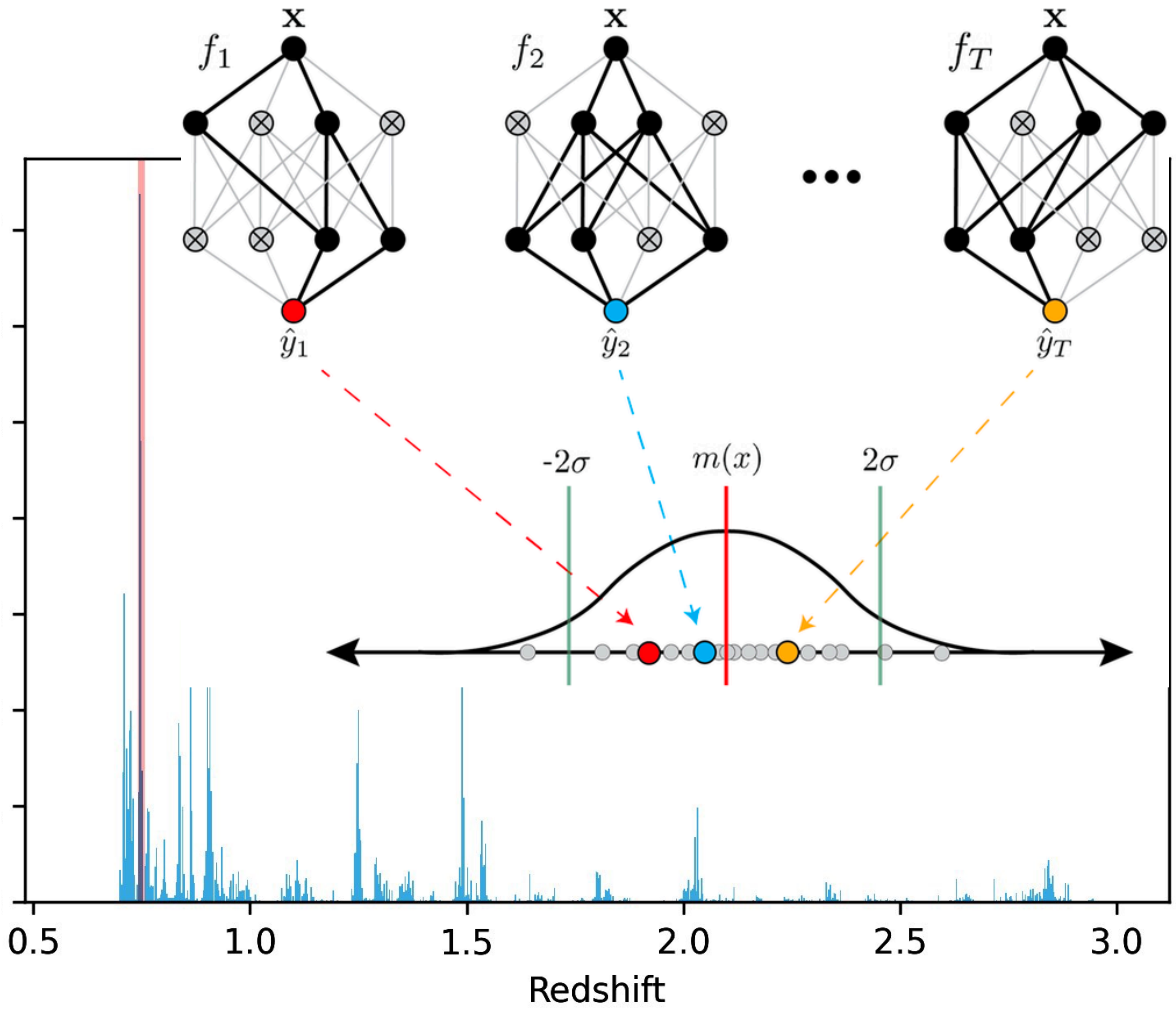



# Improve uncertainties with Monte Carlo Dropout

## MC Dropout at work

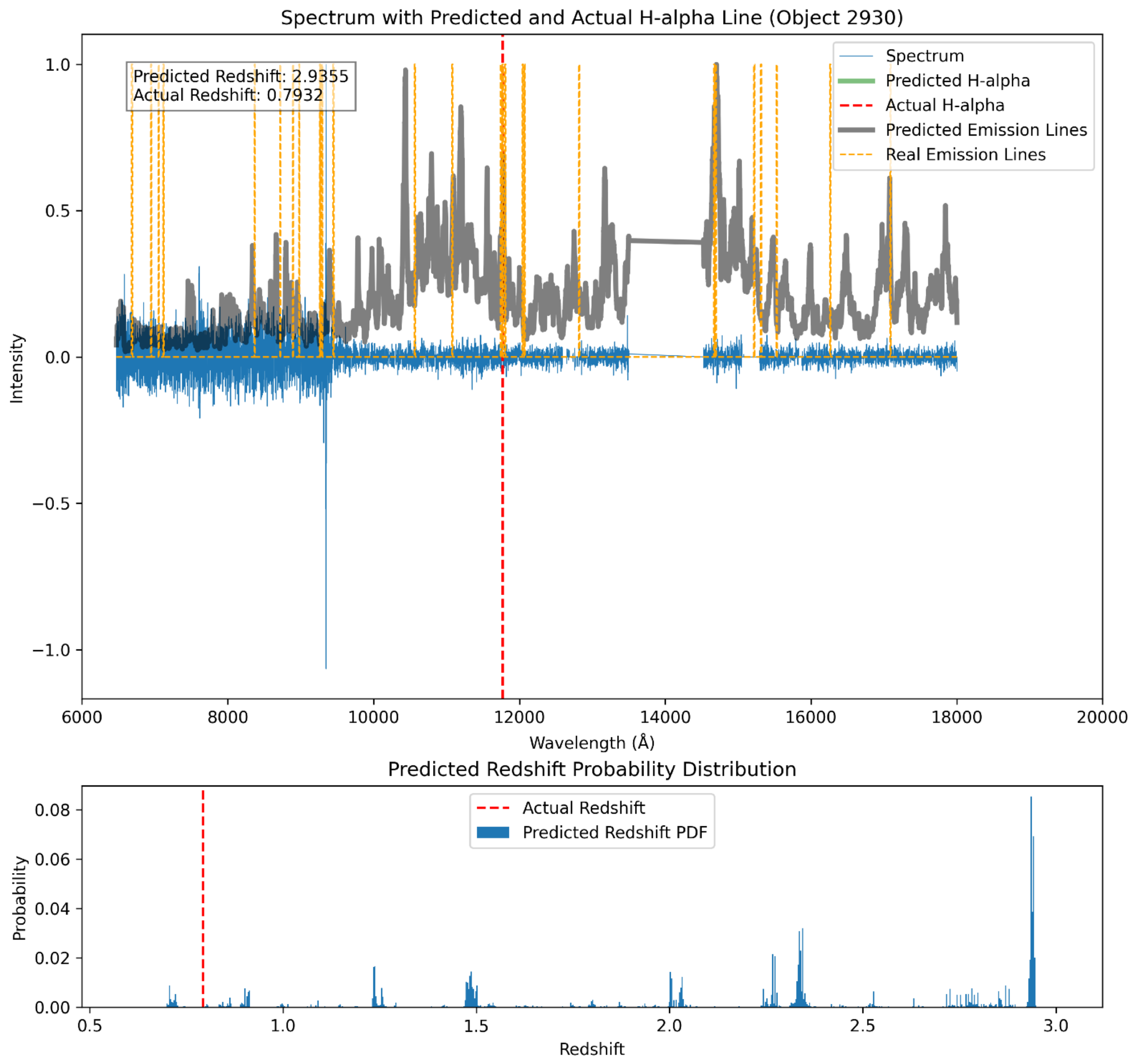


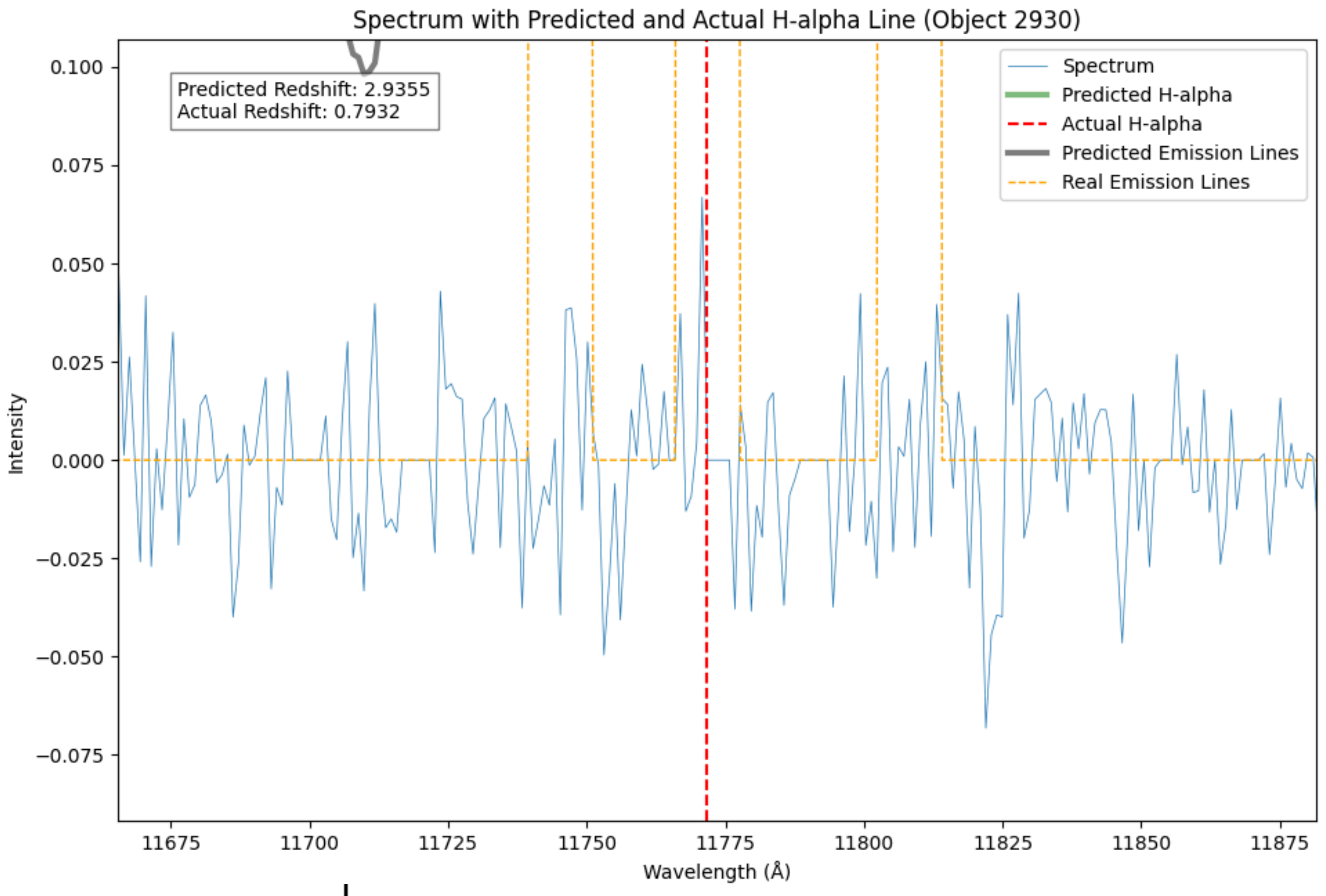

