Universidad de La Laguna

XIII Estallidos Workshop

Classifying rest-frame low-z SEDs in the PAU survey using Machine Learning clustering

Ana Luisa González Morán

Instituto de Astrofísica de Canarias (IAC)
Postdoctoral researcher

ana.gonzalez@iac.es

Collaboration group:

Pablo Arrabal Haro, NOIRLab-US
Casiana Muñoz Tuñón, IAC-Spain
Jose Miguel Rodríguez Espinosa, IAC-Spain
Jorge Sánchez Almeida, IAC-Spain
PAU Collaboration

PAU

Machine learning clustering of high-z galaxy SEDs in GOODS-N and COSMOS

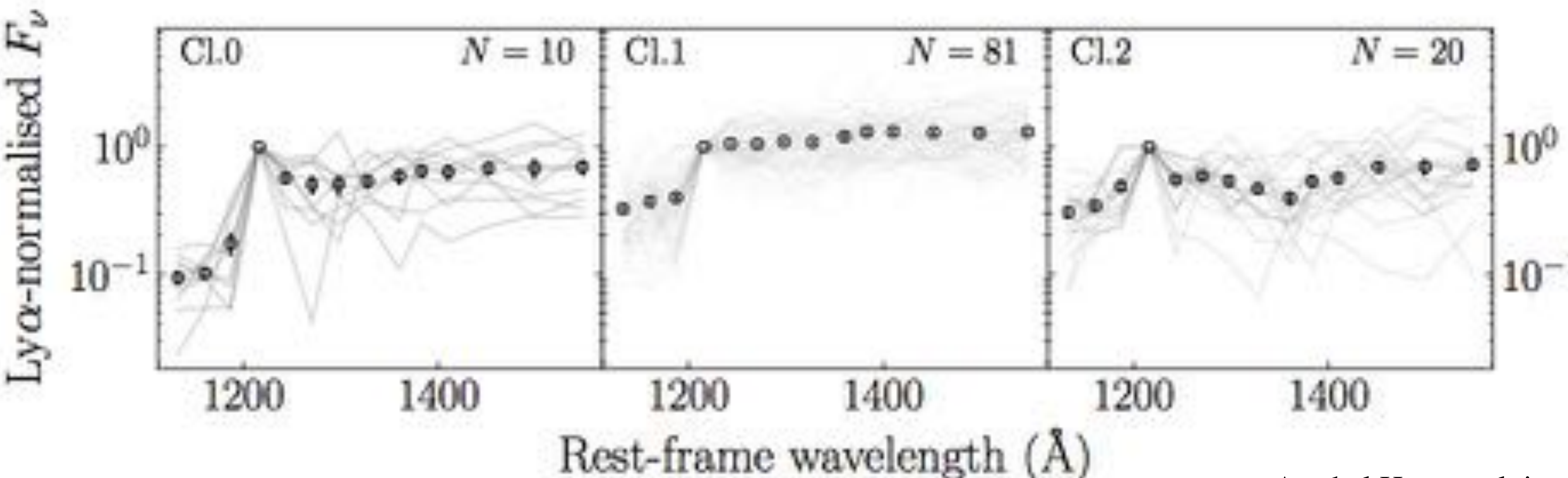
High-z SHARDS sample

1558 galaxies at 3.4 < z < 6.8 (528 LAEs and 1030 no-Ly LBGs)

SC4K-COSMOS sample

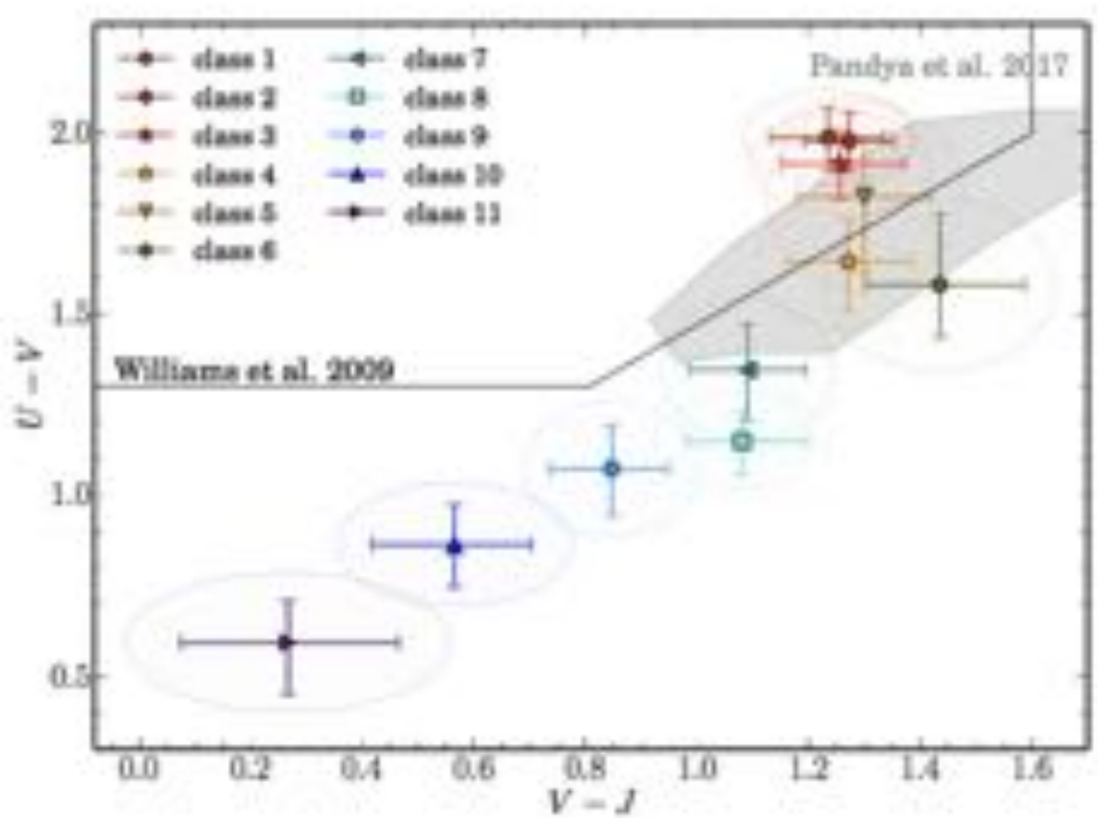
3908 LAEs at 2 < z < 6

z ~ 5 sample from SHARDS

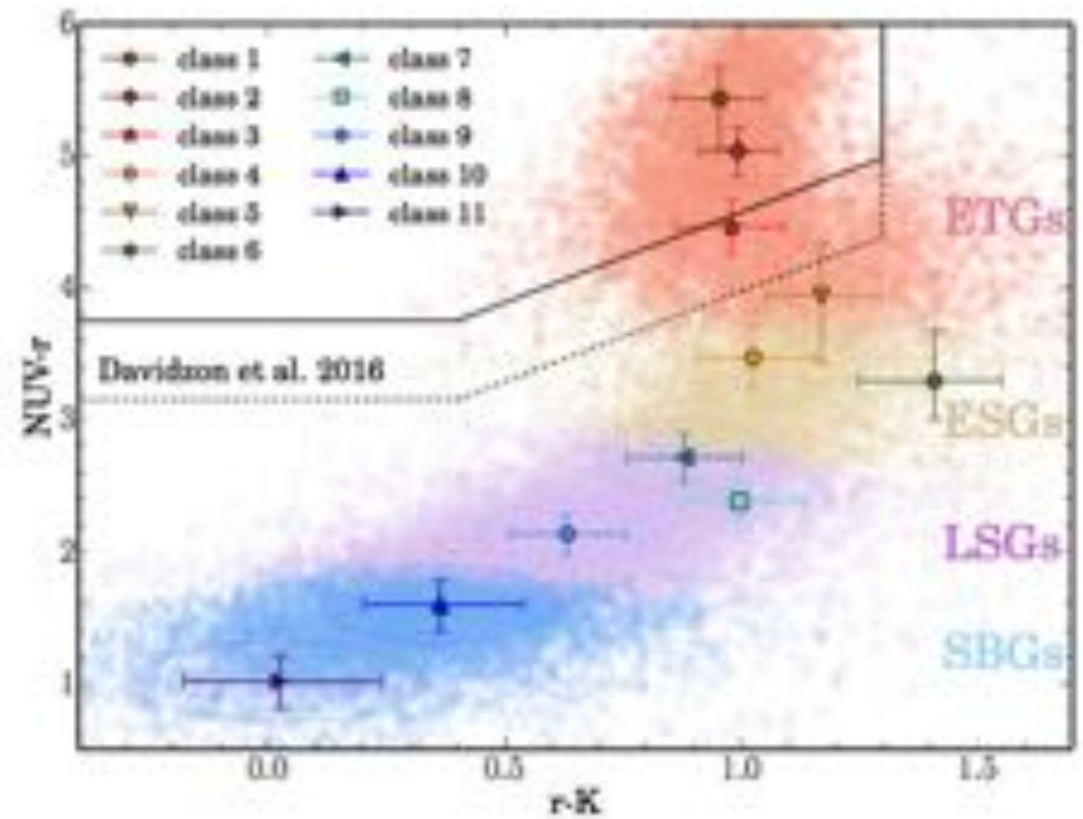


Arrabal Haro et al. in prep

The VIMOS Public Extragalactic Redshift Survey (VIPERS)

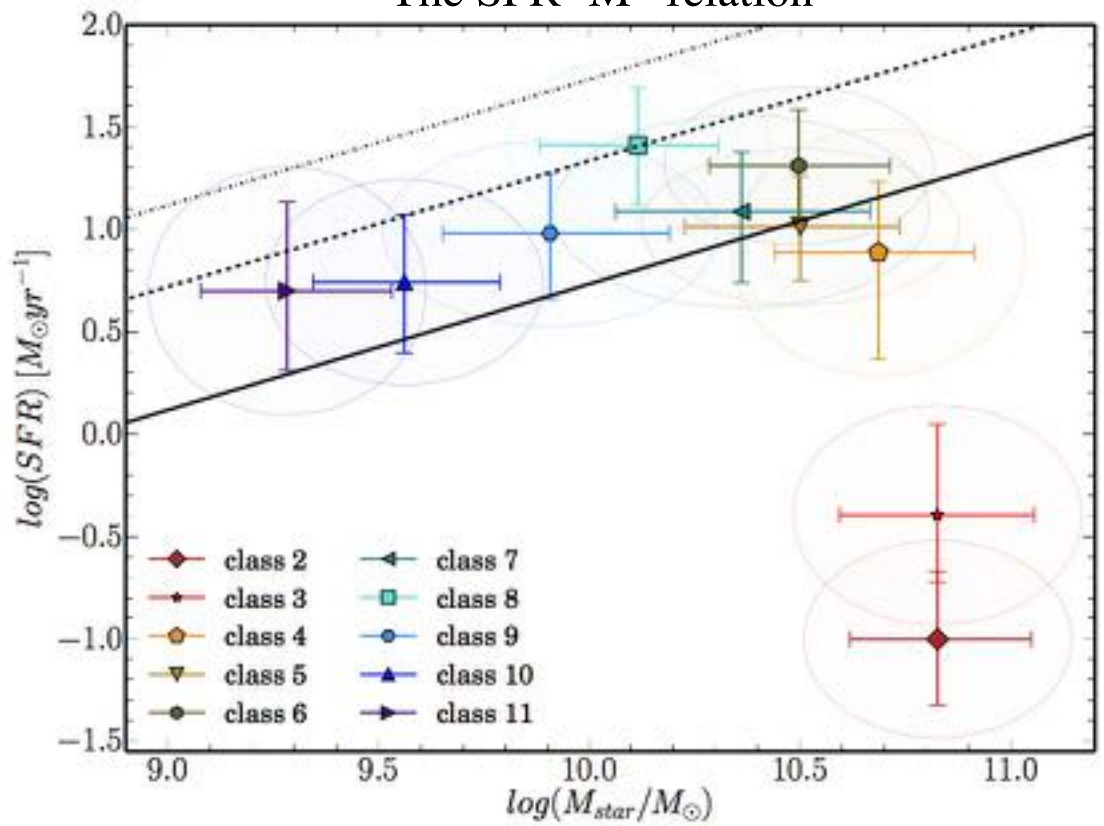


(a) *UVJ* diagram;



