Unsupervised classification of galaxy spectra up to z~1.2

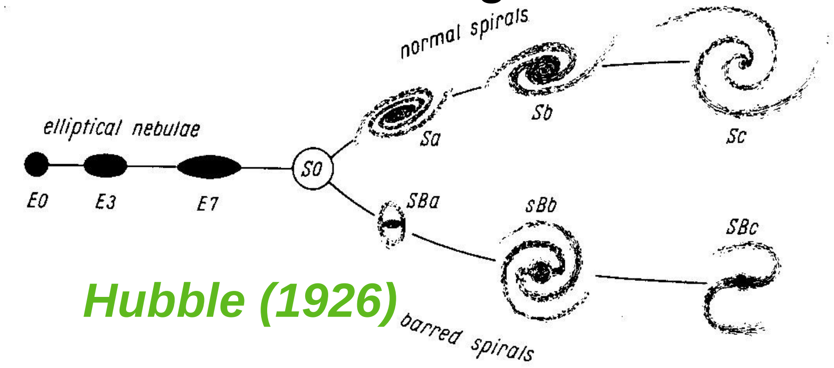
Towards galaxy spectral evolution ?

Jihane Moultaka
(IRAP, Observatoire Midi-Pyrénées, Toulouse, France)

Didier Fraix-Burnet *(IPAG, Grenoble, France)*
Julien Dubois *(PhD student)*
Cécilia Beissière-Thygesen *(Master student)*

Classification of galaxies

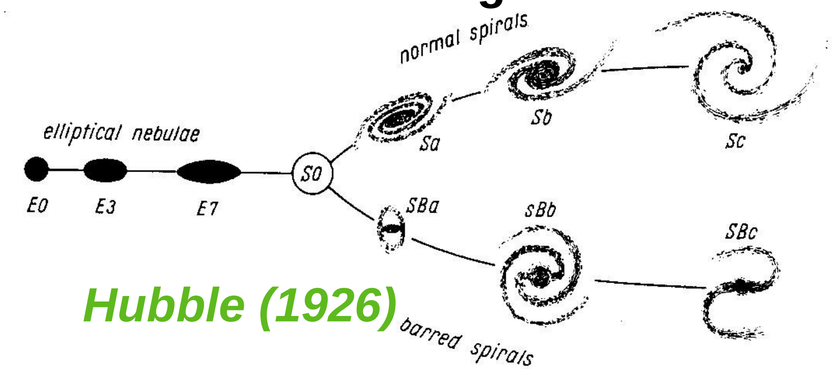
The Hubble tuning fork



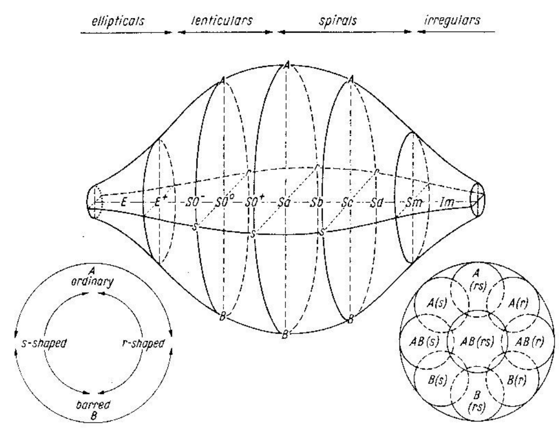
Hubble (1926)

Classification of galaxies

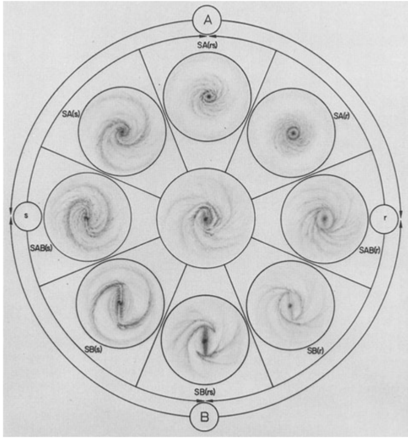
The Hubble tuning fork



Hubble (1926)

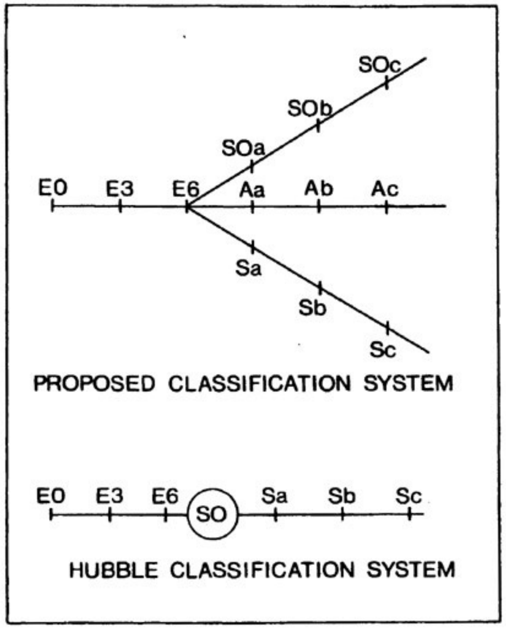


De Vaucouleurs (1959)



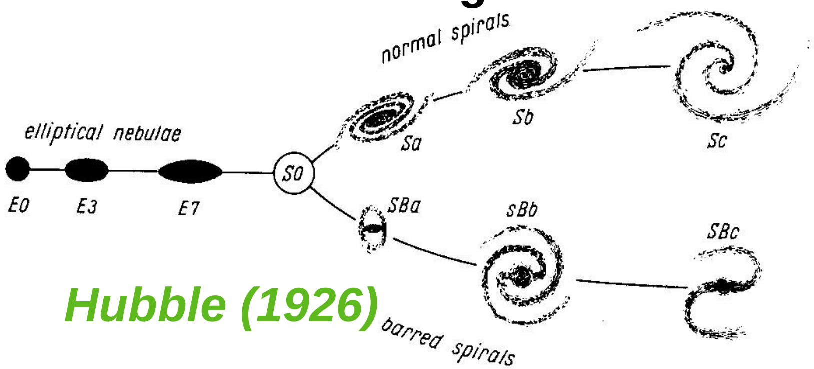
G. and A. De Vaucouleurs (1963)

Van den Bergh (1976)
''trifork'' diagram

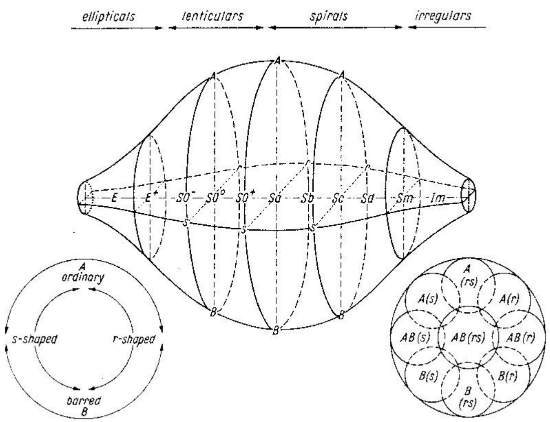


Classification of galaxies

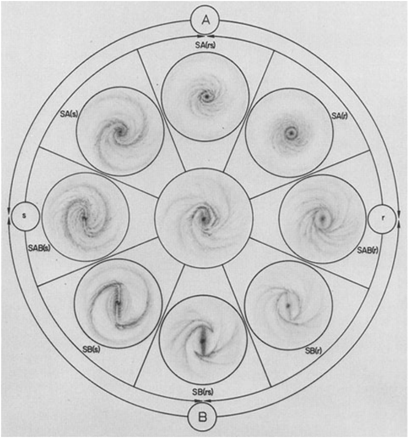
The Hubble tuning fork



Hubble (1926)

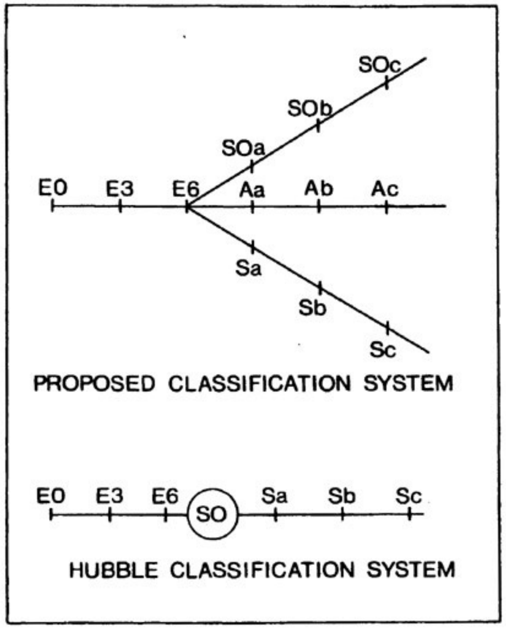


De Vaucouleurs (1959)