*Gal+15; Gal+16*

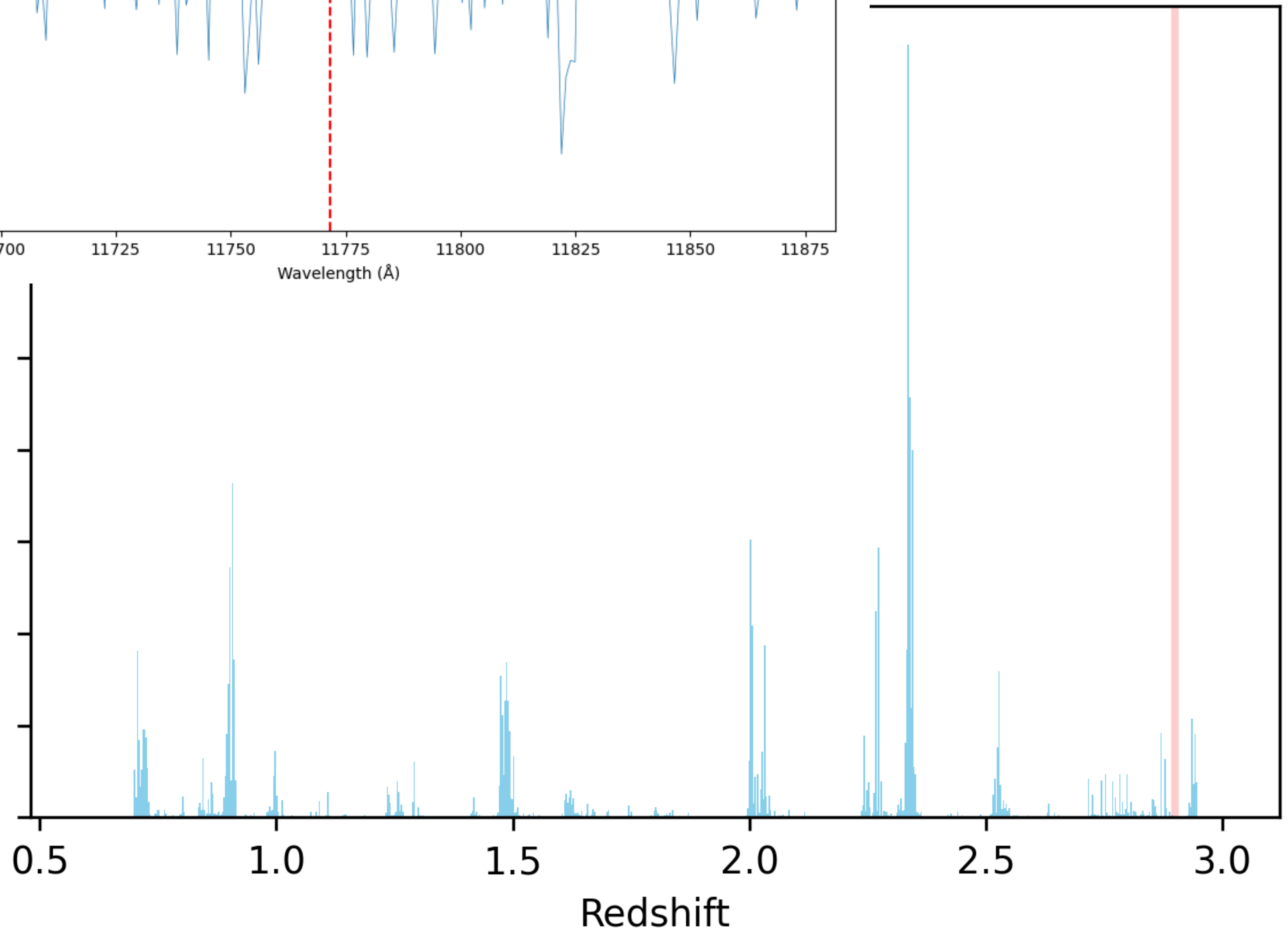


# Improve uncertainties with Monte Carlo Dropout

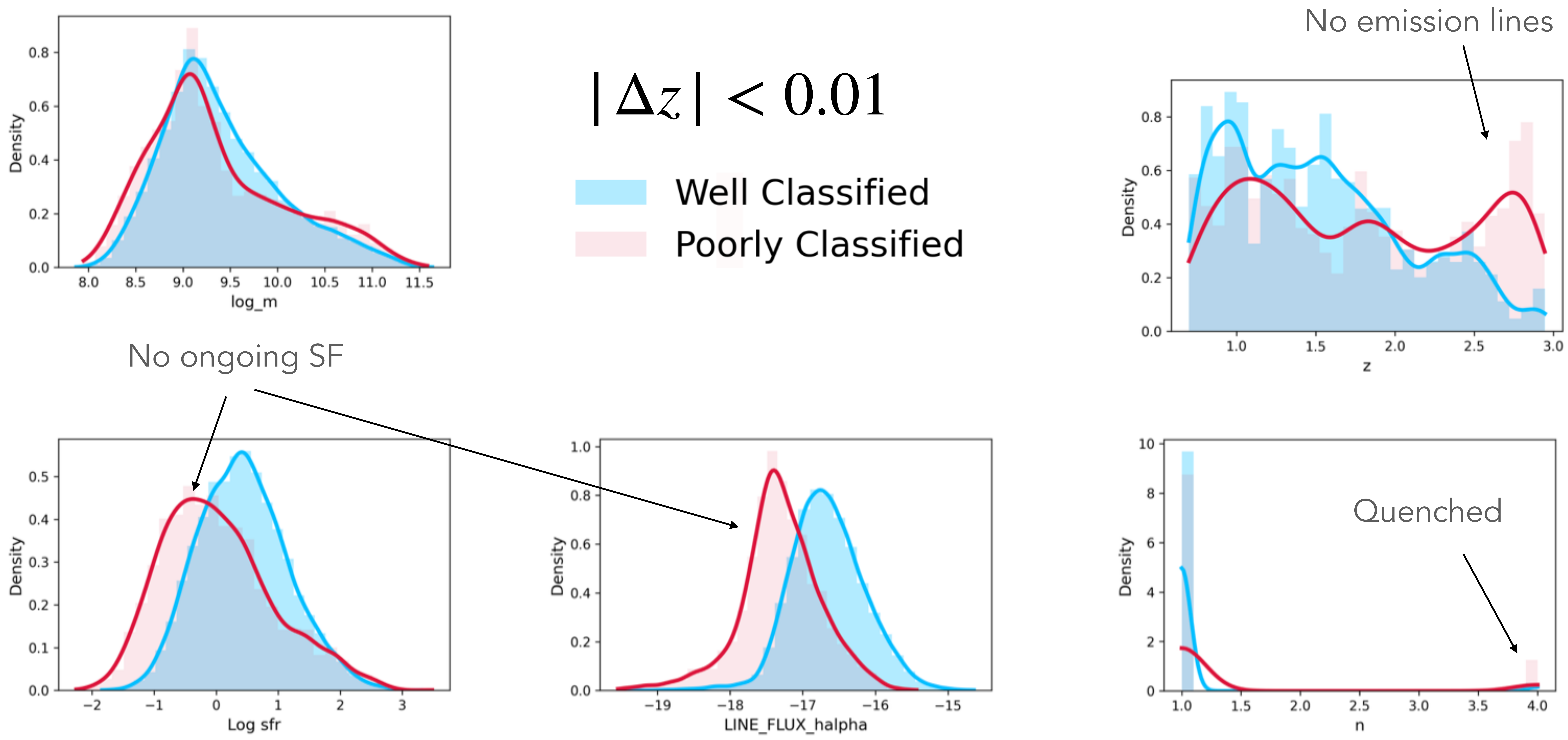
MC Dropout can't make it



# Differences in the distributions