(b) *NUVrK* diagram;

The SFR−M* relation



Siudek et al. 2018

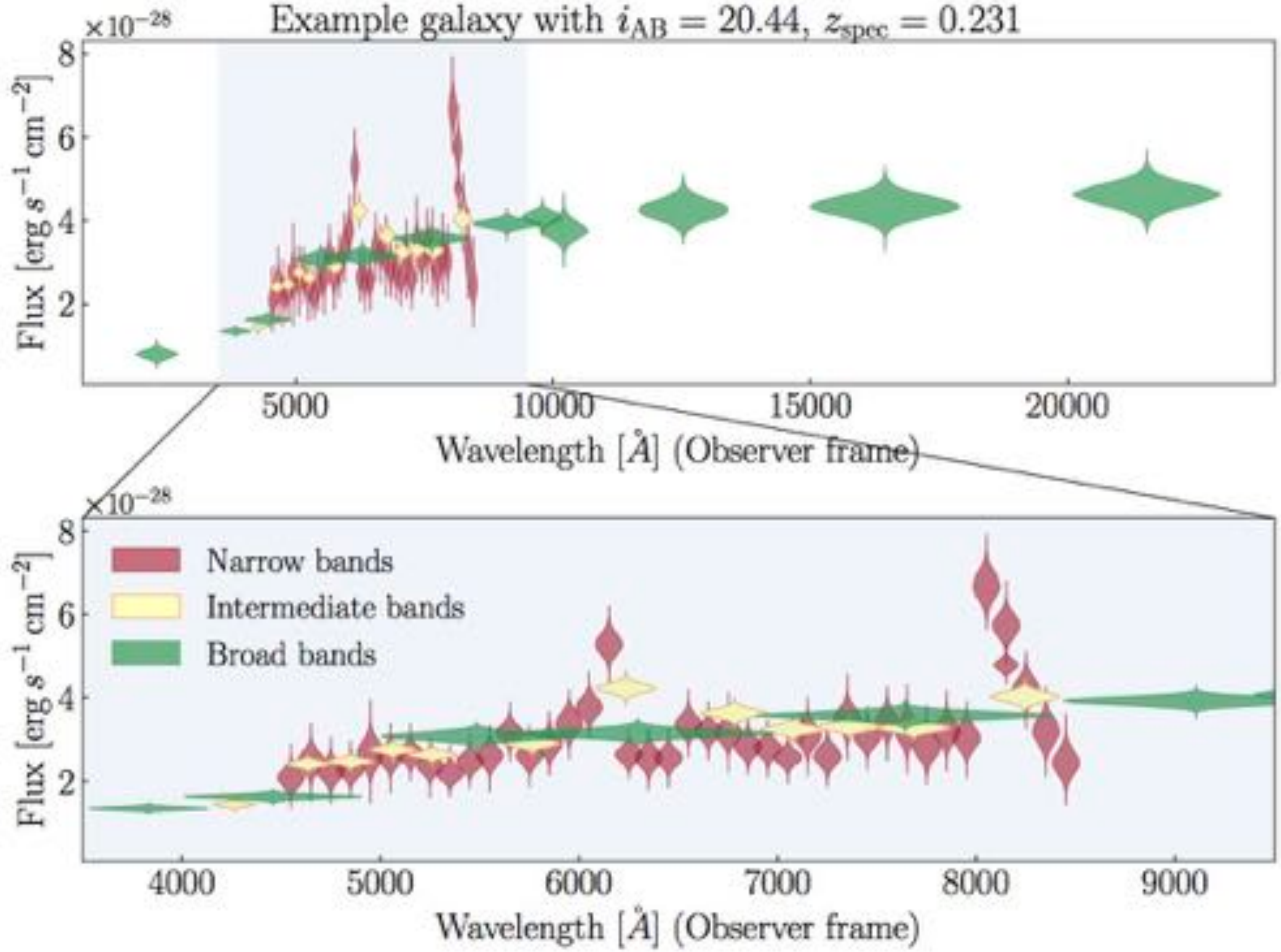
Main objective

To perform a characterization of galaxies using an unsupervised Machine Learning clustering on Spectral Energy Distributions (SEDs) from the PAU Survey in the COSMOS field.

PAU Survey

It uses a camera with a system of 40 narrow-band (of 10 nm) filters spanning the wavelength range from 450nm to 850nm.

WHT



Alarcon et al. 2020

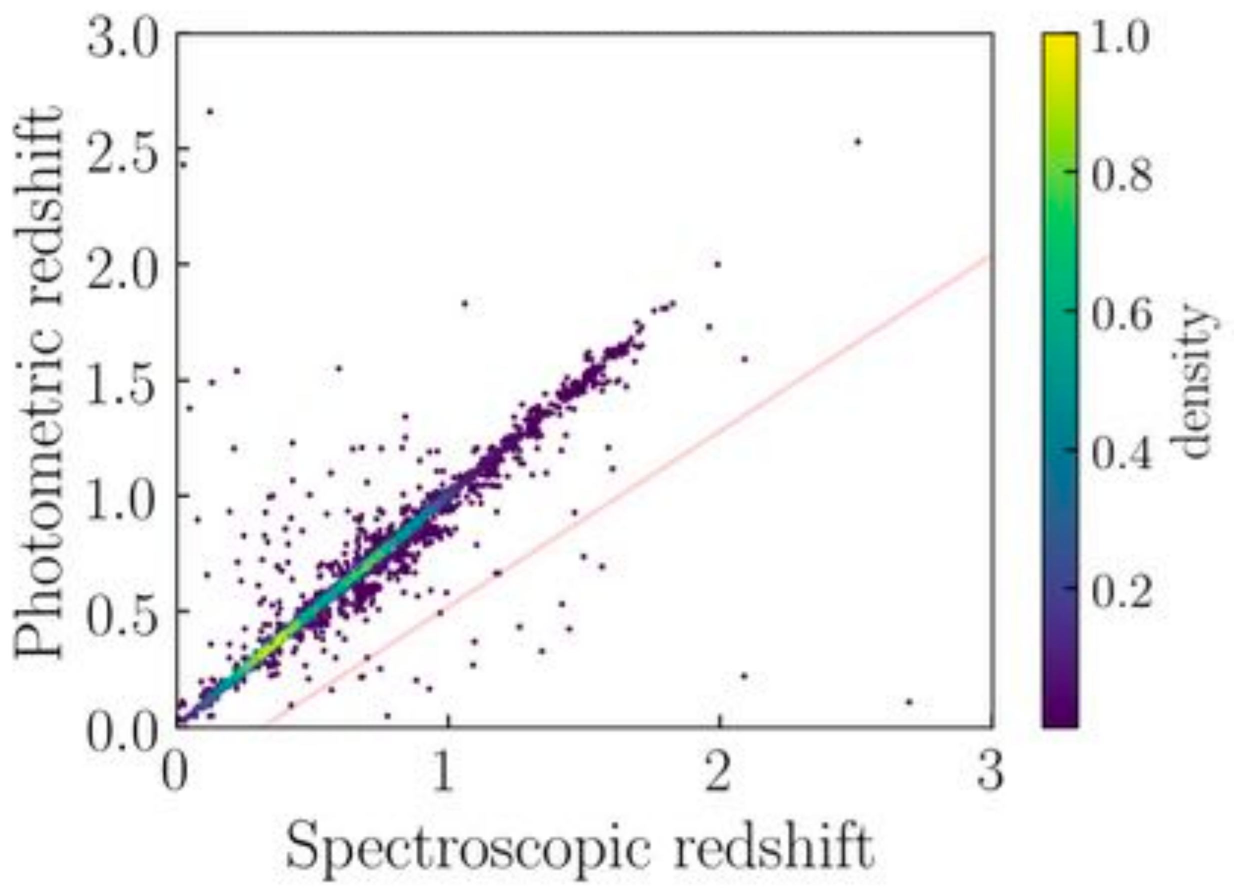
- Accurate and precise photometric redshifts.

σ68(Δz) ≈ (0.003,0.009) at iAB ~ 18 and iAB ~ 23, respectively

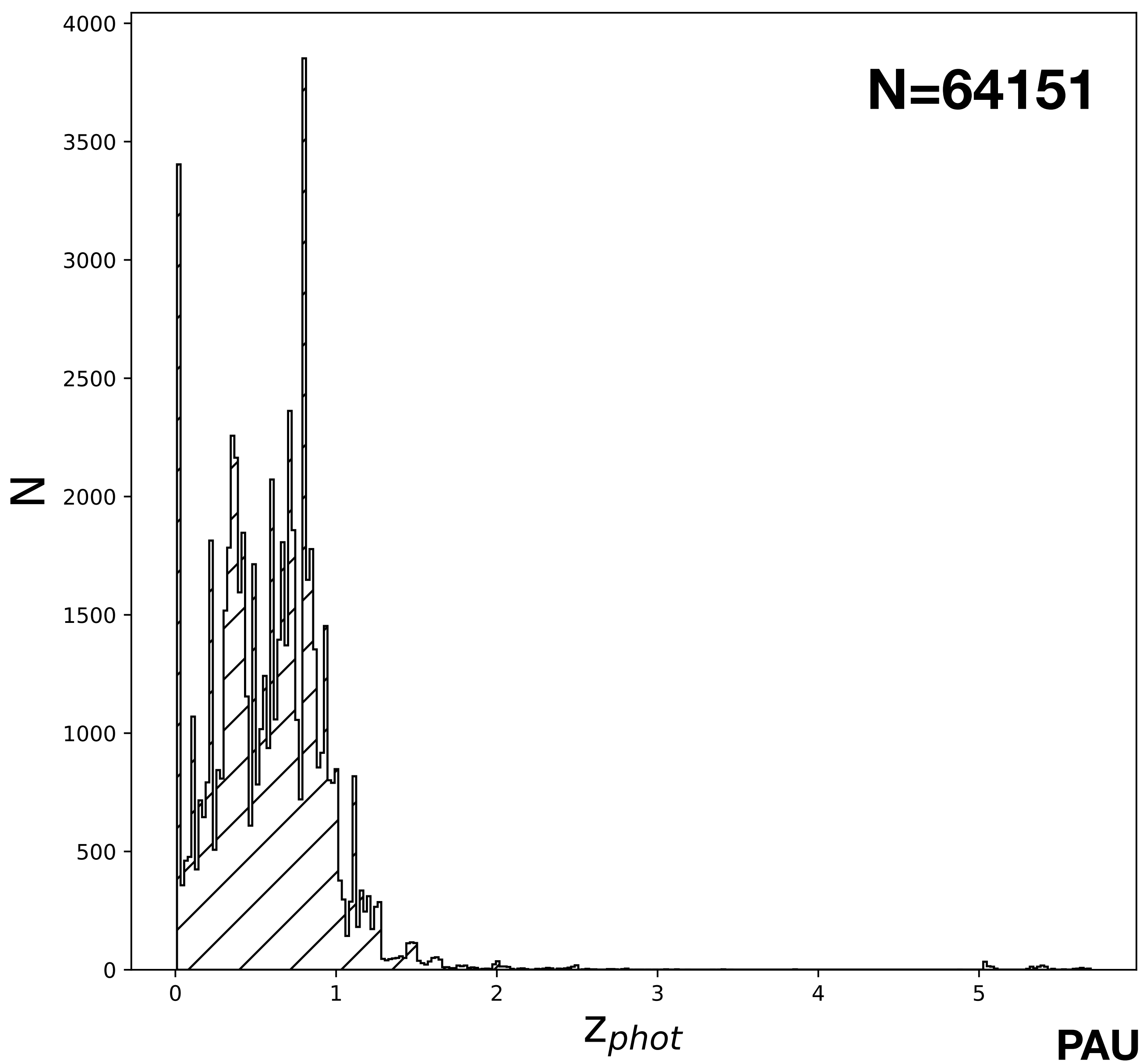
$$\Delta_z \equiv (z_{phot} - z_{spec})/(1 + z_{spec})$$