G. and A. De Vaucouleurs (1963)

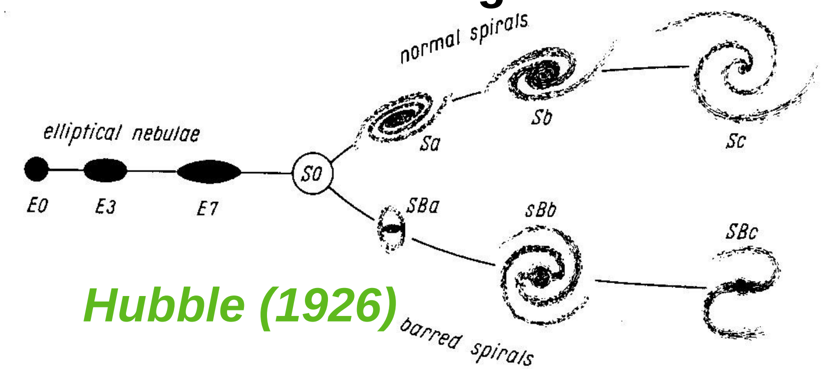
Van den Bergh (1976) ''trifork'' diagram



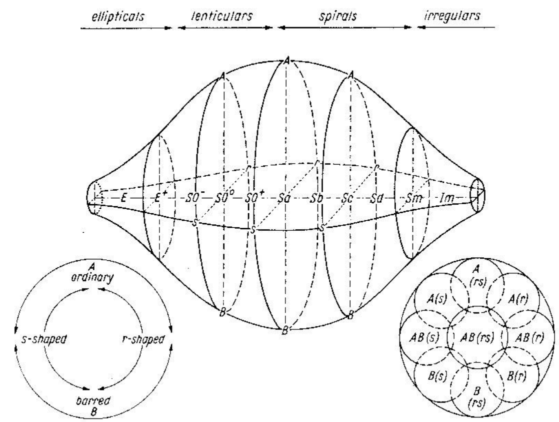
→ **Morphologies**

Classification of galaxies

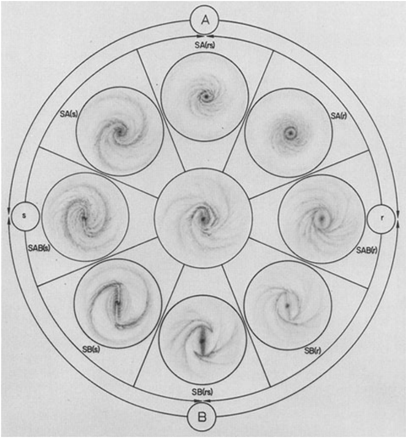
The Hubble tuning fork



Hubble (1926)

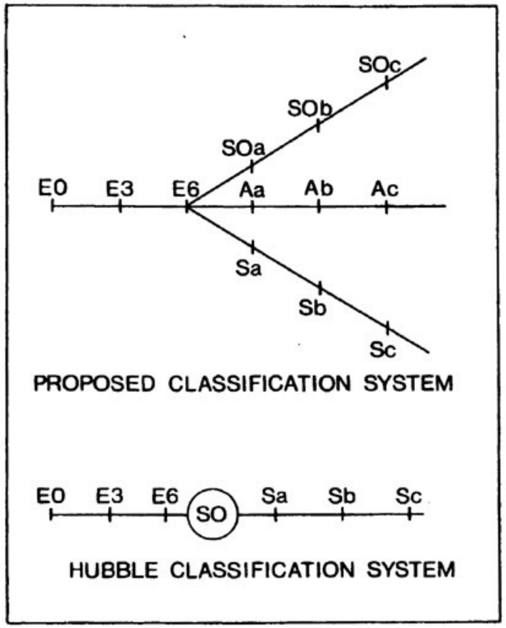


De Vaucouleurs (1959)



G. and A. De Vaucouleurs (1963)

Van den Bergh (1976) ''trifork'' diagram



… Automated classification with Machine Learning and Deep Learning

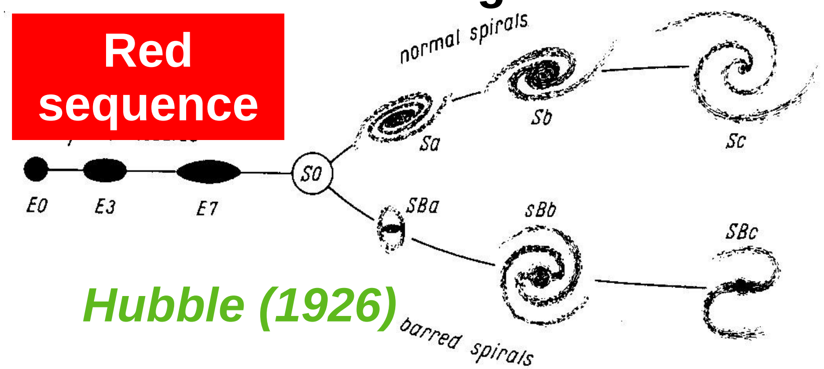
e.g : ***Fraix-Burnet (2023) , Zhang et al (2022), Cavanagh et al. (2021), Cheng et al. (2020), Khramtsov et al. (2019) ...***

Morphologies

Classification of galaxies

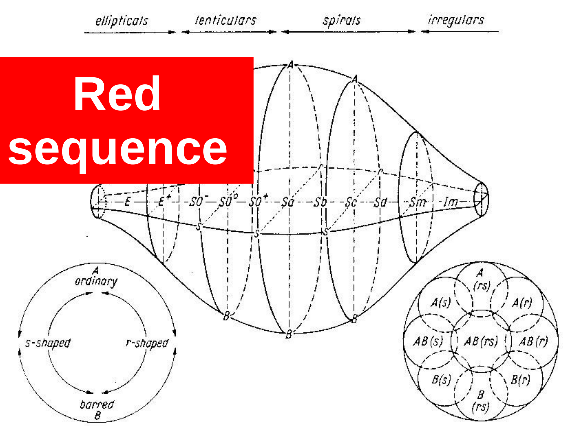
The Hubble tuning fork

Red sequence

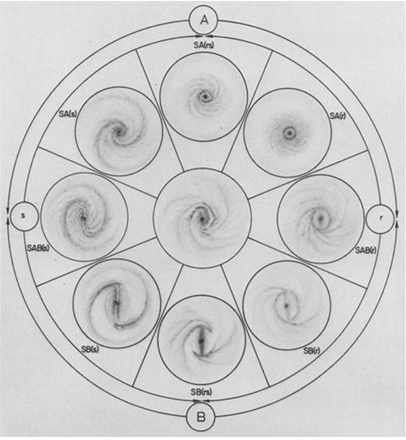


Hubble (1926)

Red sequence



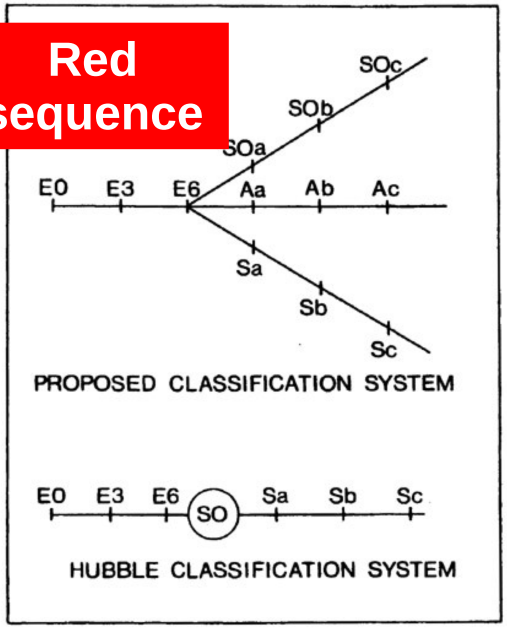
De Vaucouleurs (1959)



G. and A. De Vaucouleurs (1963)

Van den Bergh (1976) ''trifork'' diagram

Red sequence



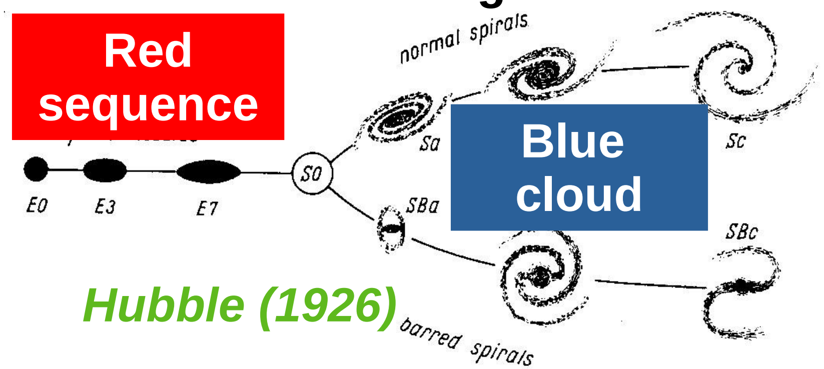
… Automated classification with Machine Learning and Deep Learning

e.g : ***Fraix-Burnet (2023) , Zhang et al (2022), Cavanagh et al. (2021), Cheng et al. (2020), Khramtsov et al. (2019) ...***