of good & bad predictions



$$|\Delta z| < 0.01$$

Well Classified

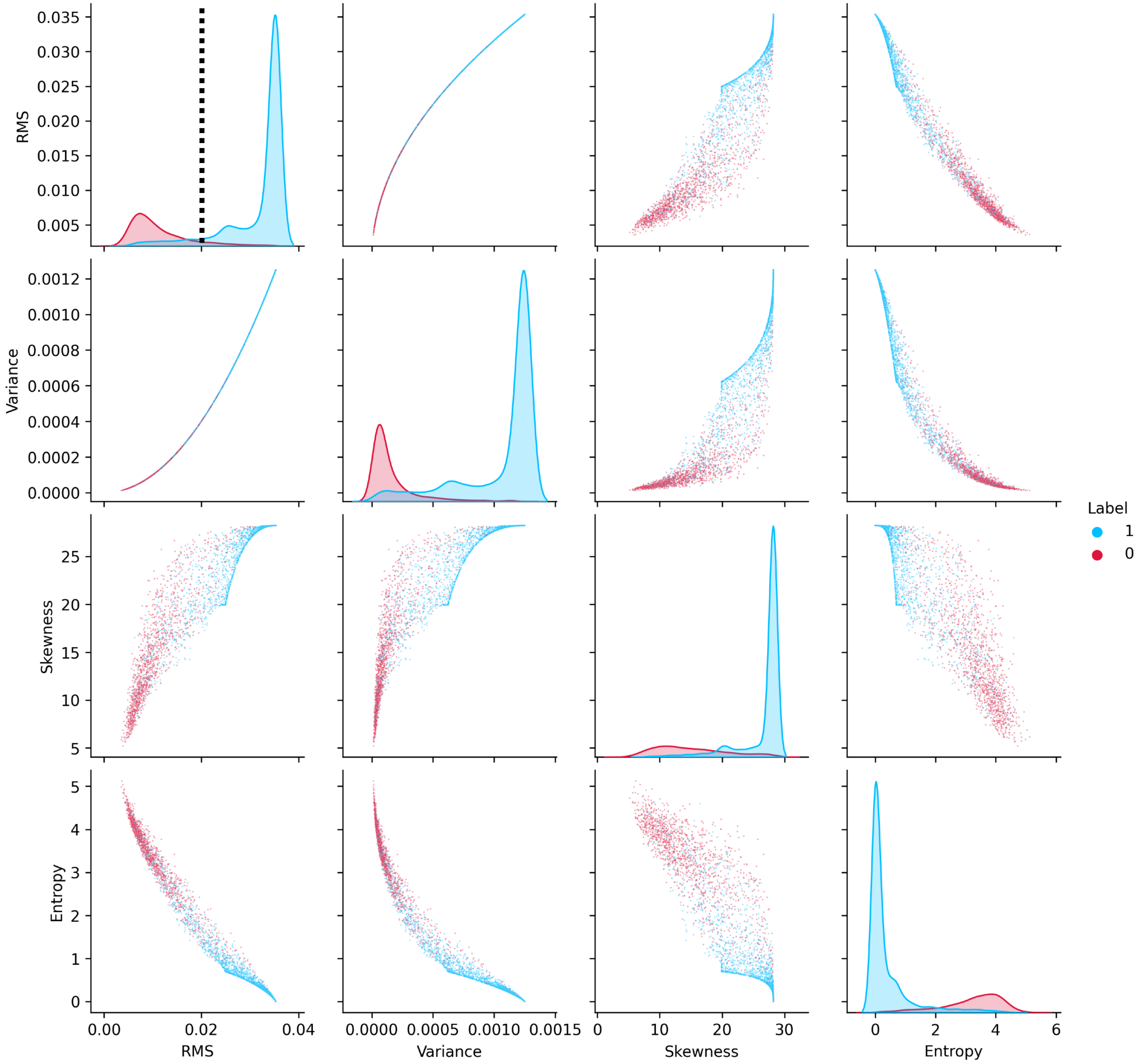
Poorly Classified

accuracy before filtering: 0.77
accuracy after filtering: 0.97