- **Redshift precision: the combination of 40 narrow bands from PAU, intermediate and broad bands from different instruments from the COSMOS survey.**

Data



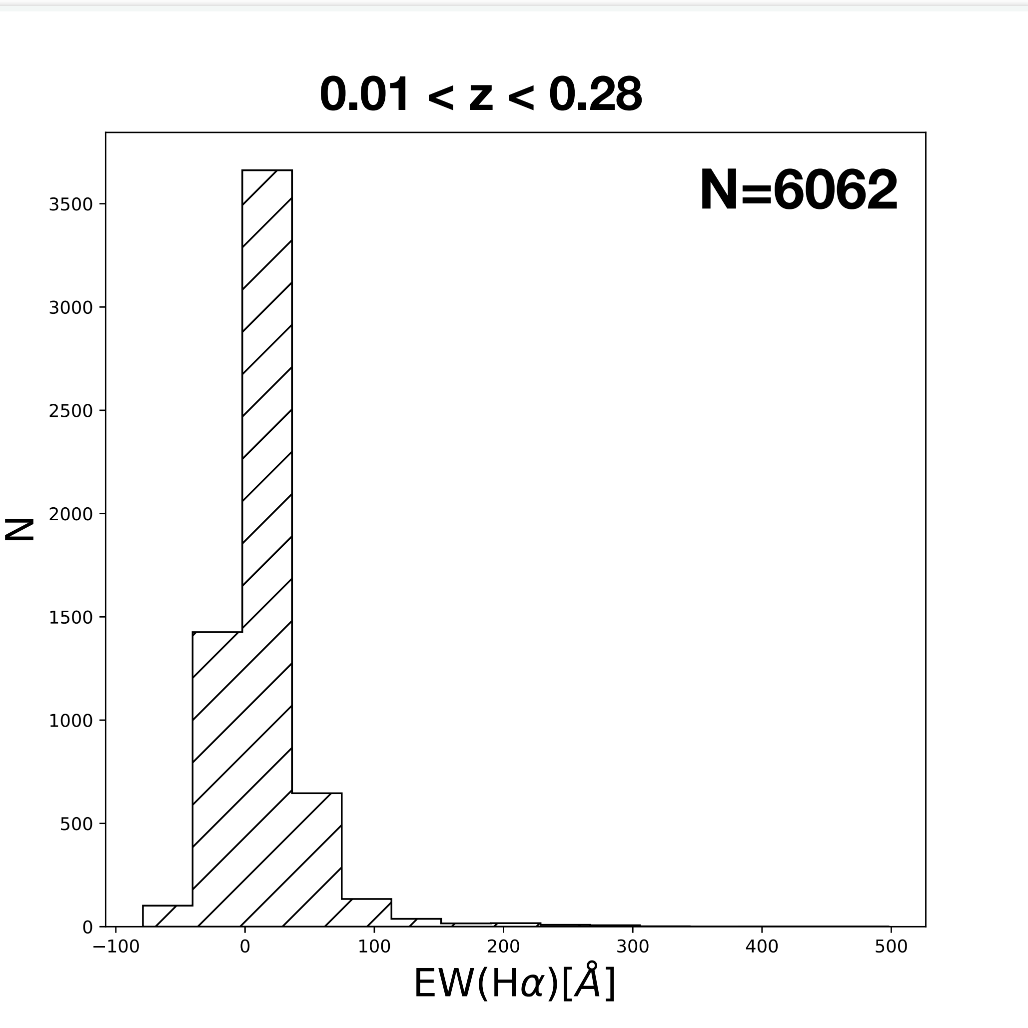
Alarcon et al. 2020



PAU Collaboration (the entire database)

Methodology

The classification has been carried out through the Gaussian Mixtures (GM) method and is focused in the search for significant differences in the shape of normalised rest-frame SEDs.

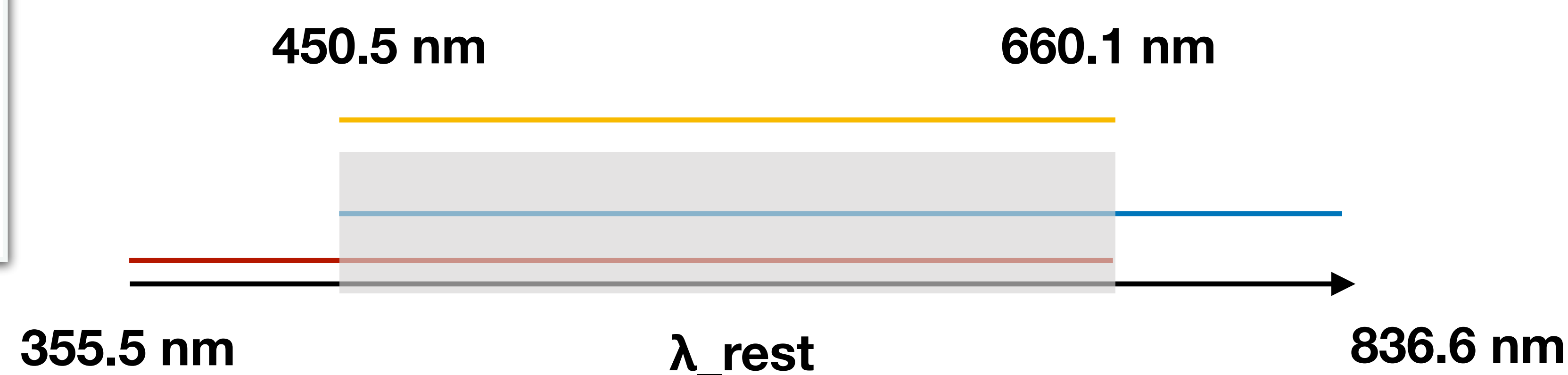


SEDs sample preparation

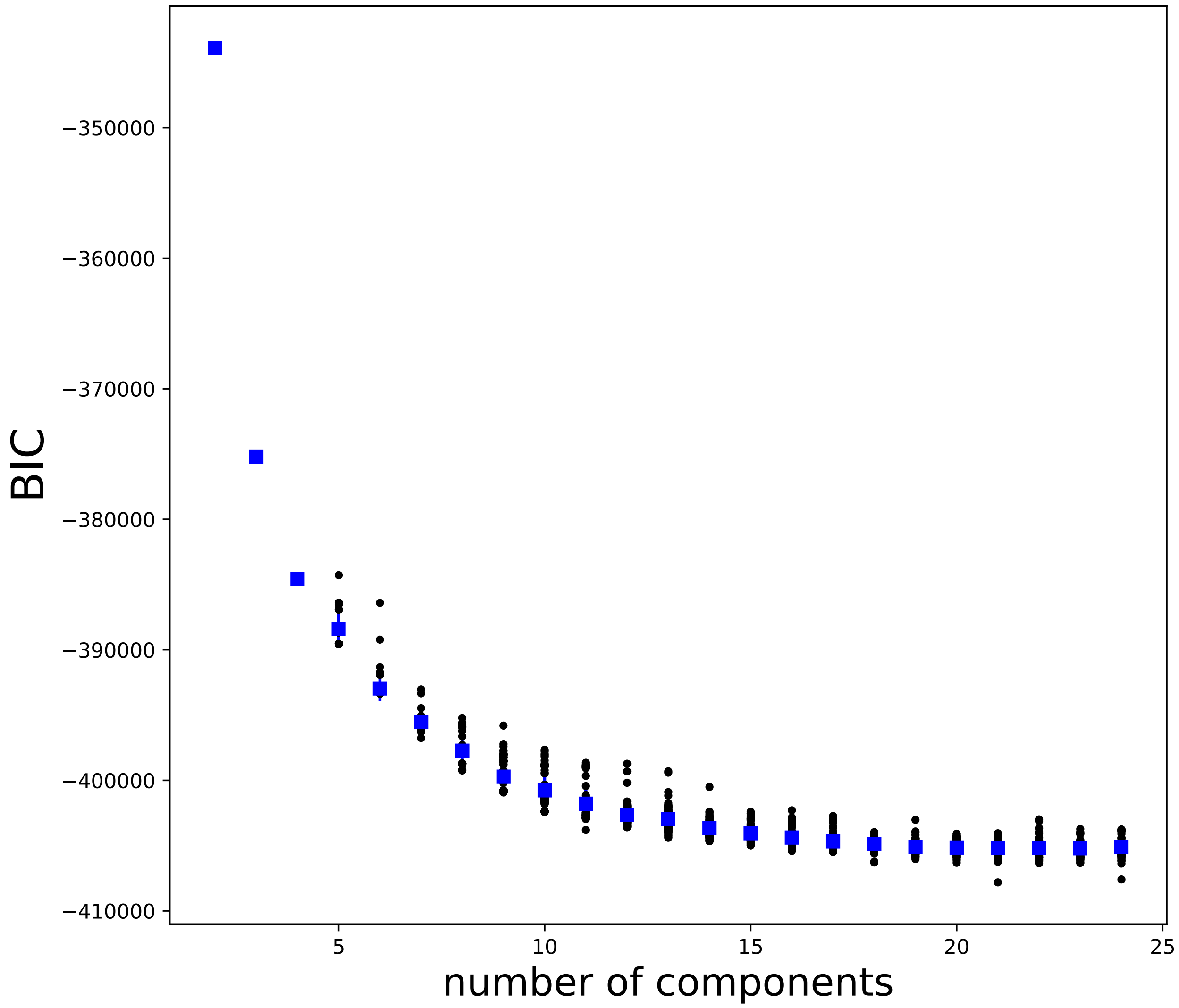
For 0.01 < z < 0.28:

F455 corresponds to λ_rest=450.5 nm
F845 corresponds to λ_rest=836.6 nm

F455 corresponds to λ_rest=355.5 nm
F845 corresponds to λ_rest=660.1 nm



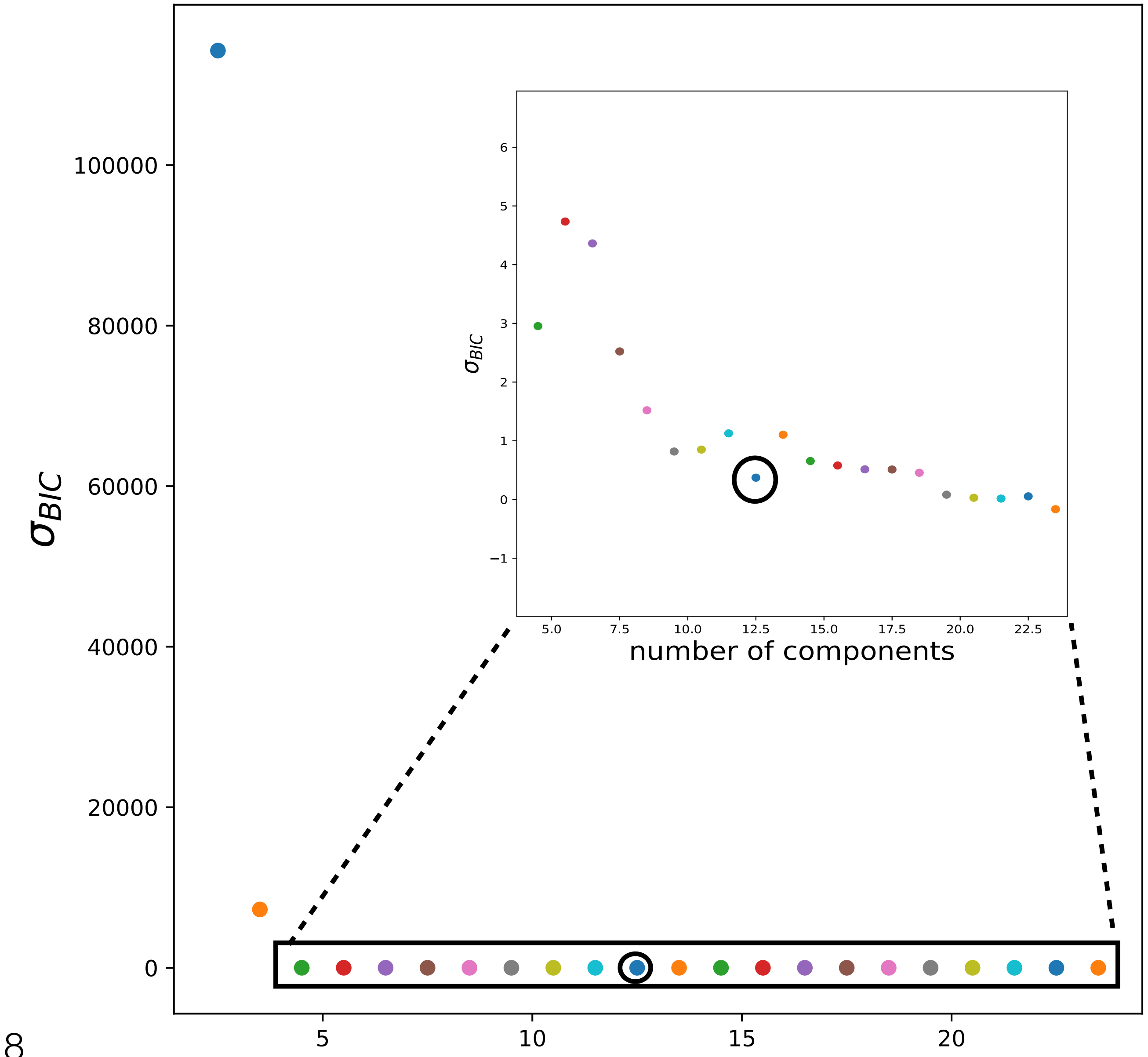
Methodology



$$\sigma(BIC) = \frac{mean(BIC_{n_i}) - mean(BIC_{n_{i+1}})}{error_{mean(BIC_{n_{i+1}})}}$$

$$\mathrm{BIC} = \chi^2 + q\ln(m)$$

where q is the number of parameters of the model used and m, the number of independent data points available



Silhouette Coefficient

Once the best number of clusters has been estimated, we make use of the silhouette coefficient to select the best GM model among all the different runs resulting in that optimal number of clusters.

$$s(x_i) = \begin{cases} \frac{b(x_i)-a(x_i)}{\max\{a(x_i),b(x_i)\}} & \text{if } |C_k| > 1 \\ 0 & \text{if } |C_k| = 1. \end{cases}$$

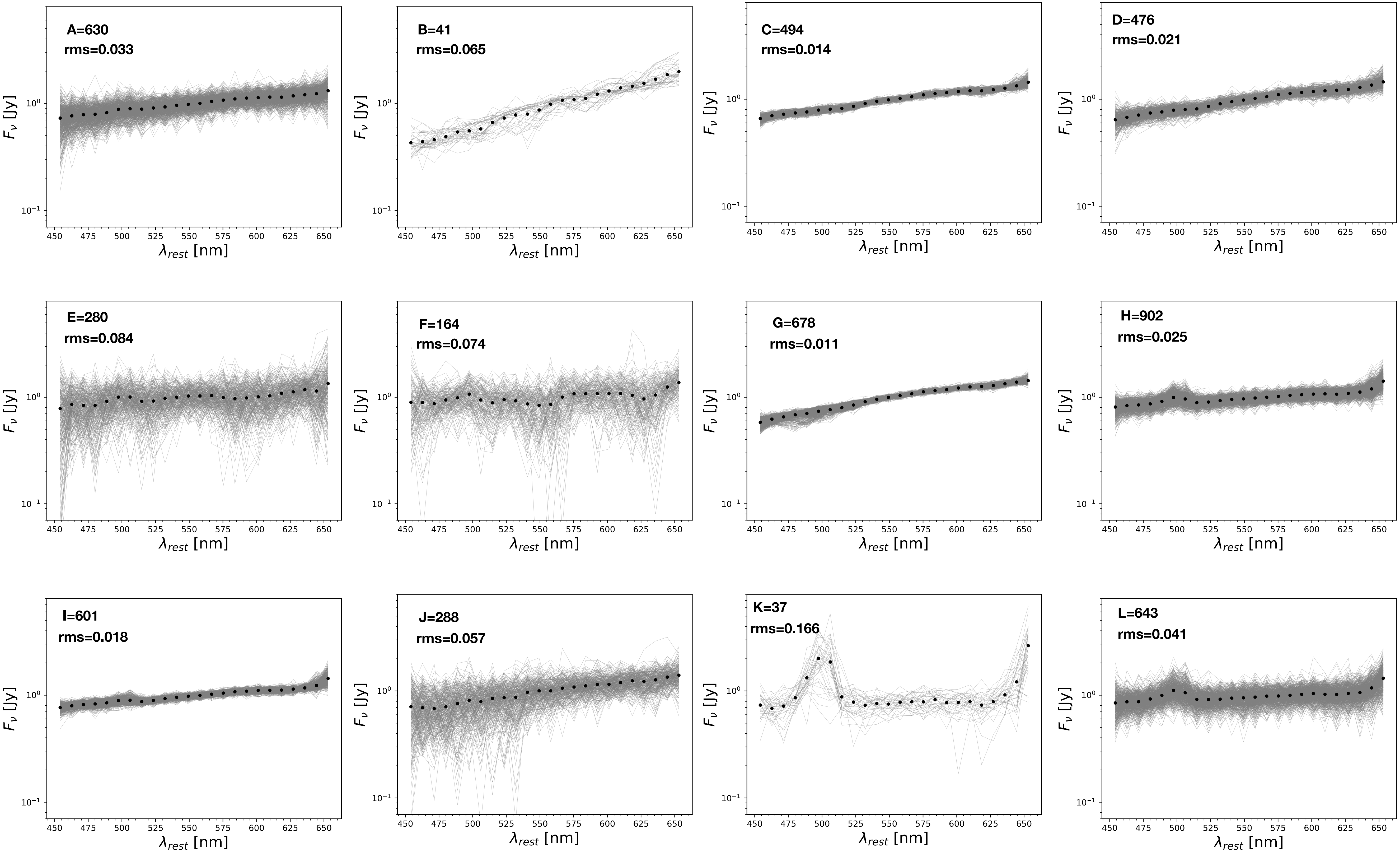
→ *This gives a perspective into the density and separation of the formed clusters* –

- The best value is 1
- The worst value is -1.
- Values near 0 indicate overlapping clusters.

The Silhouette Coefficient is a measure of how well defined the samples are clustered

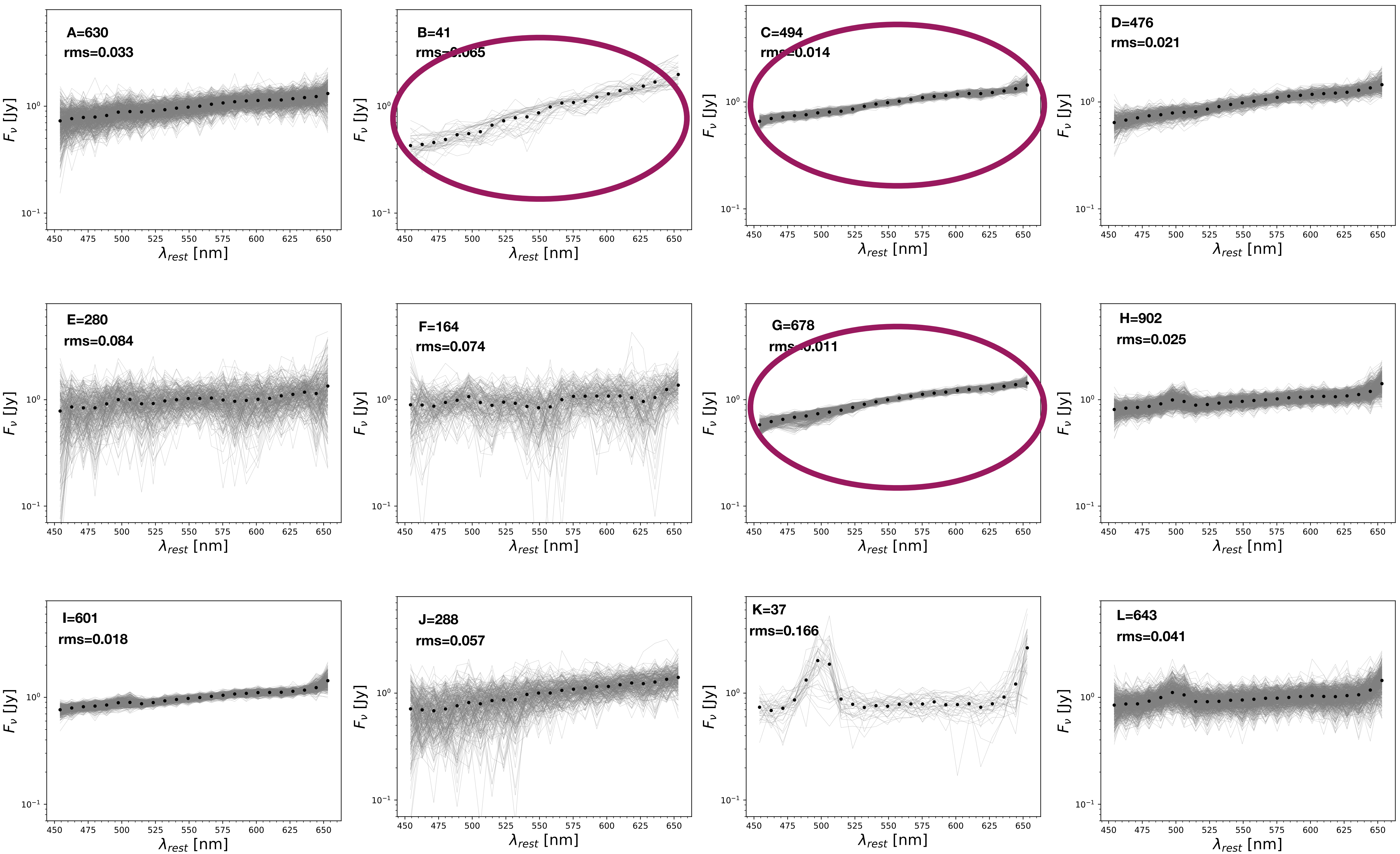
Model 1002 (12 classes, highest silhouette score)