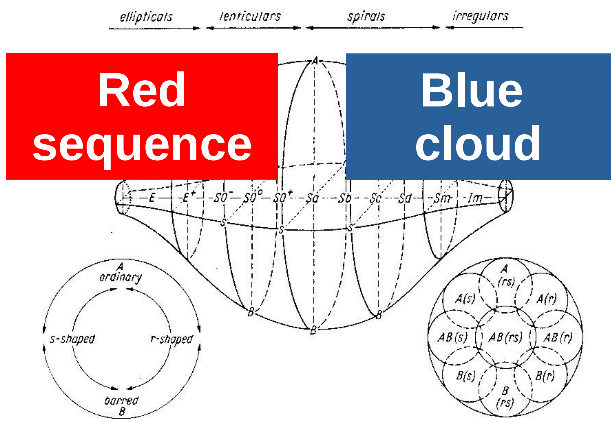
Morphologies

Classification of galaxies

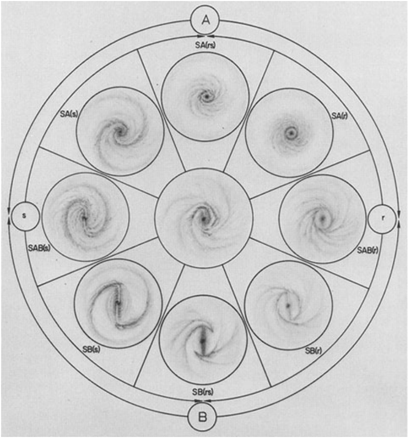
The Hubble tuning fork



Hubble (1926)

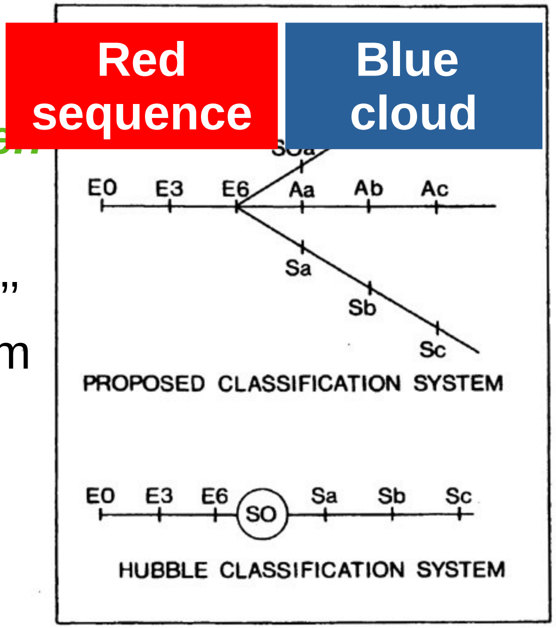


De Vaucouleurs (1959)



G. and A. De Vaucouleurs (1963)

Van den Bergh (1976) ''trifork'' diagram



… Automated classification with Machine Learning and Deep Learning

e.g : ***Fraix-Burnet (2023) , Zhang et al (2022), Cavanagh et al. (2021), Cheng et al. (2020), Khramtsov et al. (2019) ...***

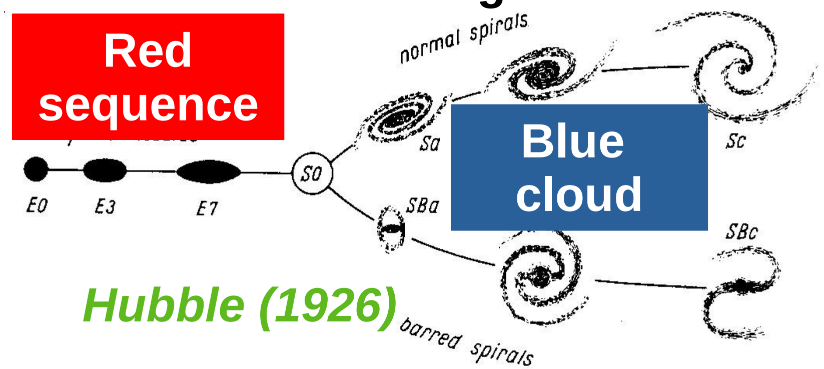
→ **Morphologies**

Classification of galaxies

The Hubble tuning fork

Red sequence

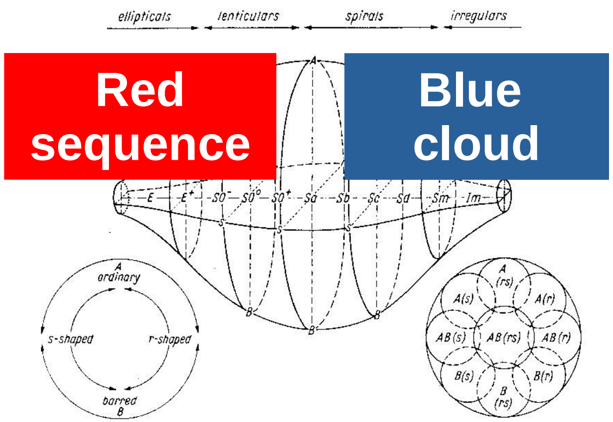
Blue cloud



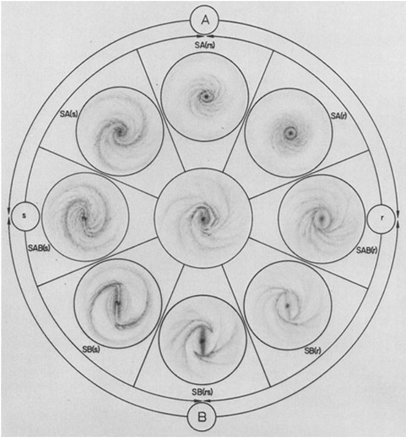
Hubble (1926)

Red sequence

Blue cloud



De Vaucouleurs (1959)

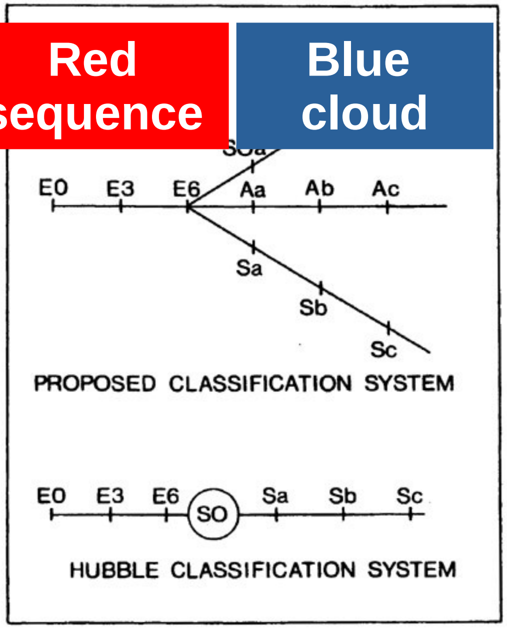


G. and A. De Vaucouleurs (1963)

Van den Bergh (1976)
''trifork'' diagram

Red sequence

Blue cloud



Bimodal distribution of galaxies

What about spectral classification ?

Visual inspection

Morgan & Mayall (1957), Veron (1976), Kennicut (1992), Dobos et al. (2012), Wang et al. (2017) ...