Filtered-out spectra ~27%

rms cut = 0.02

$$t_{exp} = 2\ h$$



$|\Delta z| < 0.01$

Well Classified

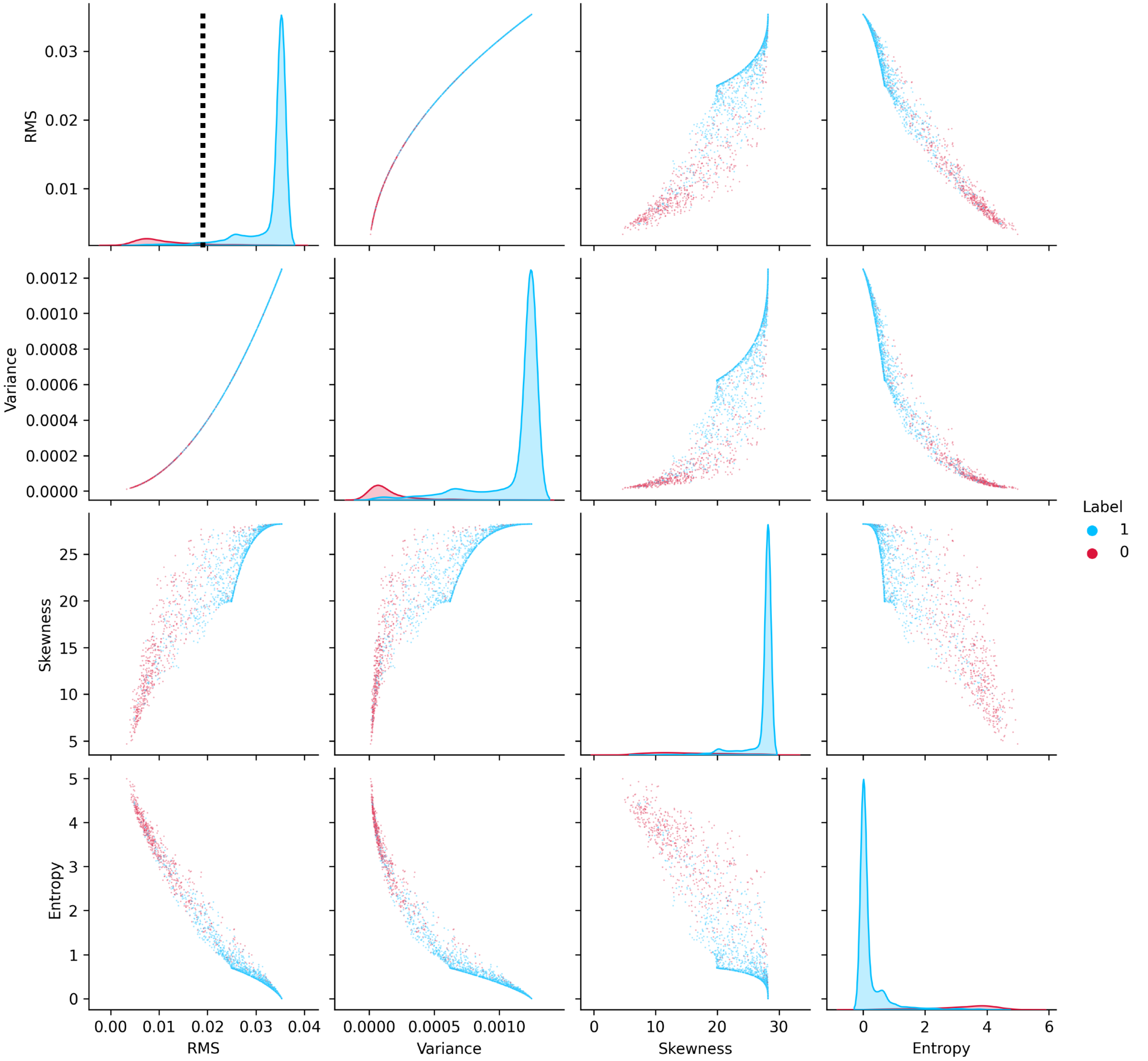
Poorly Classified

accuracy before filtering: 0.9
accuracy after filtering: 0.99