Preliminary results



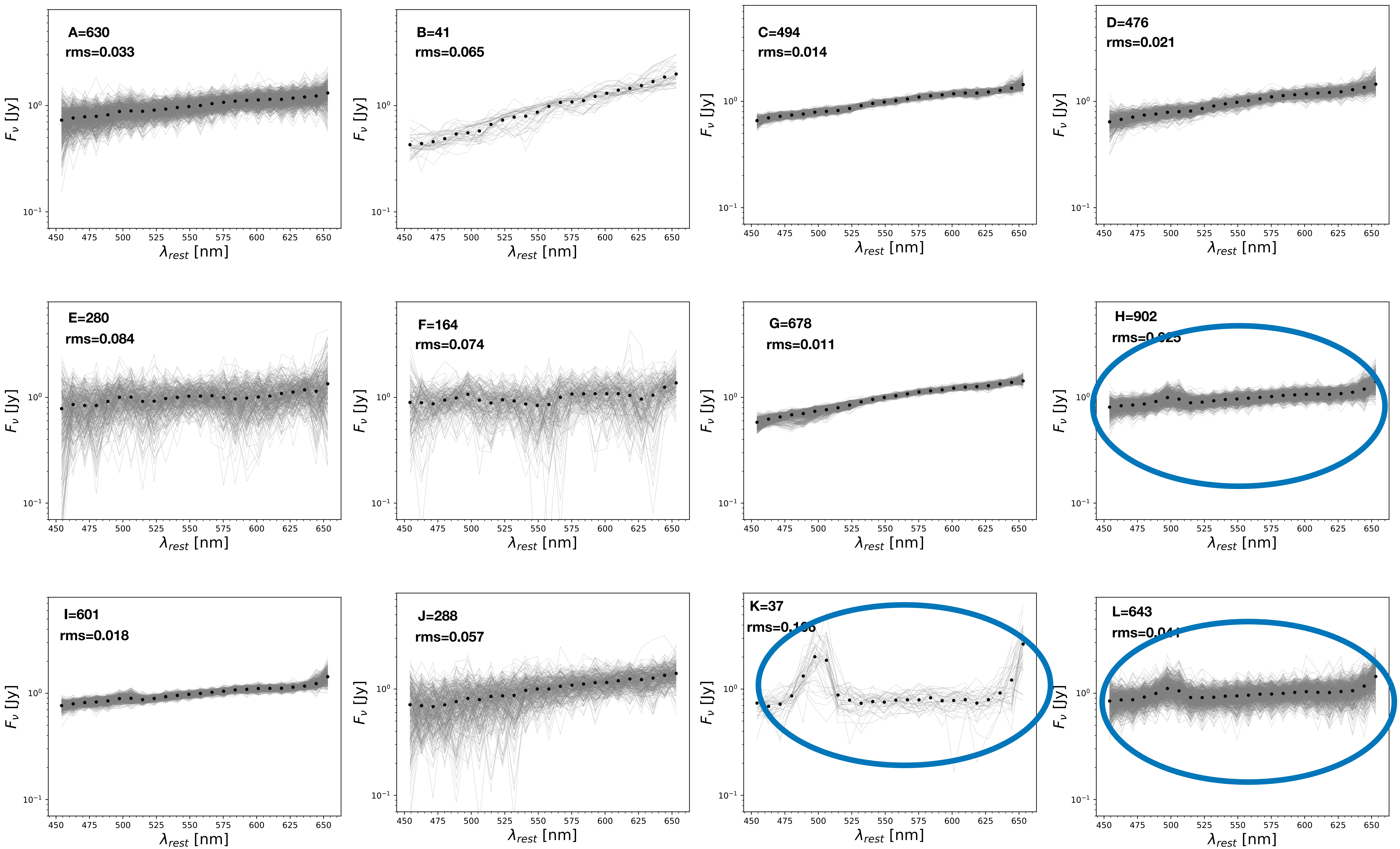
Model 1002 (12 classes, highest silhouette score)

Preliminary results



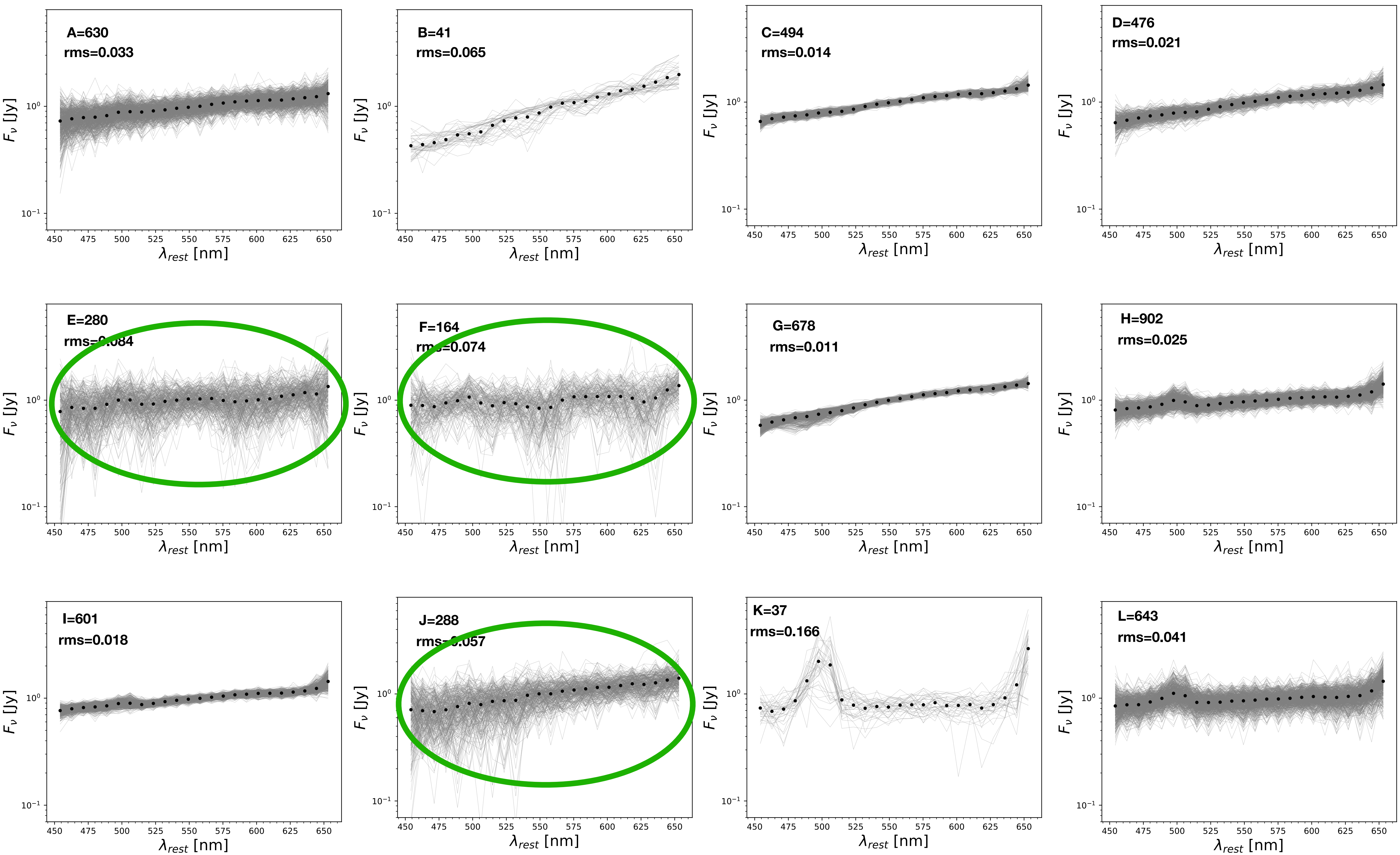
Model 1002 (12 classes, highest silhouette score)

Preliminary results



Model 1002 (12 classes, highest silhouette score)