Automatic inspection (Machine Learning)

Sanchez Almeida et al . (2010) : 700 000 (SDSS DR7) using K-means method

De et al. (2016) : Canopy clustering algorithm

… This talk

Not much has been done

Our goal

Make a classification of spectra of galaxies at different redshifts to :

- Improve galaxy classification across cosmic times by **constraining their physics**
- Find **unknown classes** of galaxy spectra ➡ Refine the bimodal distribution of galaxies

Our goal

Make a classification of spectra of galaxies at different redshifts to :

- Improve galaxy classification across cosmic times by **constraining their physics**
- Find **unknown classes** of galaxy spectra → Refine the bimodal distribution of galaxies

→ No preconceived physical model – **No** human **bias**

Our goal

Make a classification of spectra of galaxies at different redshifts to :

- Improve galaxy classification across cosmic times by **constraining their physics**
- Find **unknown classes** of galaxy spectra → Refine the bimodal distribution of galaxies

→ No preconceived physical model – **No** human **bias**

→ **Unsupervised approach**

Our goal

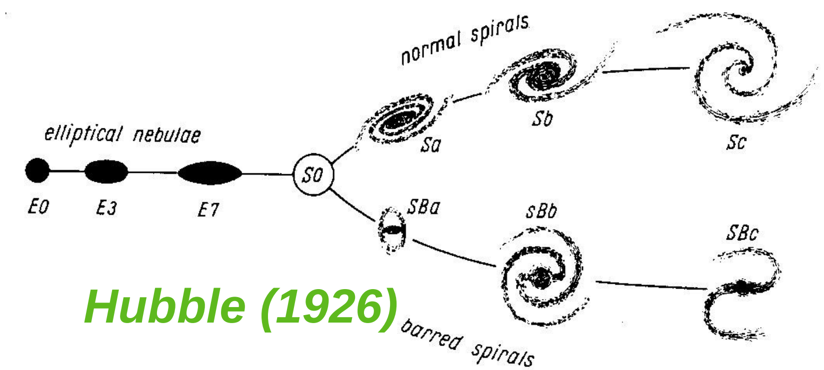
Make a classification of spectra of galaxies at different redshifts to :

- Improve galaxy classification across cosmic times by **constraining their physics**
- Find **unknown classes** of galaxy spectra ➡ Refine the bimodal distribution of galaxies

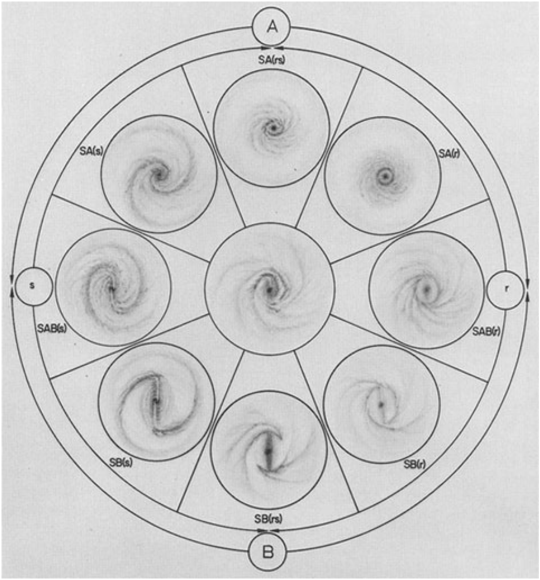
➡ No preconceived physical model – **No** human **bias**

➡ **Unsupervised approach**

Just like ...



Hubble (1926)



G. & A. De Vaucouleurs (1963)

Our goal

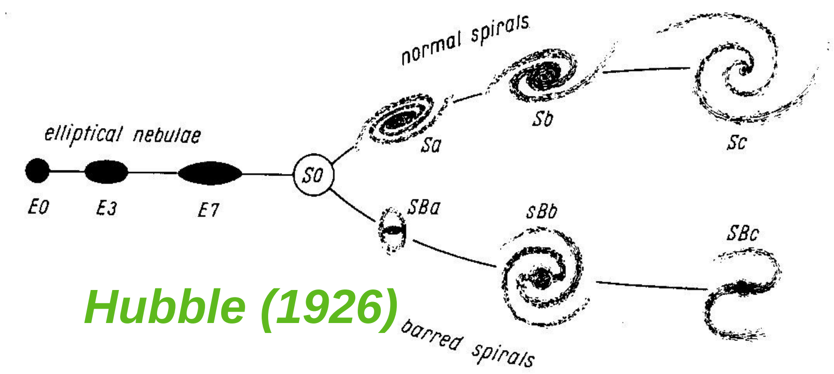
Make a classification of spectra of galaxies at different redshifts to :

- Improve galaxy classification across cosmic times by **constraining their physics**
- Find **unknown classes** of galaxy spectra → Refine the bimodal distribution of galaxies

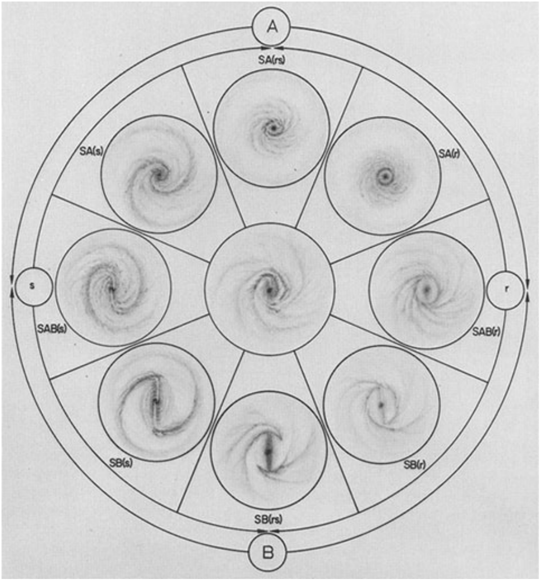
→ No preconceived physical model – **No** human **bias**

→ **Unsupervised approach**

Just like ...



Hubble (1926)

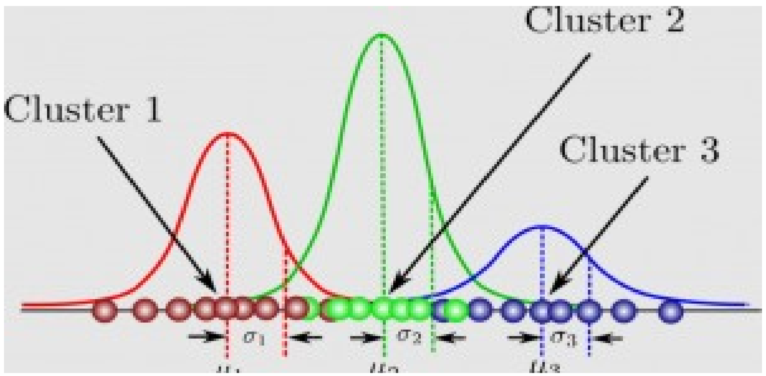


G. & A. De Vaucouleurs (1963)

→ Derive **templates** of galaxy **spectra** that may be used as training templates for supervised approaches of ML ...

Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301



Gaussian Mixture Model (GMM)

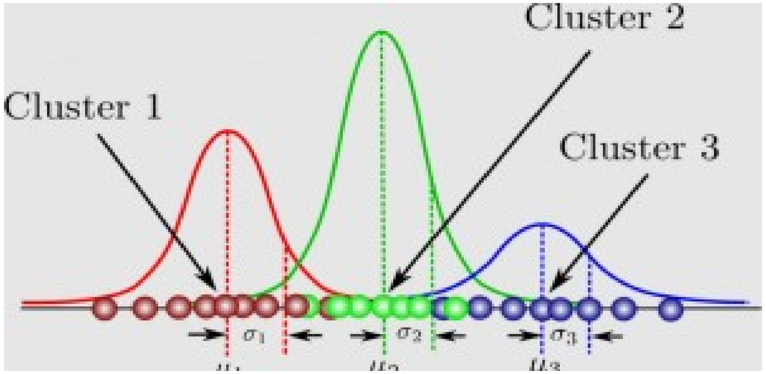
Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301

First guess (**GMM**)