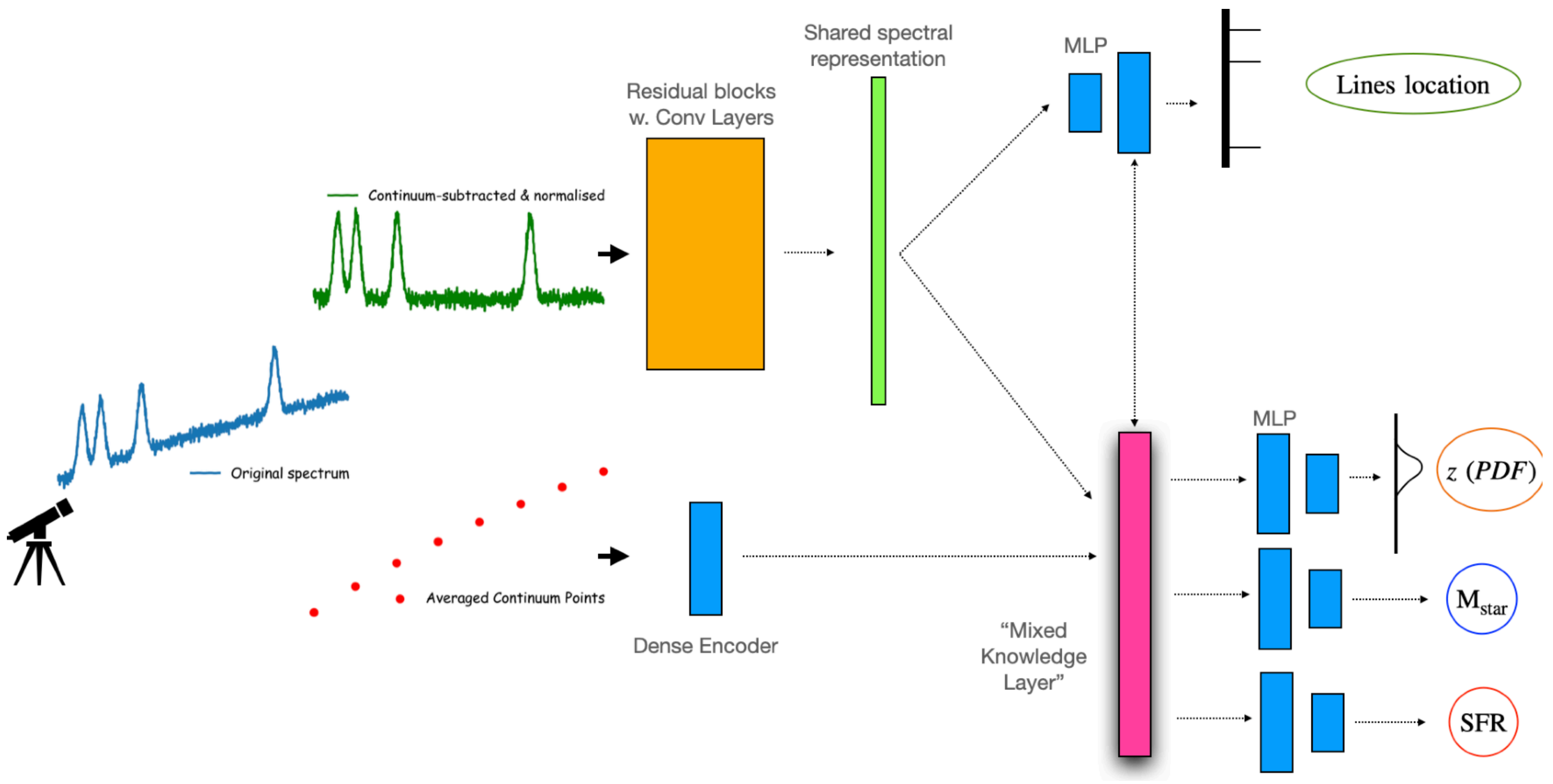
Filtered-out spectra ~12%

rms cut = 0.02

$$t_{exp} = 8\ h$$



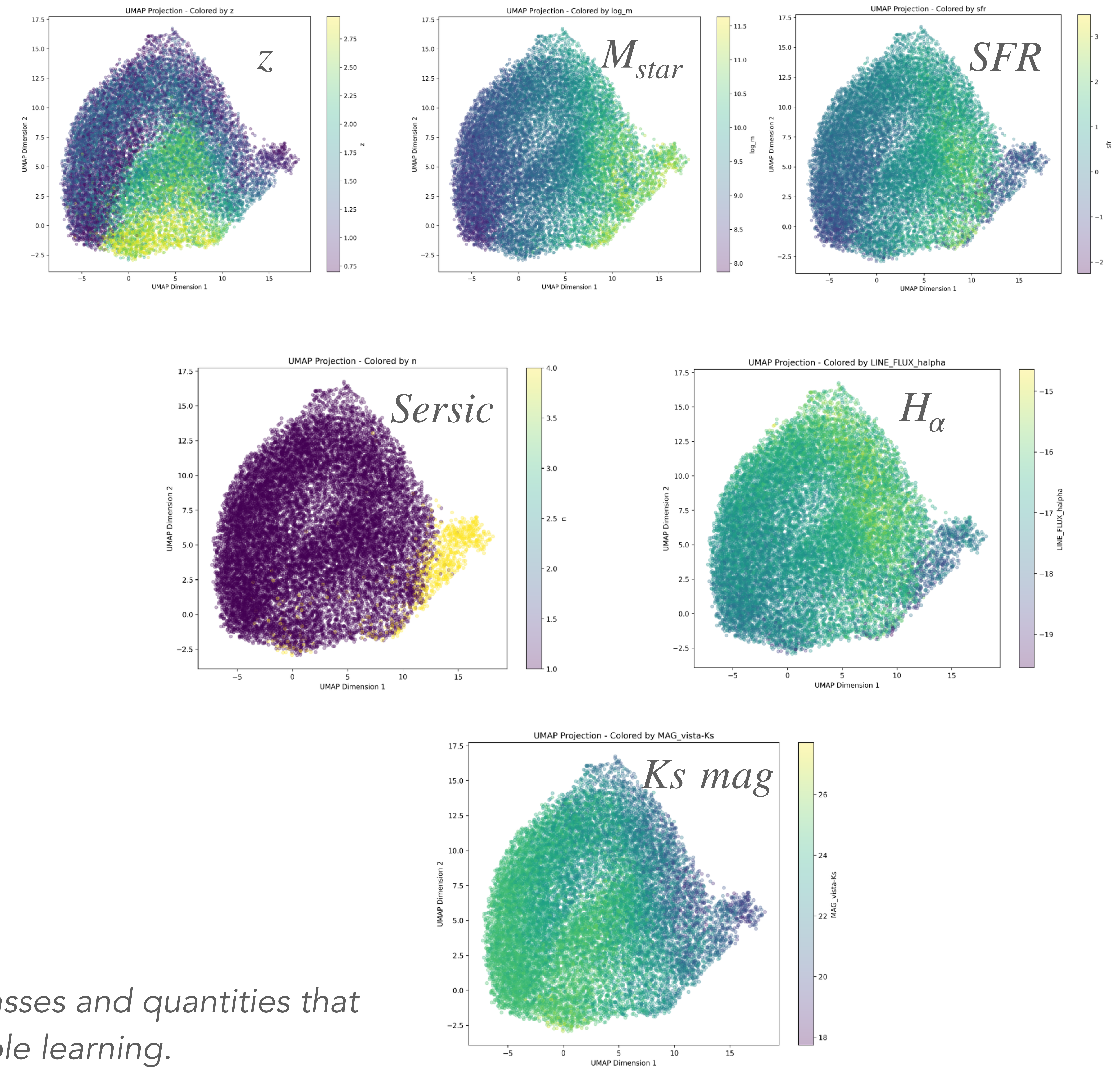
# Information encoded in the last embedding layers