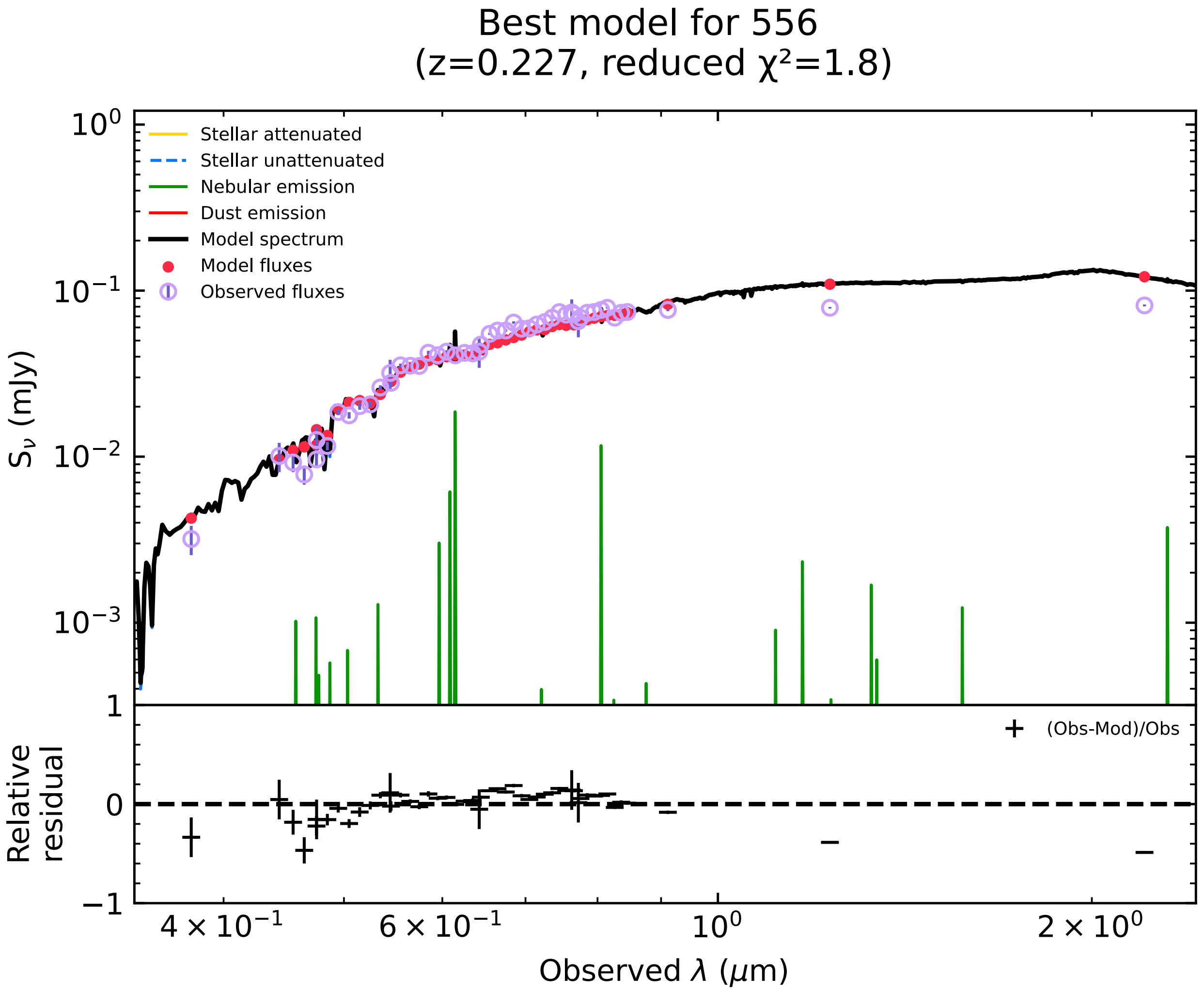
Preliminary results

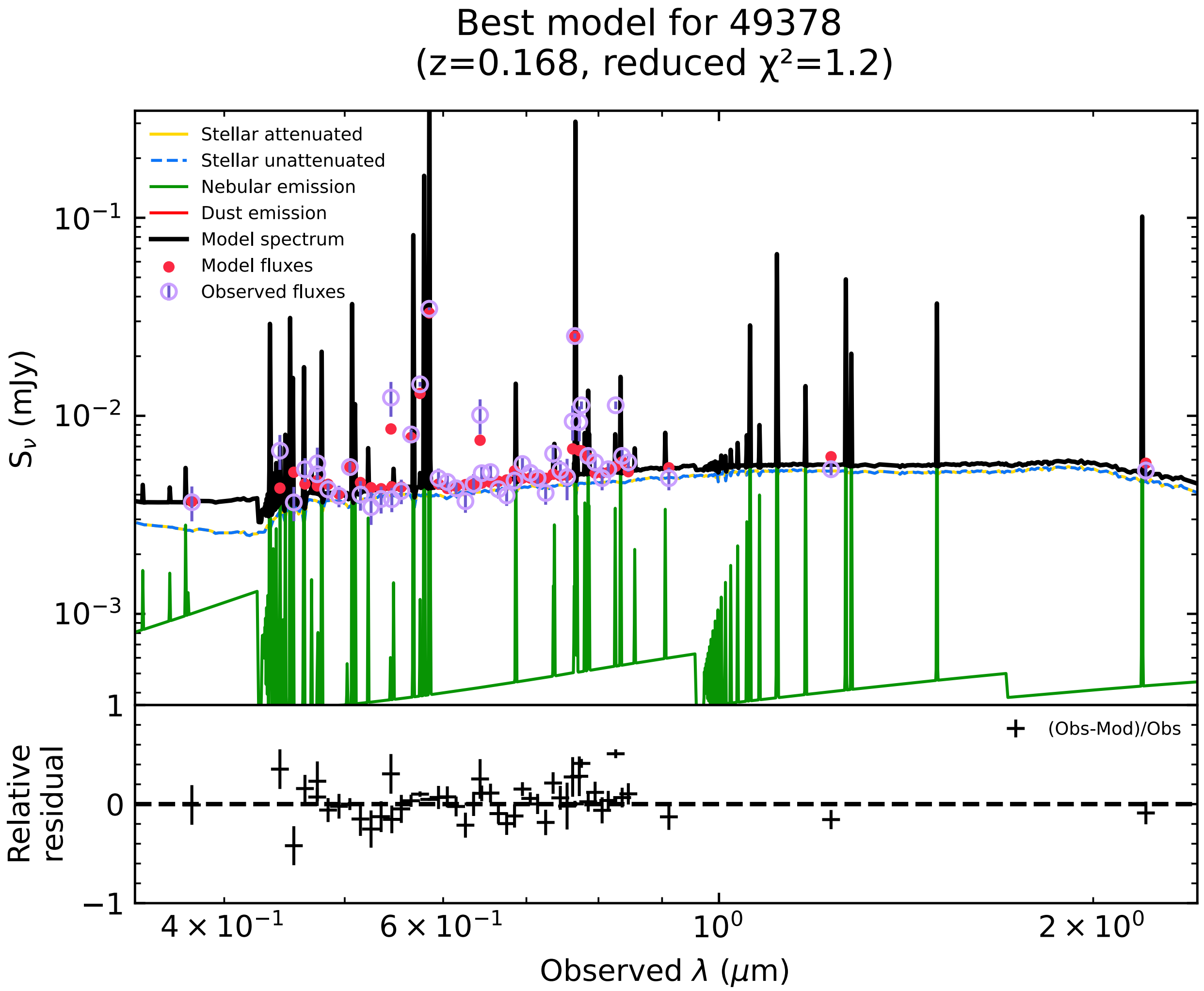


Preliminary results

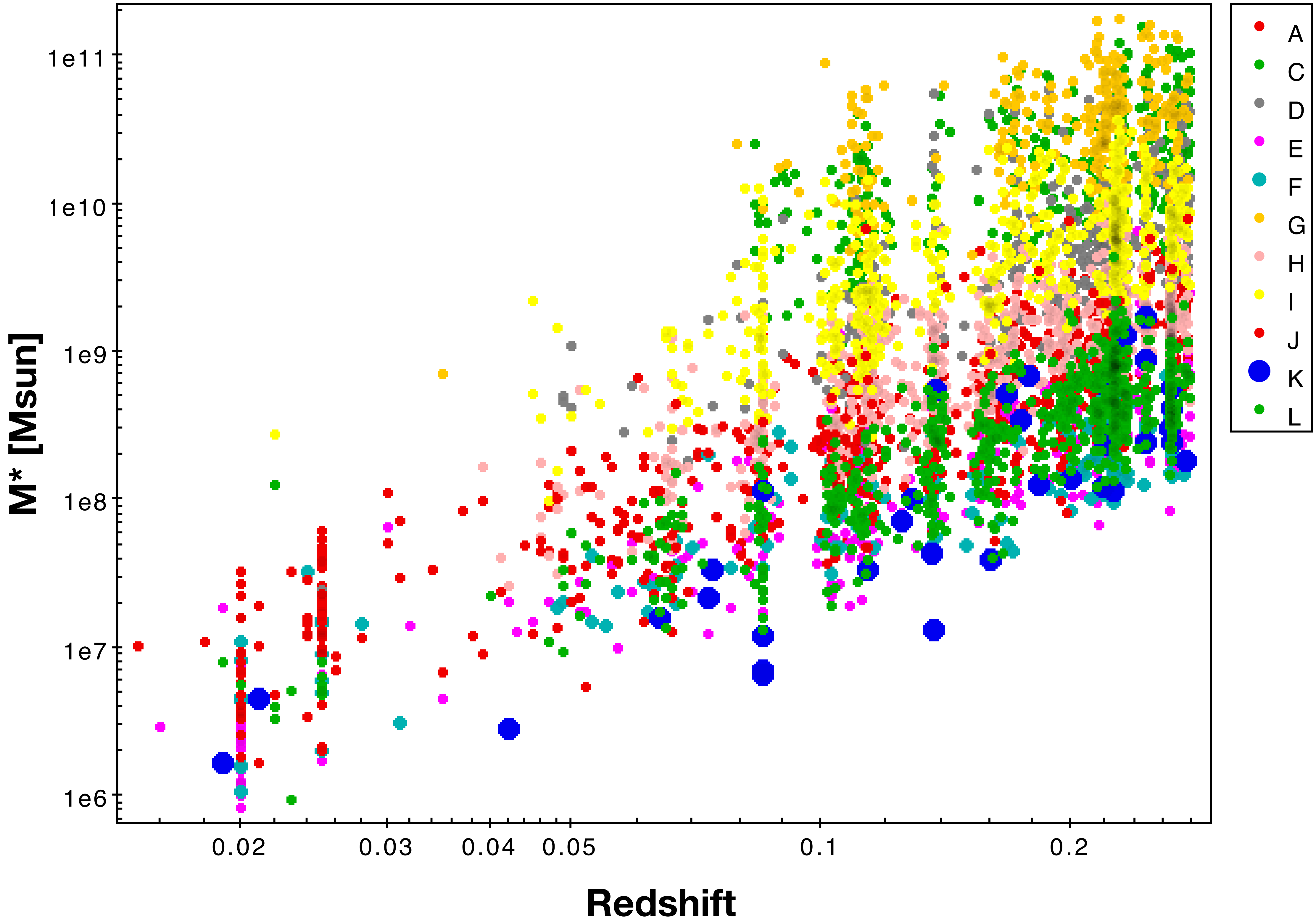
CIGALE

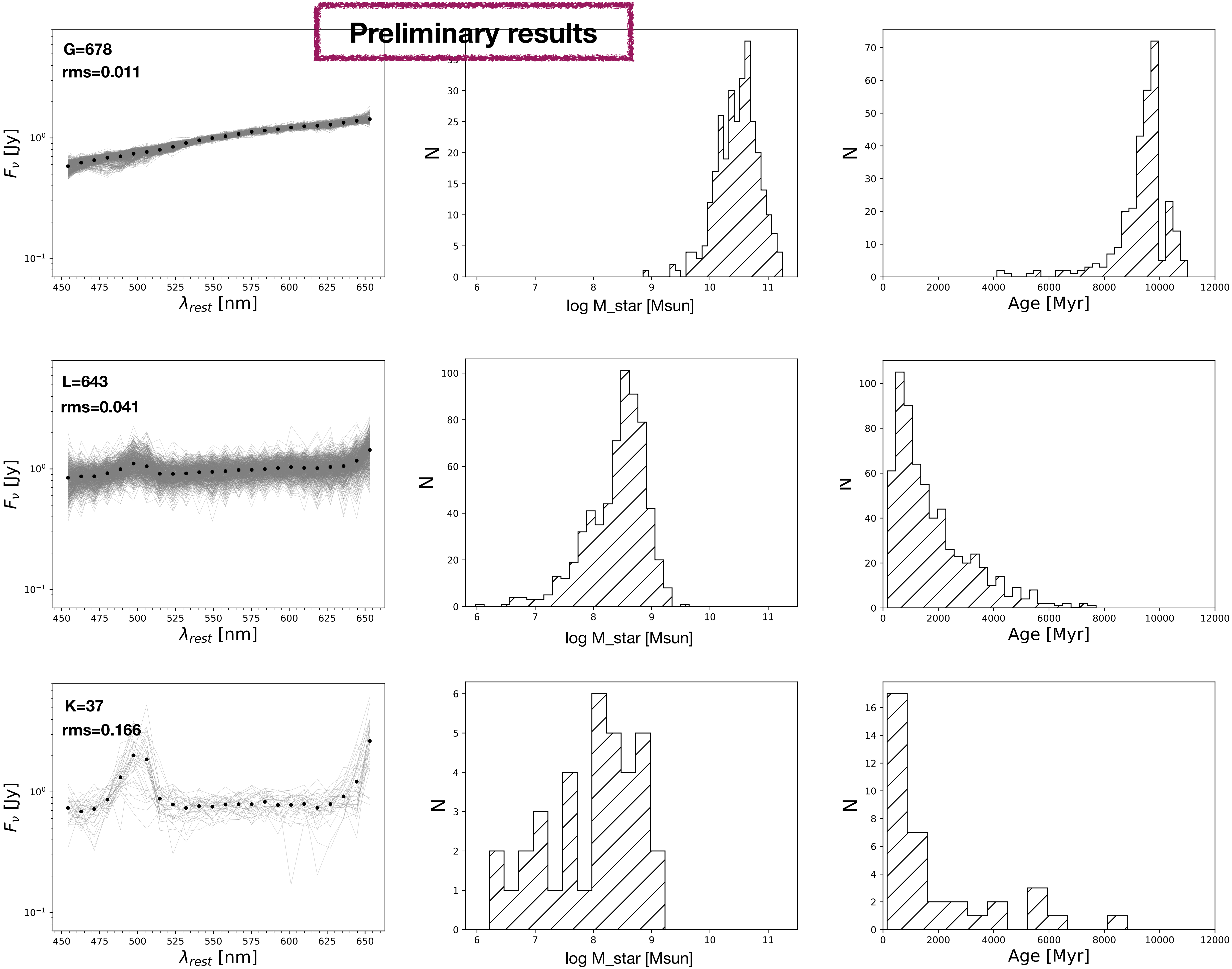
We run CIGALE adding broad band filters also provided by the PAU survey