Number of classes **K**

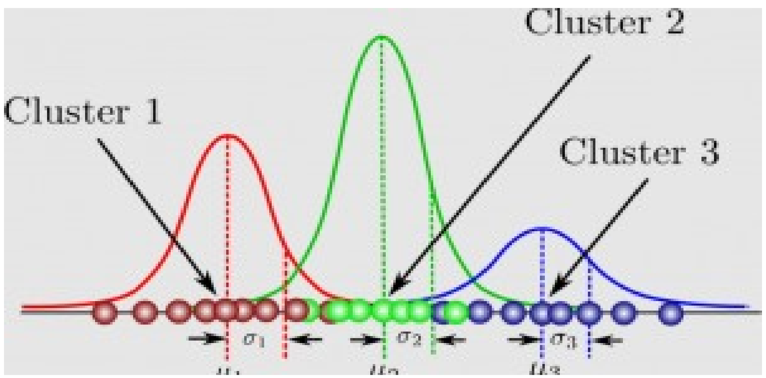
Input



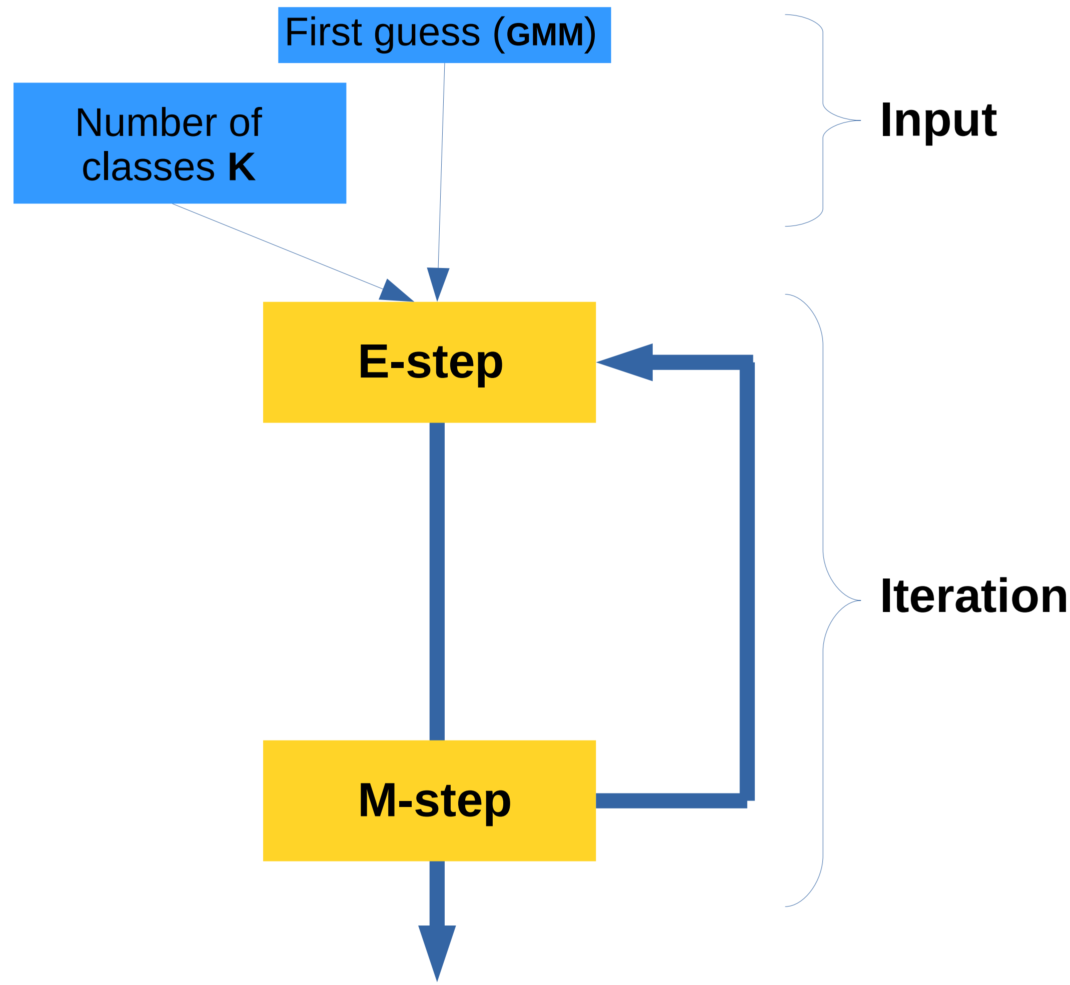
Gaussian Mixture Model (GMM)

Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301

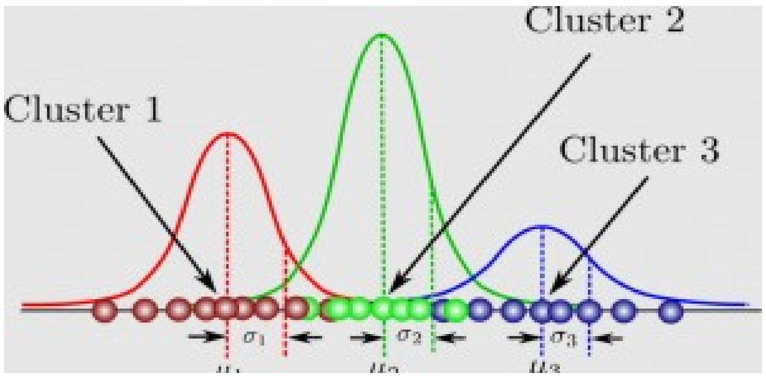


Gaussian Mixture Model (GMM)

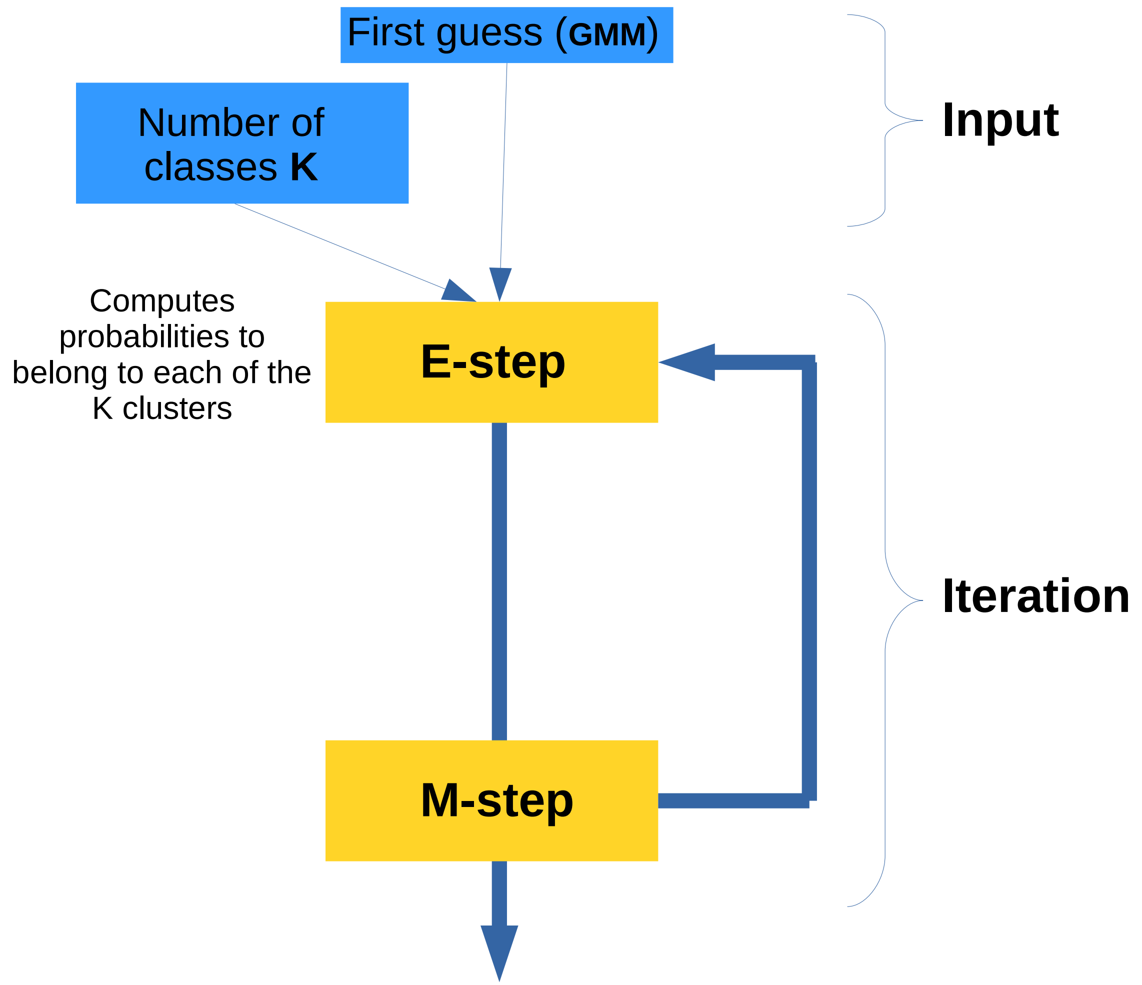


Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301

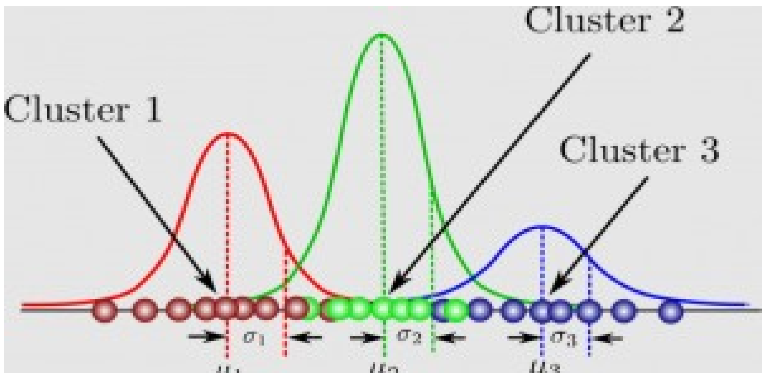


Gaussian Mixture Model (GMM)