~100 D space

UMAP dimensionality reduction

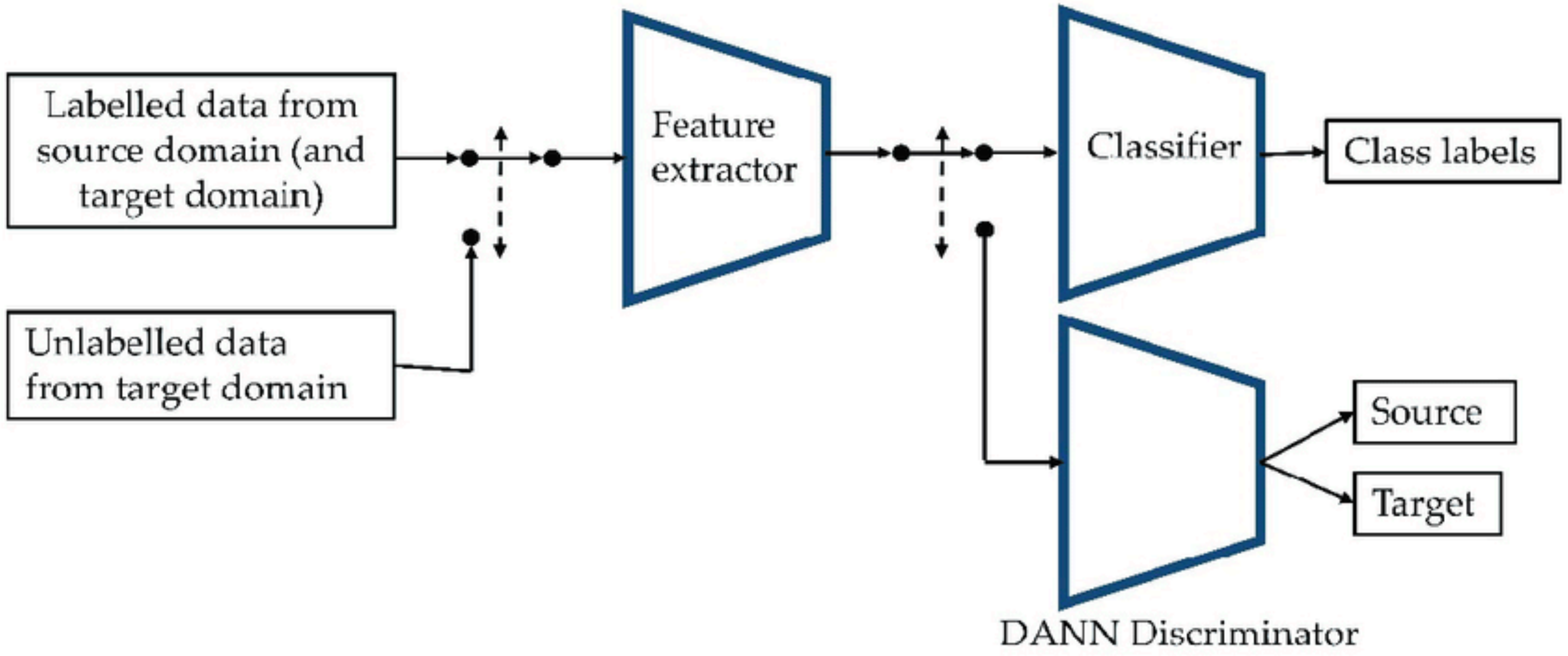
2-D projection

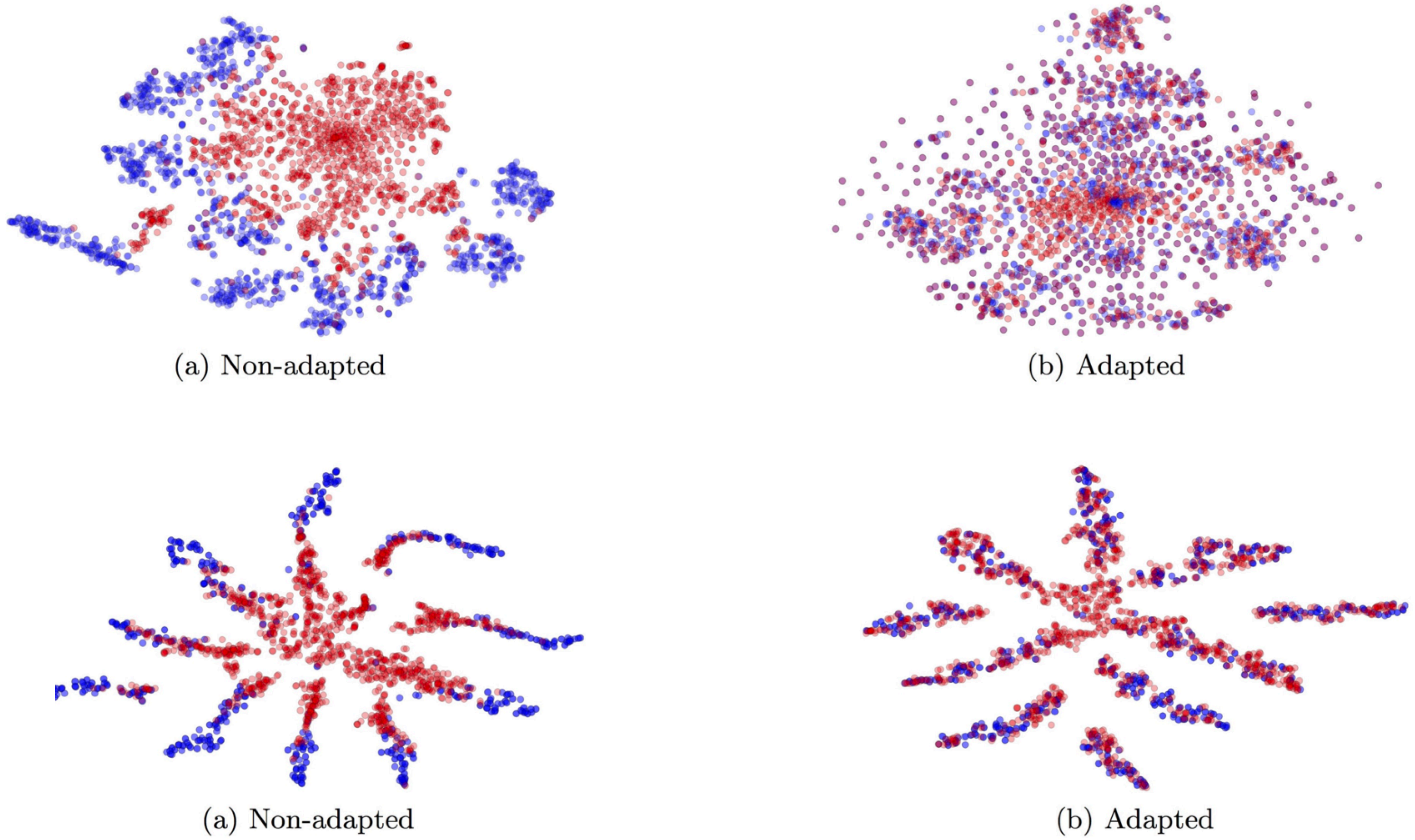



*The model encodes information that can be described through classes and quantities that “make sense” and the model has not seen in training — explainable learning.*