Preliminary results





Preliminary results
G=678
rms=0.011
L=643
rms=0.041
K=37
rms=0.166
F_ν [Jy]
λ_{rest} [nm]
N
log M_star [Msun]
Age [Myr]

16

Conclusions

- **Characterization of galaxies down to $z < 0.3$ in the COSMOS field using an unsupervised Machine Learning clustering of a photometric galaxy survey (PAU survey) with a quasi-spectroscopic redshift precision**

- **Differences in the shapes of normalised rest-frame SEDs have been found (different slopes, some have emission lines like Hb, [O III] and Ha, others have absorption lines like Mg and Na I)**

- **Age and M* estimated with CIGALE for each class are in agreement with the shape observed in the normalised rest-frame low-z SEDs.**

- **Identification of a small population with intense emission lines.**

- **Thanks to the accurate and precise photometric redshifts given by the PAU survey we were able to do this analysis.**

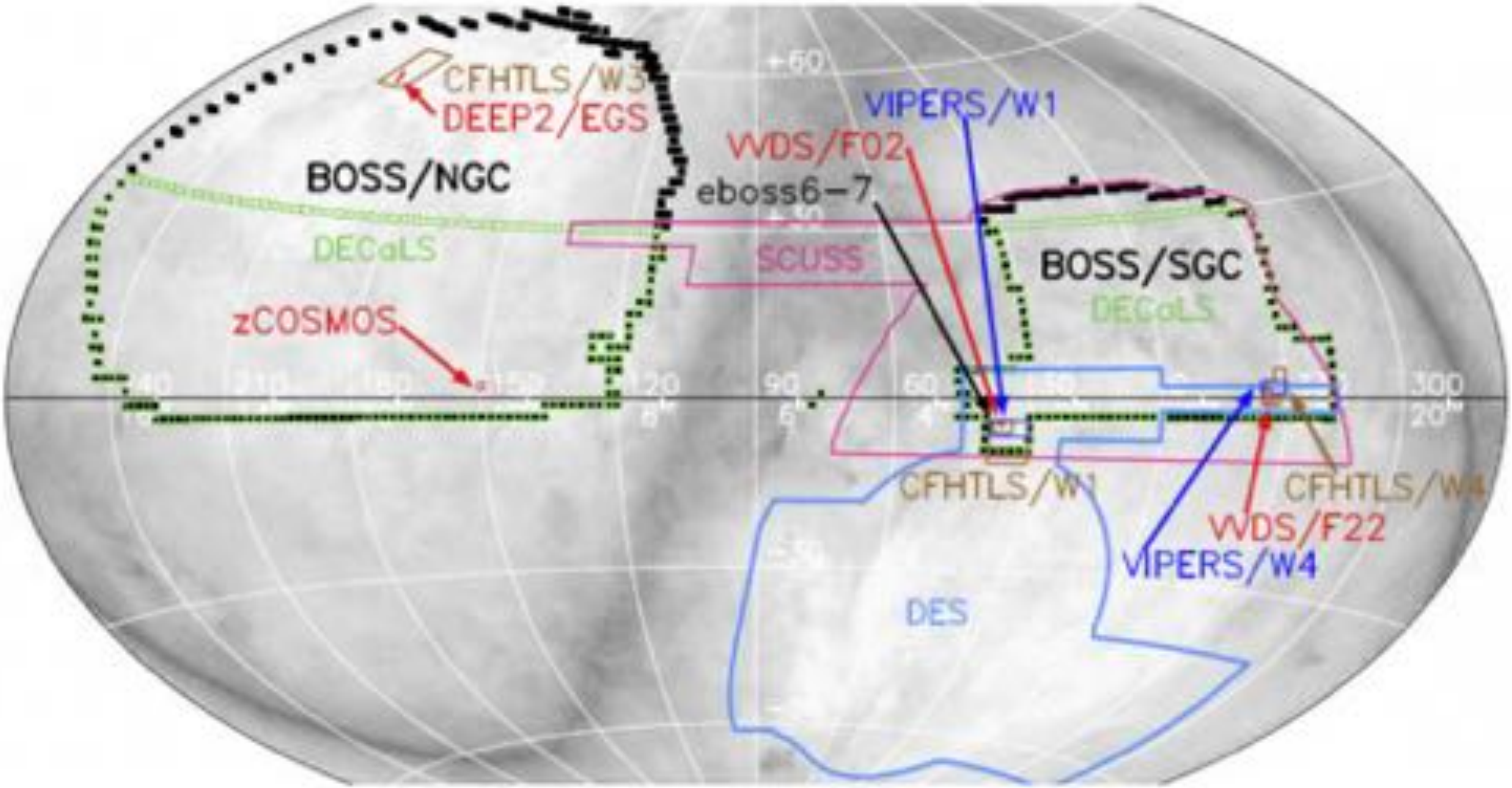
- **Work in progress**

Gracias!

supplementary material

CIGALE

- SFH (double exponential)
- SSP (Bruzual and Charlot 2003) with a Salpeter Initial mass function
- Nebular emission (continuum and line nebular emission)
- Dust attenuation (modified Charlot & Fall 2000 attenuation law)
- Dust emission (Dale et al. 2014 dust emission templates)
- tau_main = 50, 100, 250, 500, 1000, 2000, 4000, 6000, 8000, 15000, 30000
(e-folding time of the main stellar population model in Myr.)
- f_burst = 0.0, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2
(Mass fraction of the late burst population.)
- age = 10, 150, 250, 500, 1000, 2000, 4000, 8000, 10000, 12000
(Age of the main stellar population in the galaxy in Myr)
- metallicity = 0.0001, 0.0004, 0.004, 0.008, 0.02, 0.05
- Av_ISM = 0., 0.2, 0.4, 0.6, 0.8, 1., 1.2, 1.4, 1.6, 1.8, 2., 2.2, 2.4, 2.6, 2.8, 3., 3.2, 3.4, 3.6, 3.8, 4.



CFHTLS/W3
DEEP2/EGS
BOSS/NGC
DECaLS
zCOSMOS
VIPERS/W1
VVDS/F02
eboss6-7
SCUSS
BOSS/SGC
DECaLS
CFHTLS/W1
CFHTLS/W4
VVDS/F22
VIPERS/W4
DES
+60
300
20h

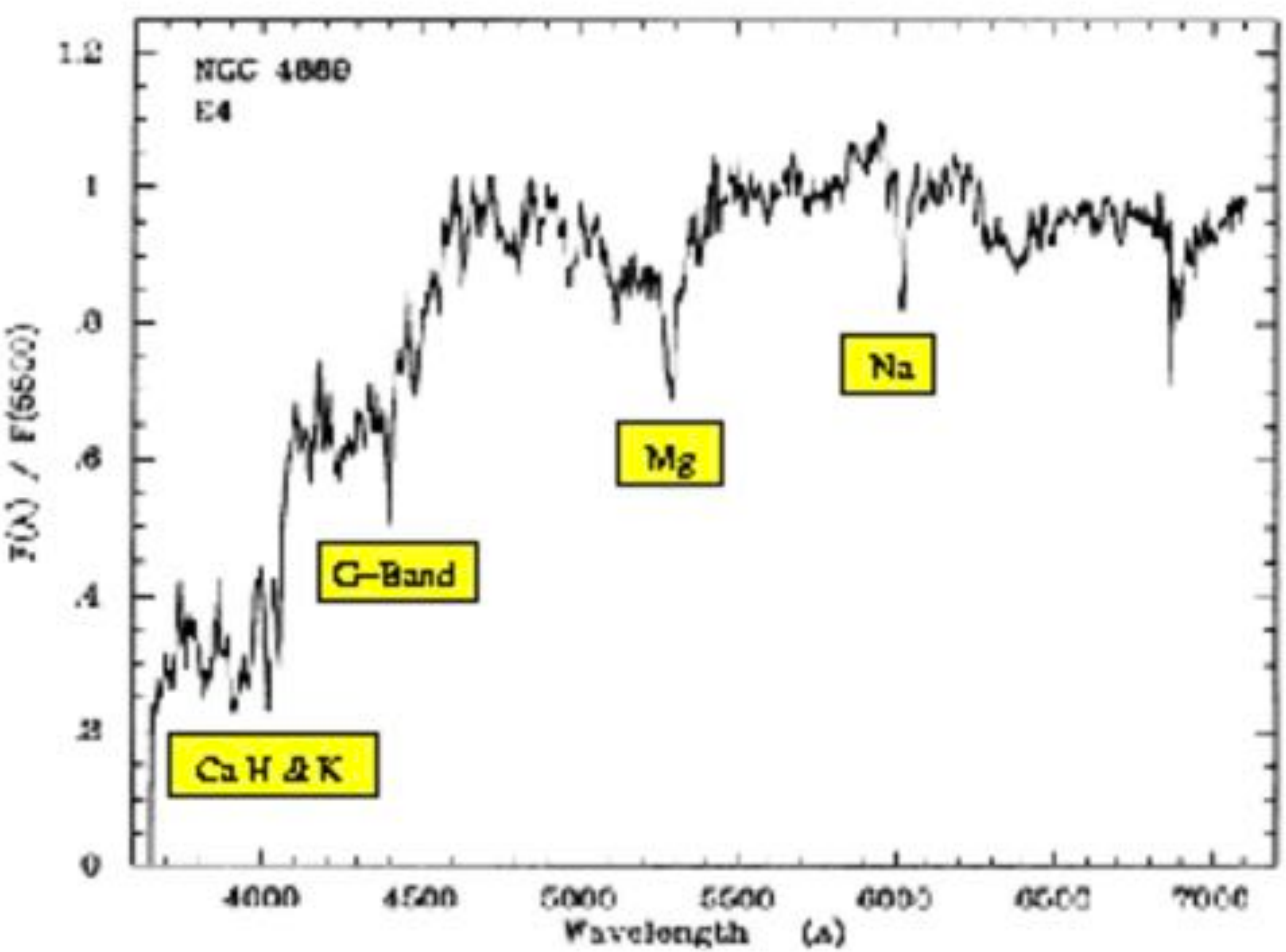
- **The flux line (and error) of the OII, OIII, Hbeta and Halpha lines is measured for each object at the photometric redshift using the best model. The measurements are corrected for internal galaxy dust extinction.**

- **The absolute magnitudes of the GALEX NUV, CFHT u, Subaru r and i bands are measured using the photometric redshift of the catalog and the best model. The measurements are corrected for internal galaxy dust extinction.**

- **The best model, extinction law and extinction value is given for each object.**

Publicly available. For extensive details see Alarcon et al. 2020.