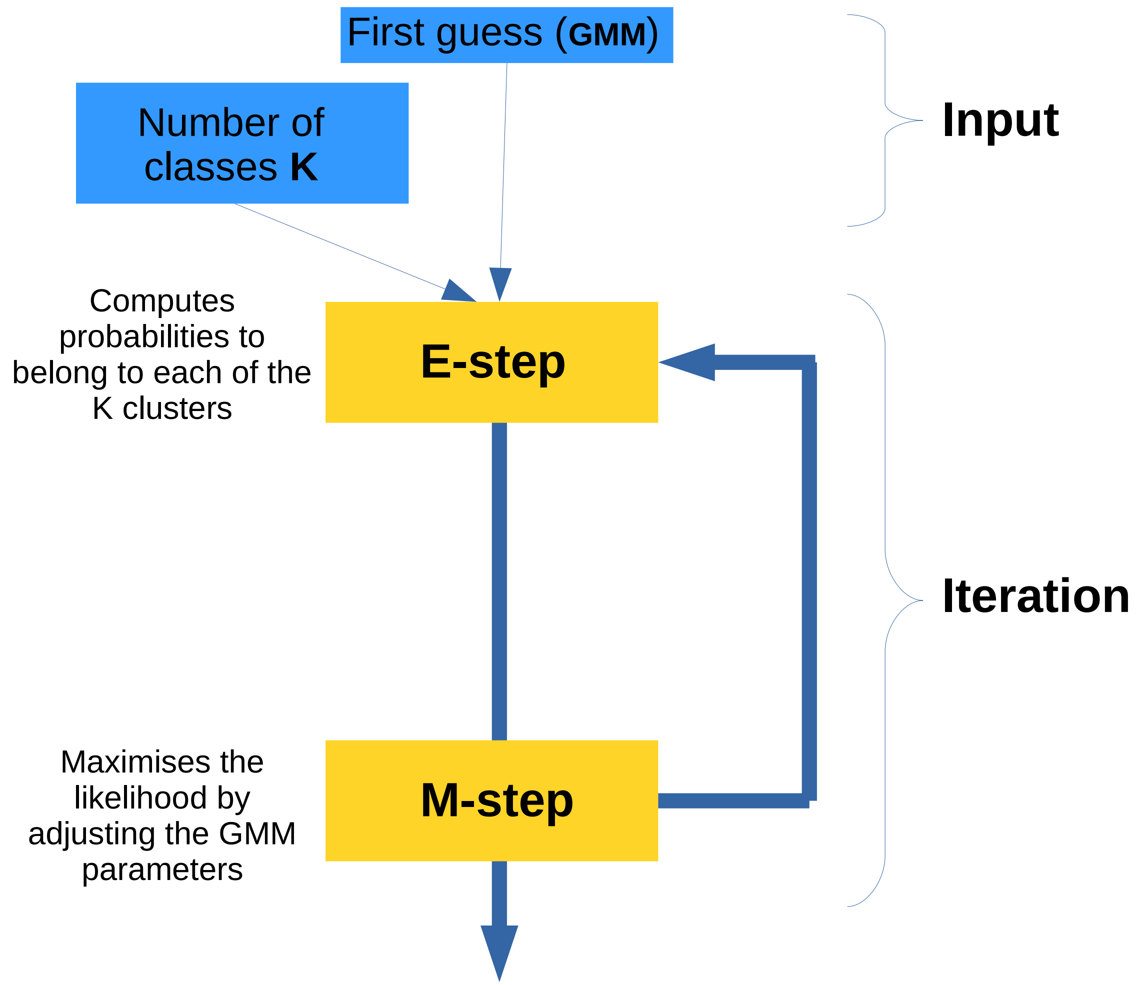


Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301

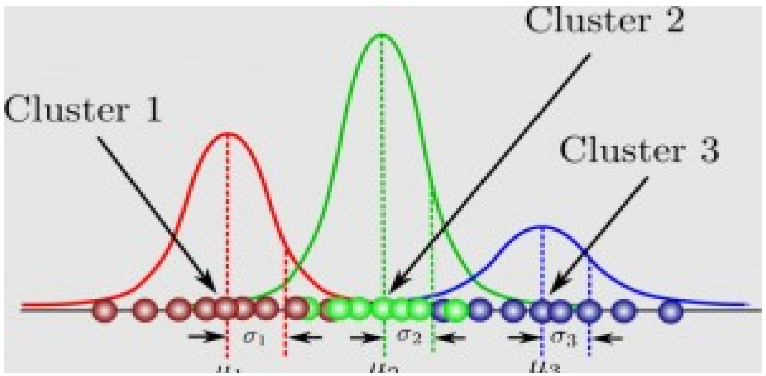


Gaussian Mixture Model (GMM)

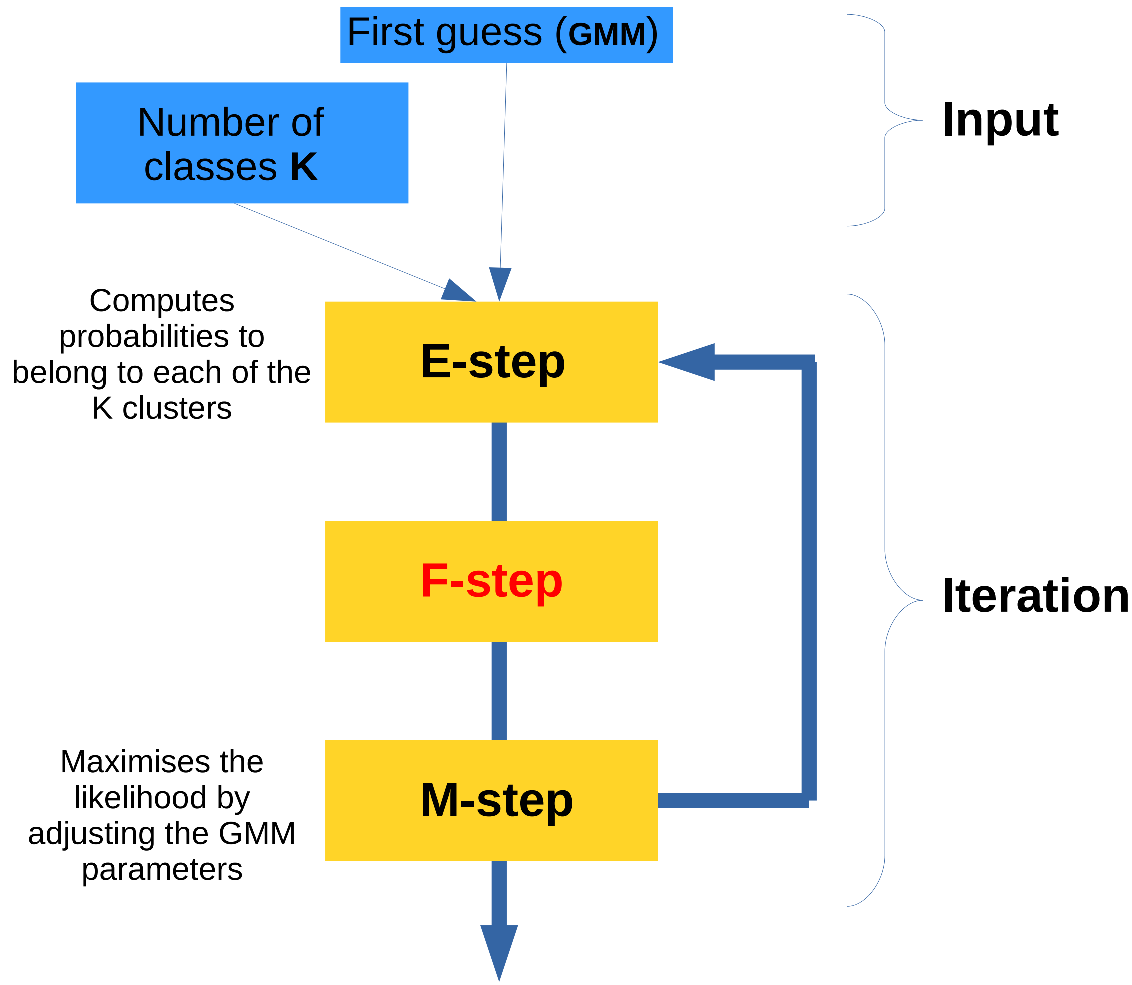


Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301

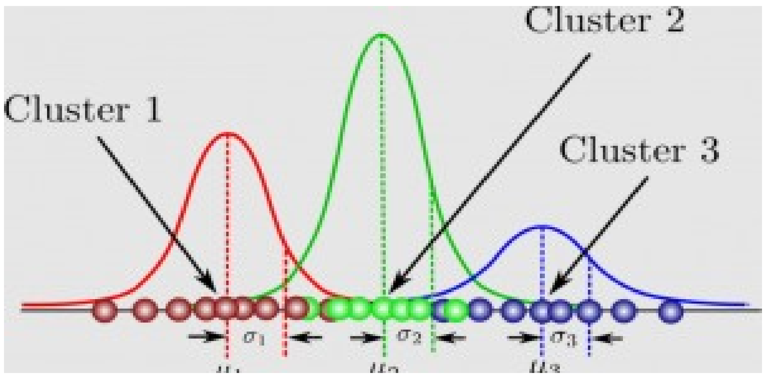


Gaussian Mixture Model (GMM)