# Future — domain adversarial training to align real data

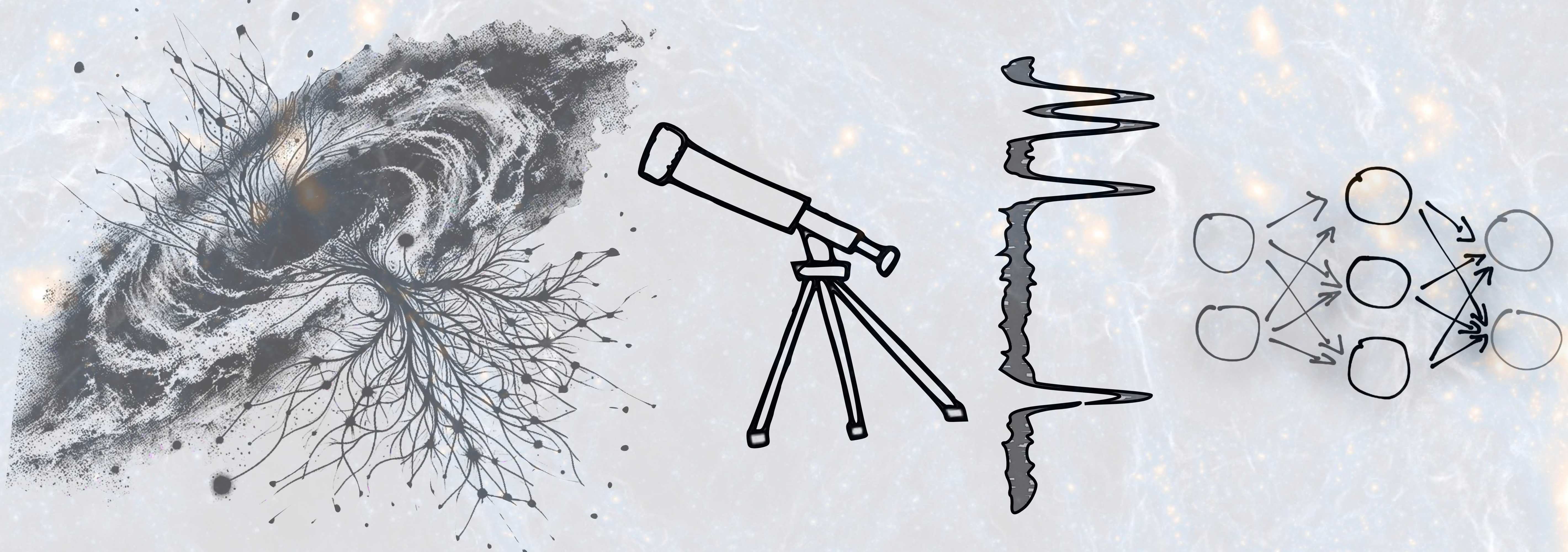
There might be a domain shift between synthetic data and real observations. Domain adaptation helps filling the gap.




Remember F. Belfiore's talk (yesterday)?

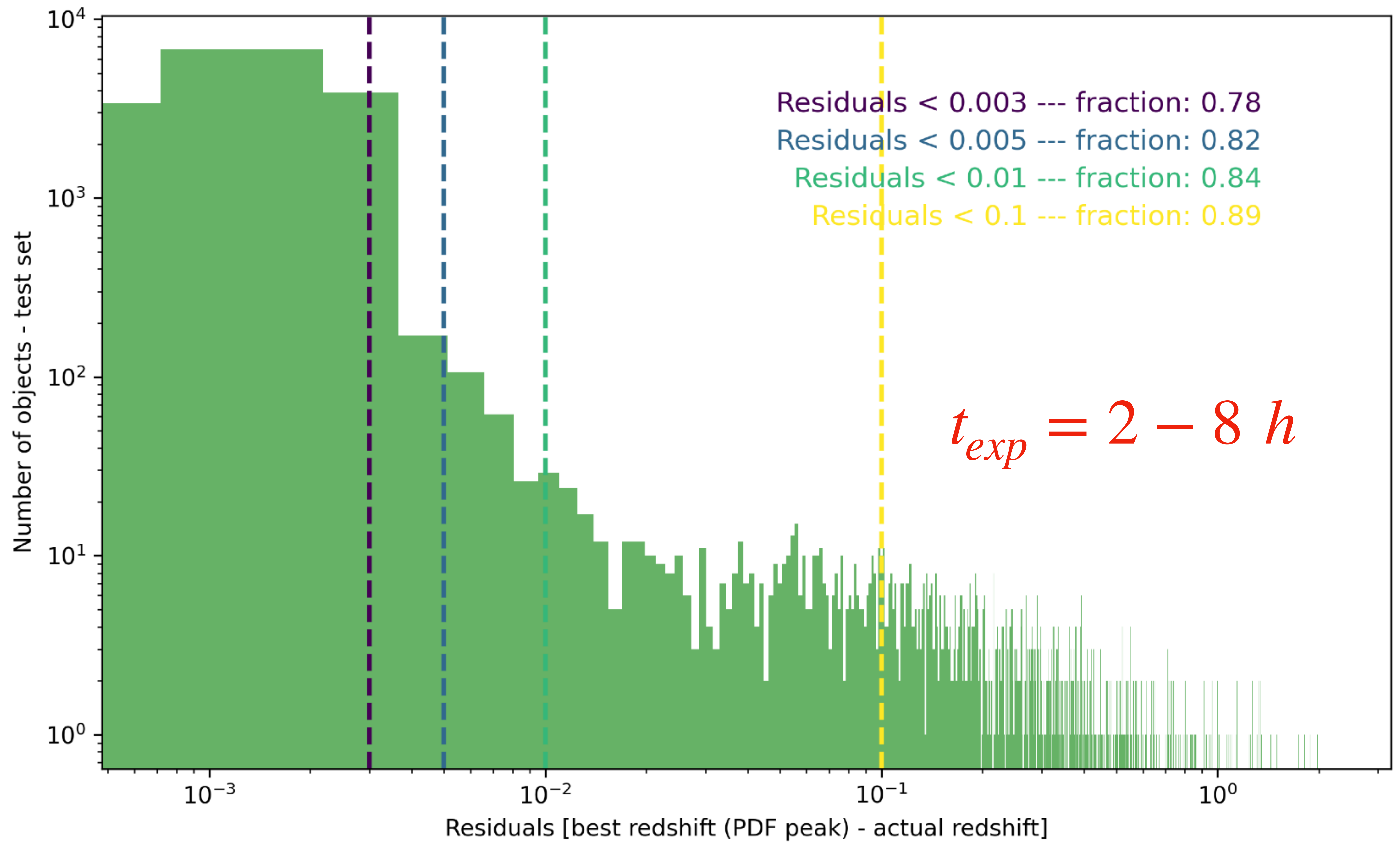
*Vilalta+18; Ćiprijanović+20; Huertas-Company+23; Belfiore & Ginolfi, in prep*