Elliptical galaxy spectra are characterized by strong absorption lines, due to metals in the stellar atmospheres of the low luminosity stellar population



Mg=5175 A

Na I=5889, 5895 AA

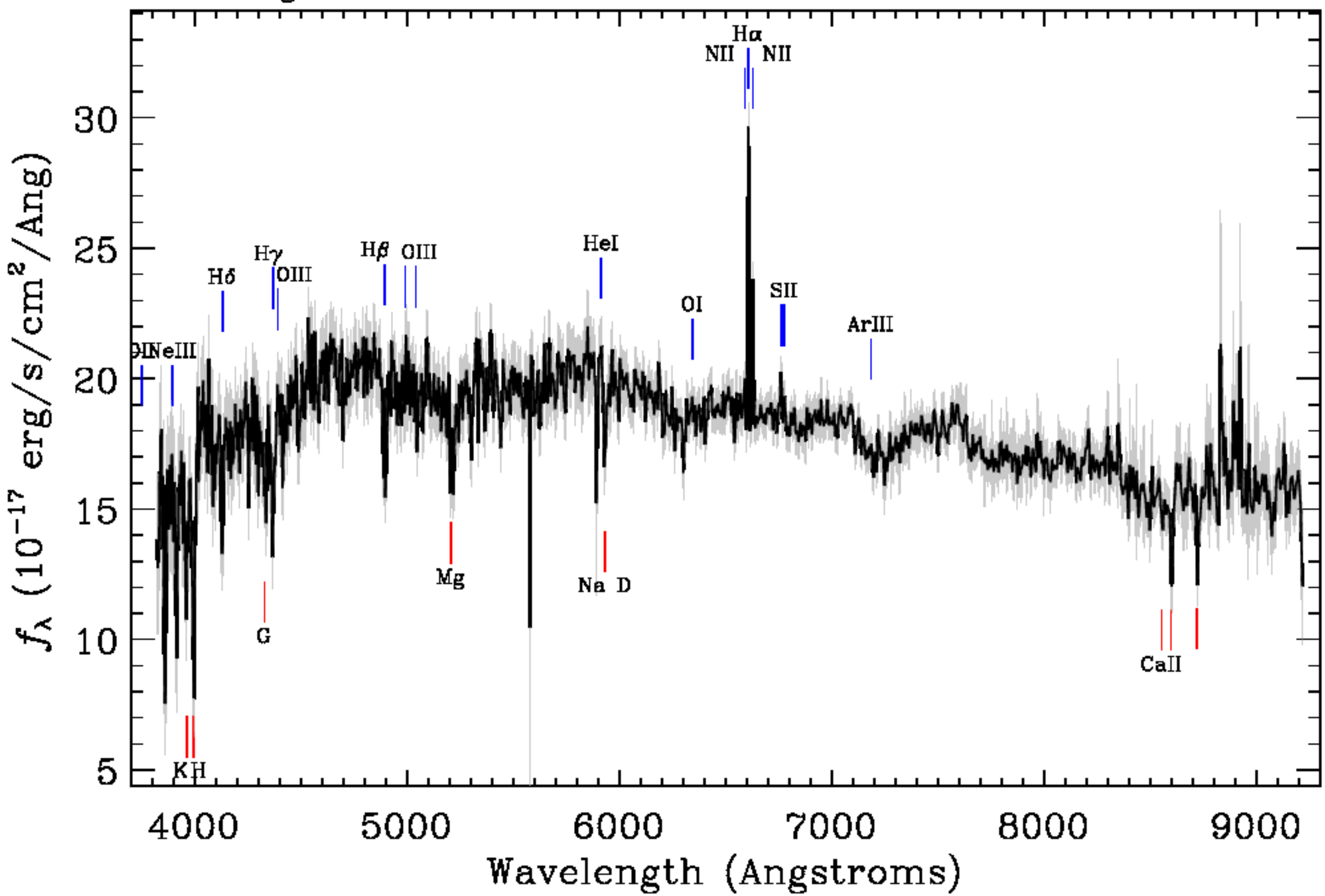
G (Ca I)=4307 A

H (Ca II)=3968 A

K (Ca II)=3933 A

Kennicutt, Robert, 1992

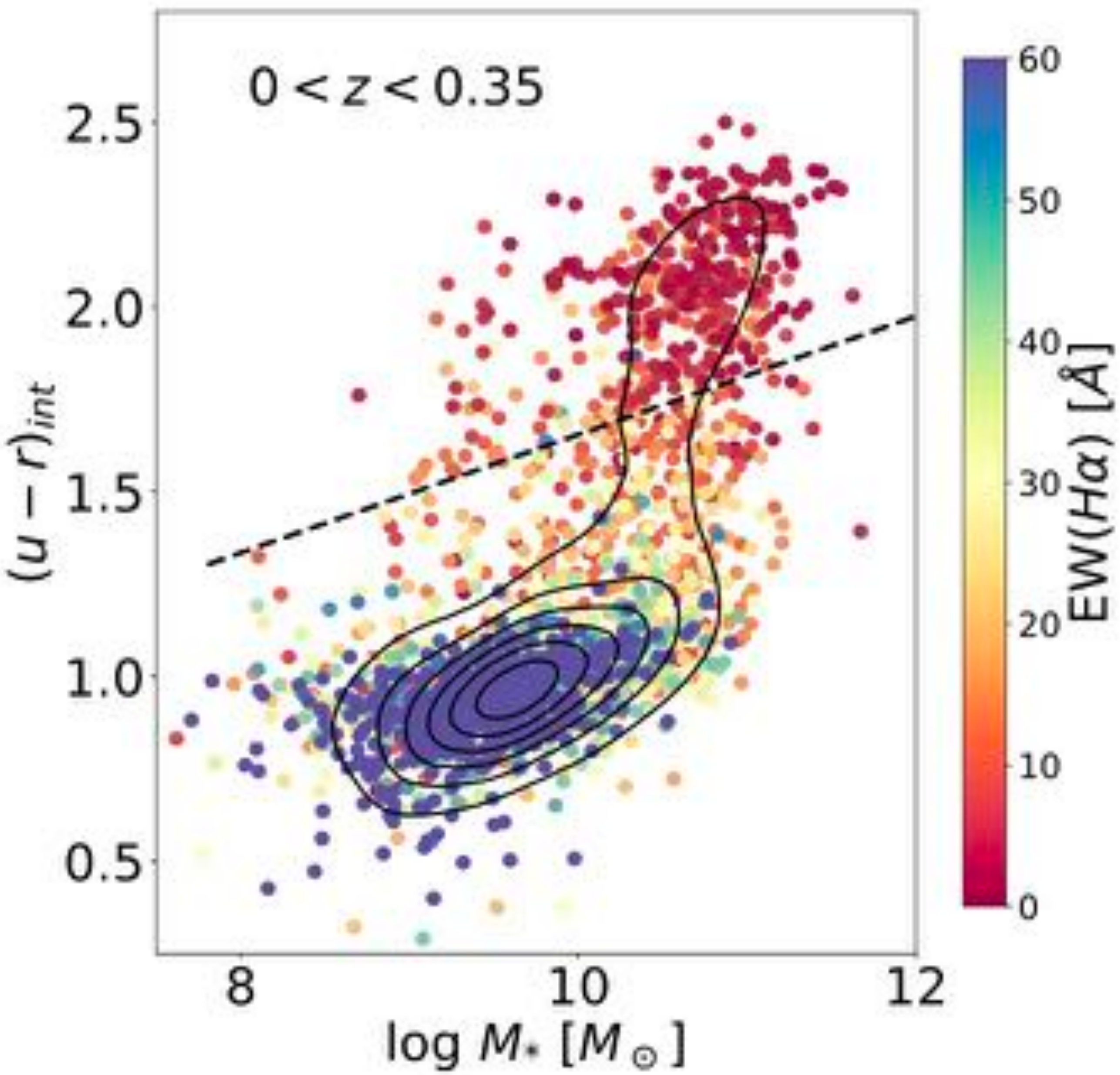
Survey: *sdss* Program: *legacy* Target: *GALAXY*
RA=188.59114, Dec=8.20465, Plate=1628, Fiber=334, MJD=53474
$z=0.00642\pm0.00002$ Class=GALAXY
No warnings.



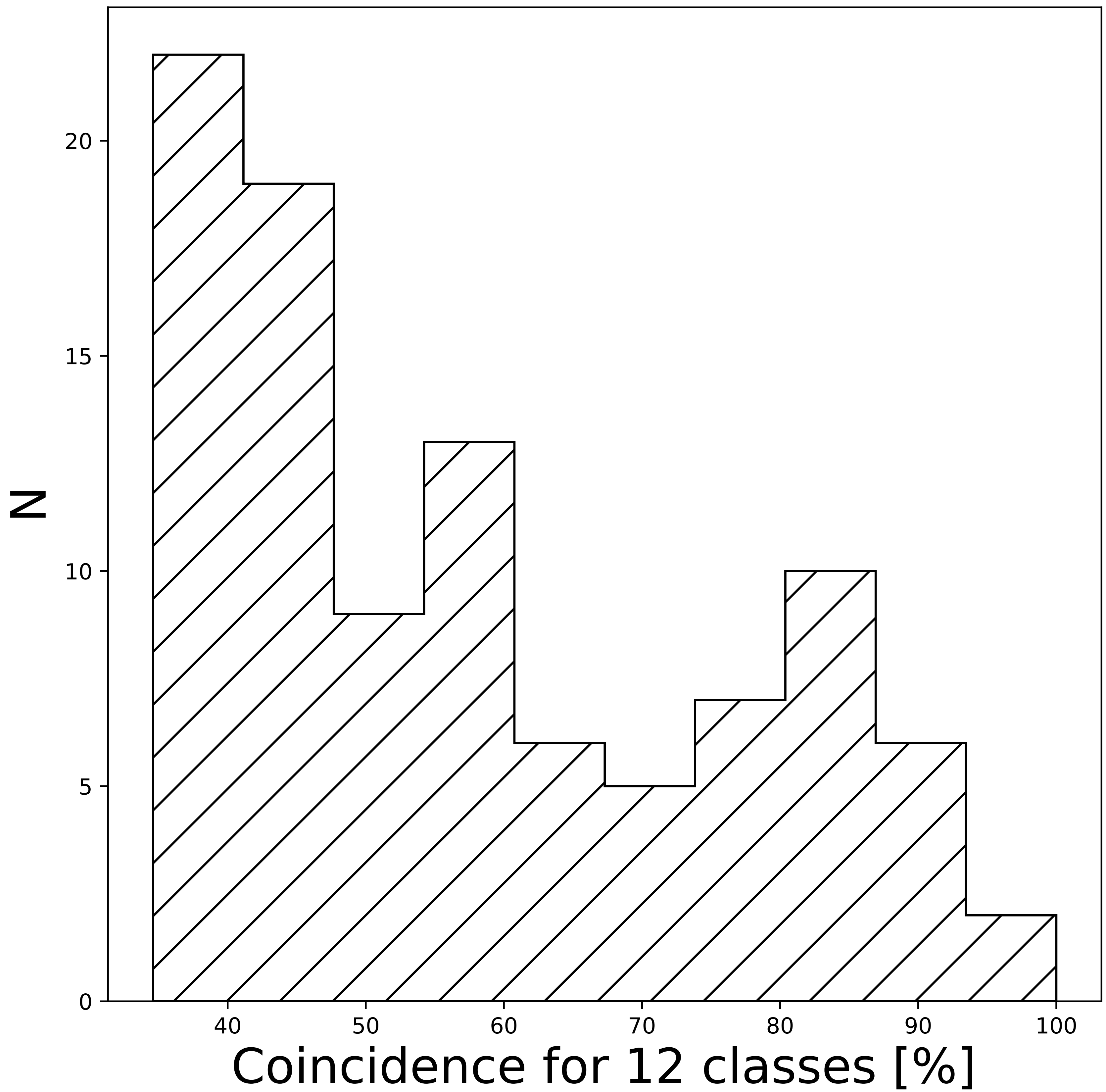
There is interference by strong airglow emission features from our atmosphere (such as [OI] 5577 Å).

These are the most likely wavelengths for unexpected peaks in a faint-object spectrum in case of imperfect sky subtraction, so any peaks or valleys at these locations may be artifacts. Airglow is the green area at the top of the atmosphere. Not aurora, but oxygen continuing to glow after sunset

The miniJPAS survey: Identification and characterization of the emission line galaxies down to z < 0.35 in the AEGIS field



Martínez-Solaeche et al. 2022


N
0
5
10
15
20
40
50
60
70
80
90
100
Coincidence for 12 classes [%]