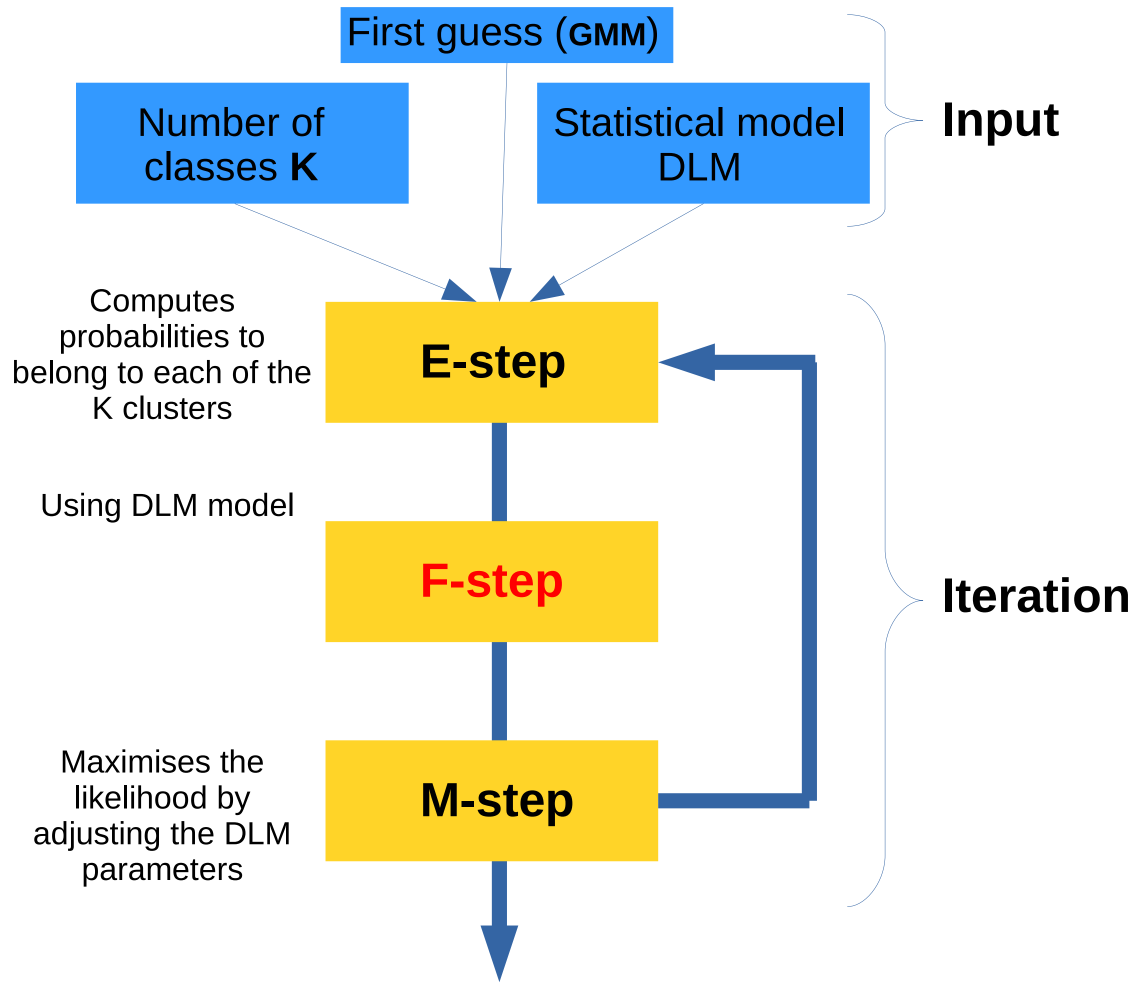


Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301

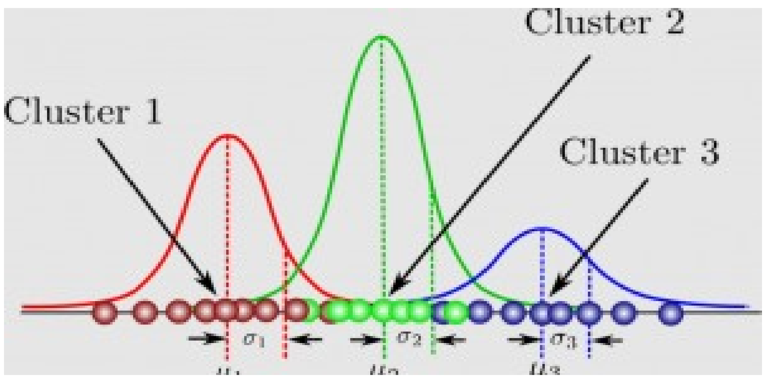


Gaussian Mixture Model (GMM)



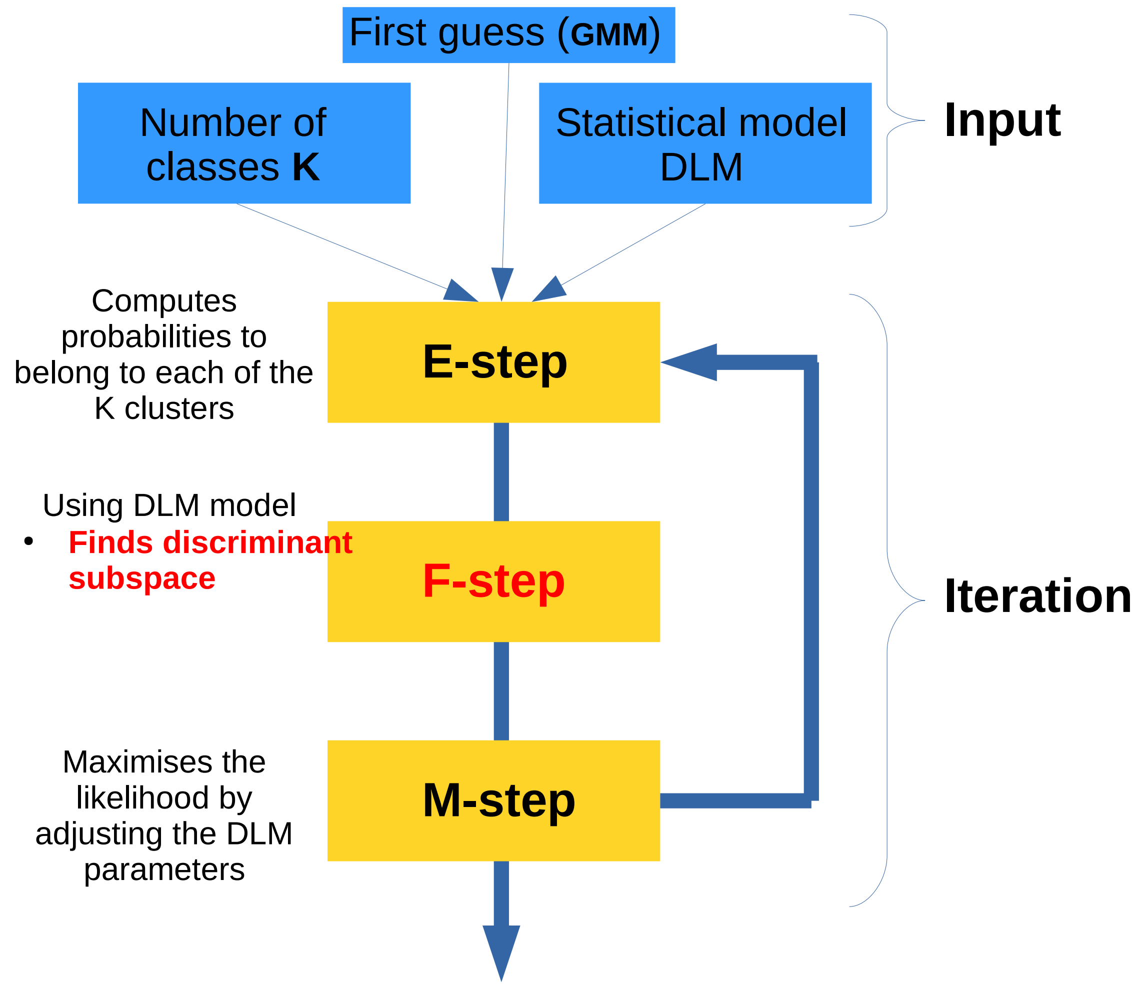
Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301



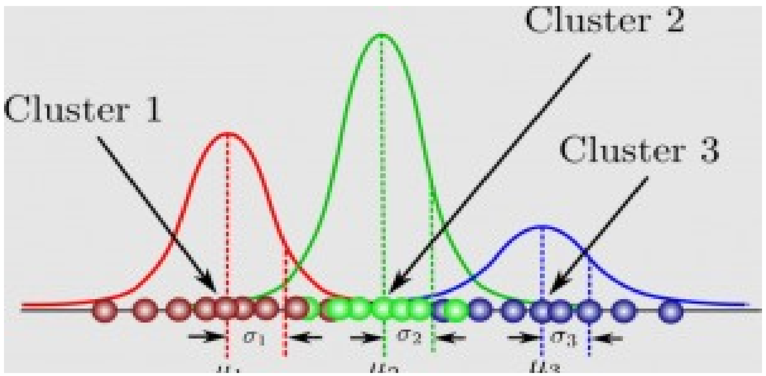
Gaussian Mixture Model (GMM)

in a subspace of smaller dimension than phys. parameter space, the « latent subspace »