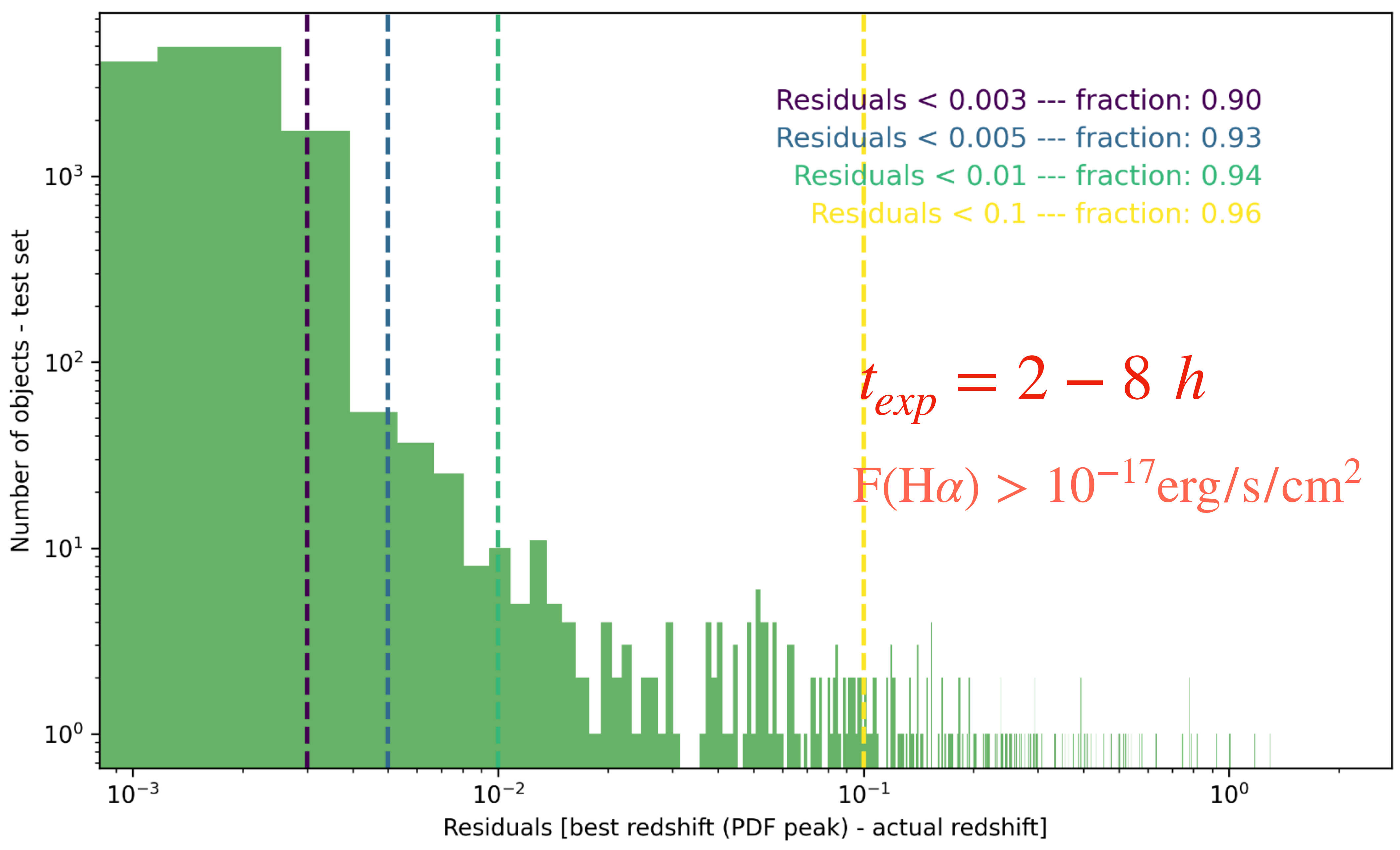


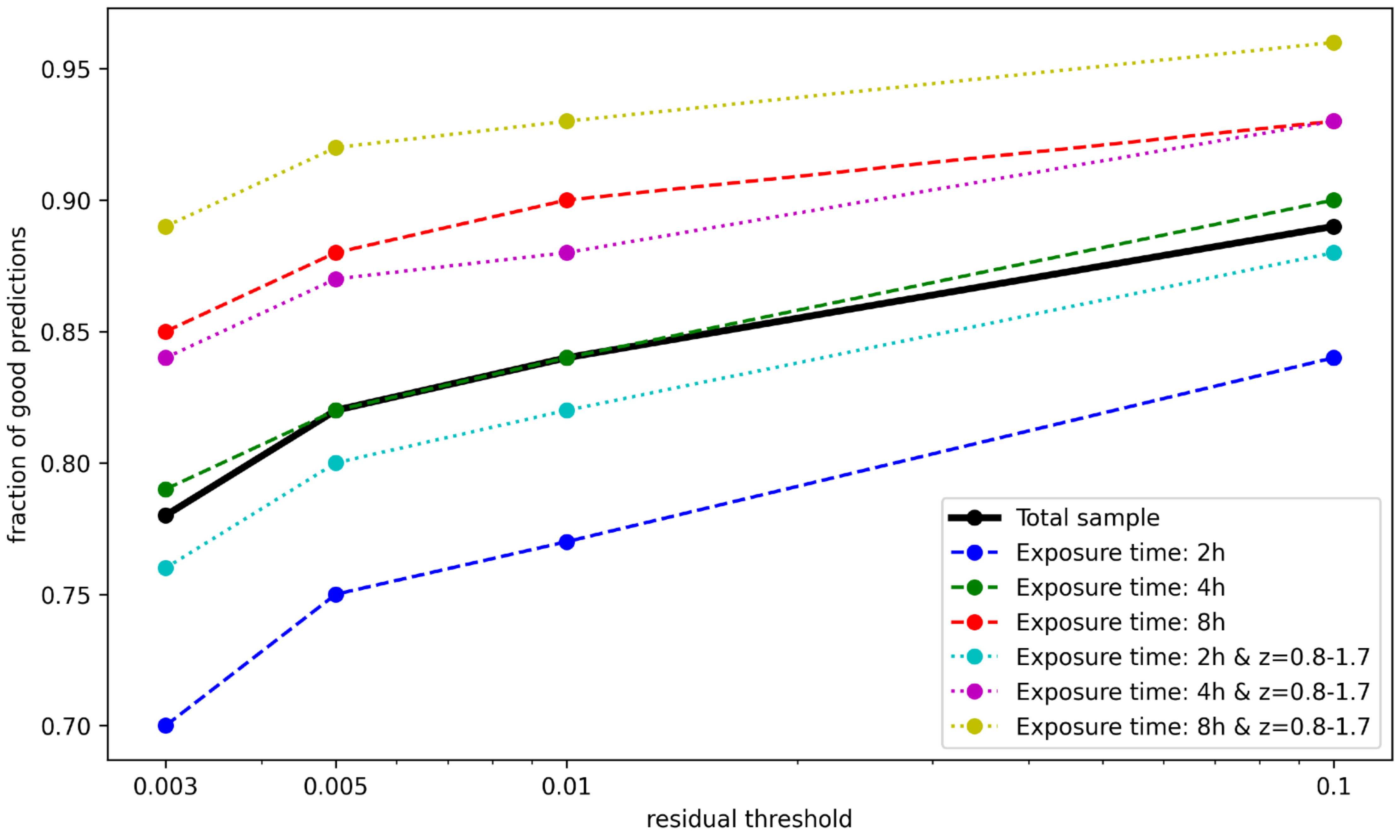
# Conclusions

- The upcoming volume and complexity of spectral data, especially around cosmic noon, call for the help of deep learning.
- We designed a conv neural net trained through multi-task learning that can accurately obtain redshift and physical properties from galaxy spectra, handling uncertainties.
- We tested our pipeline on simulated MOONS spectra, outperforming standard spectral fitting tools. Our results will help in designing observational strategies for MOONS.
- We find that an a-posteriori analysis of the redshift PDF boosts performance through straightforward contaminant removal.

# Additional slides


Residuals < 0.003 --- fraction: 0.78
Residuals < 0.005 --- fraction: 0.82
Residuals < 0.01 --- fraction: 0.84
Residuals < 0.1 --- fraction: 0.89
$t_{exp} = 2 - 8\ h$
Number of objects - test set
Residuals [best redshift (PDF peak) - actual redshift]


Residuals < 0.003 --- fraction: 0.90
Residuals < 0.005 --- fraction: 0.93
Residuals < 0.01 --- fraction: 0.94
Residuals < 0.1 --- fraction: 0.96
$t_{exp} = 2 - 8\ h$
$F(H\alpha) > 10^{-17} erg/s/cm^2$
Number of objects - test set
Residuals [best redshift (PDF peak) - actual redshift]


Total sample
Exposure time: 2h
Exposure time: 4h
Exposure time: 8h
Exposure time: 2h & z=0.8-1.7
Exposure time: 4h & z=0.8-1.7
Exposure time: 8h & z=0.8-1.7
fraction of good predictions
residual threshold