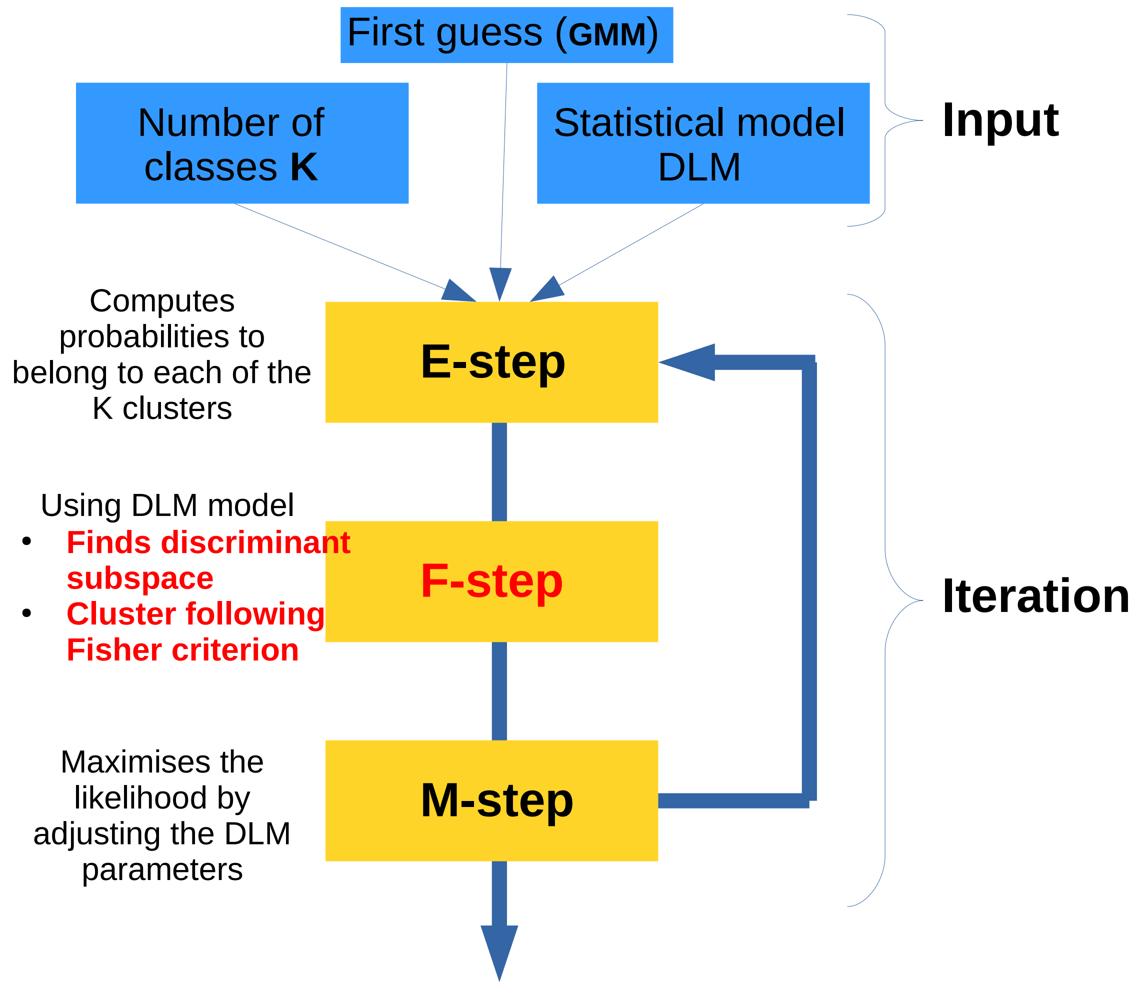
Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301



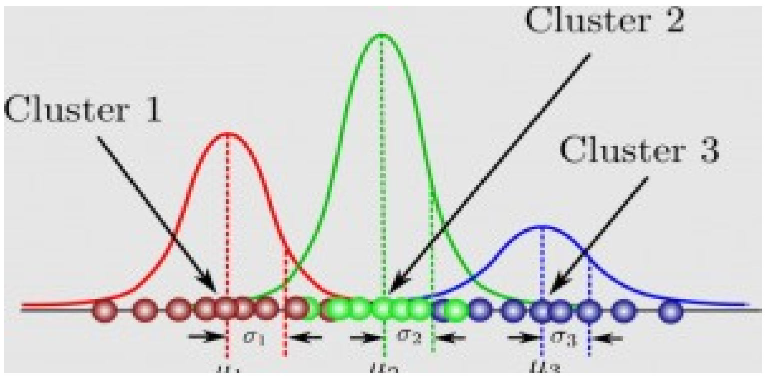
Gaussian Mixture Model (GMM)

in a subspace of smaller dimension than phys. parameter space, the « latent subspace »



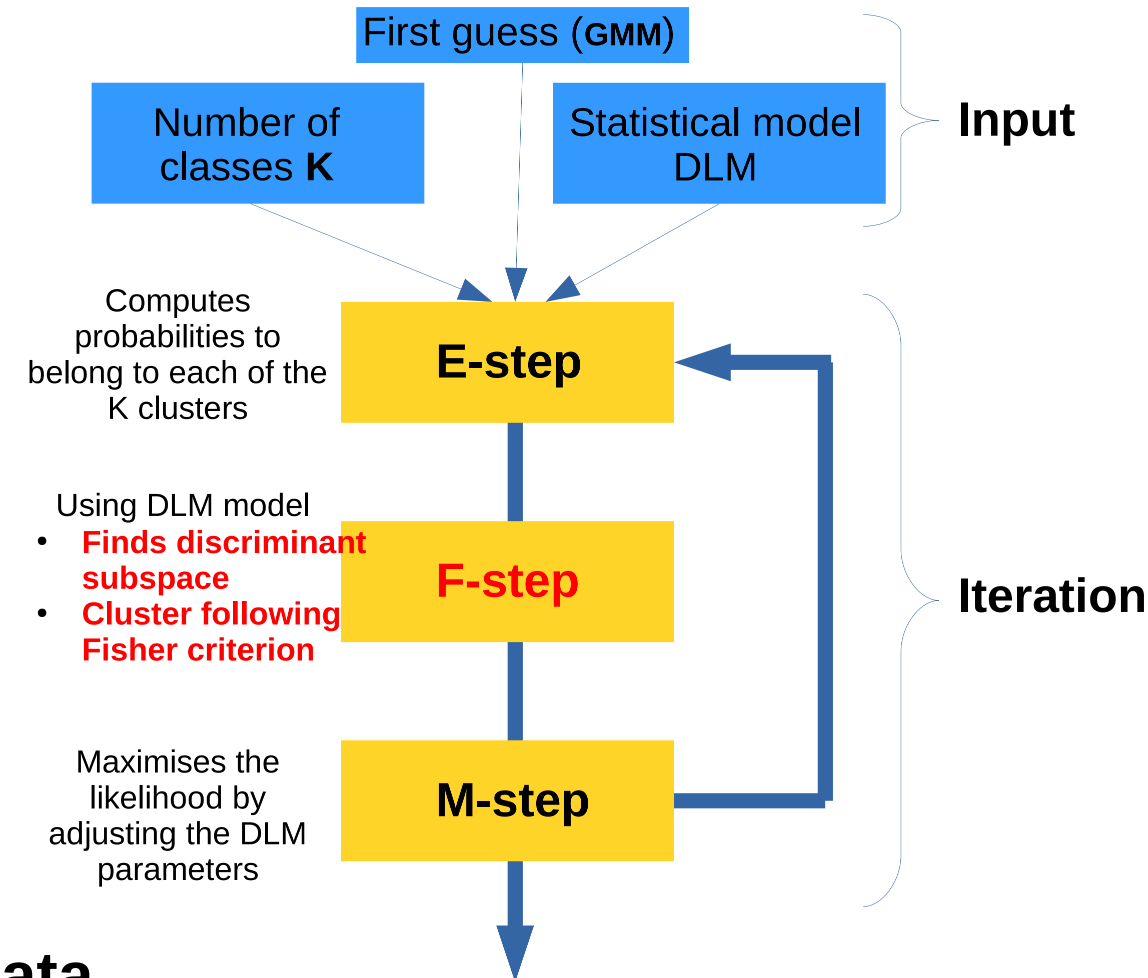
Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301



Gaussian Mixture Model (GMM)

in a subspace of smaller dimension than phys. parameter space, the « latent subspace »

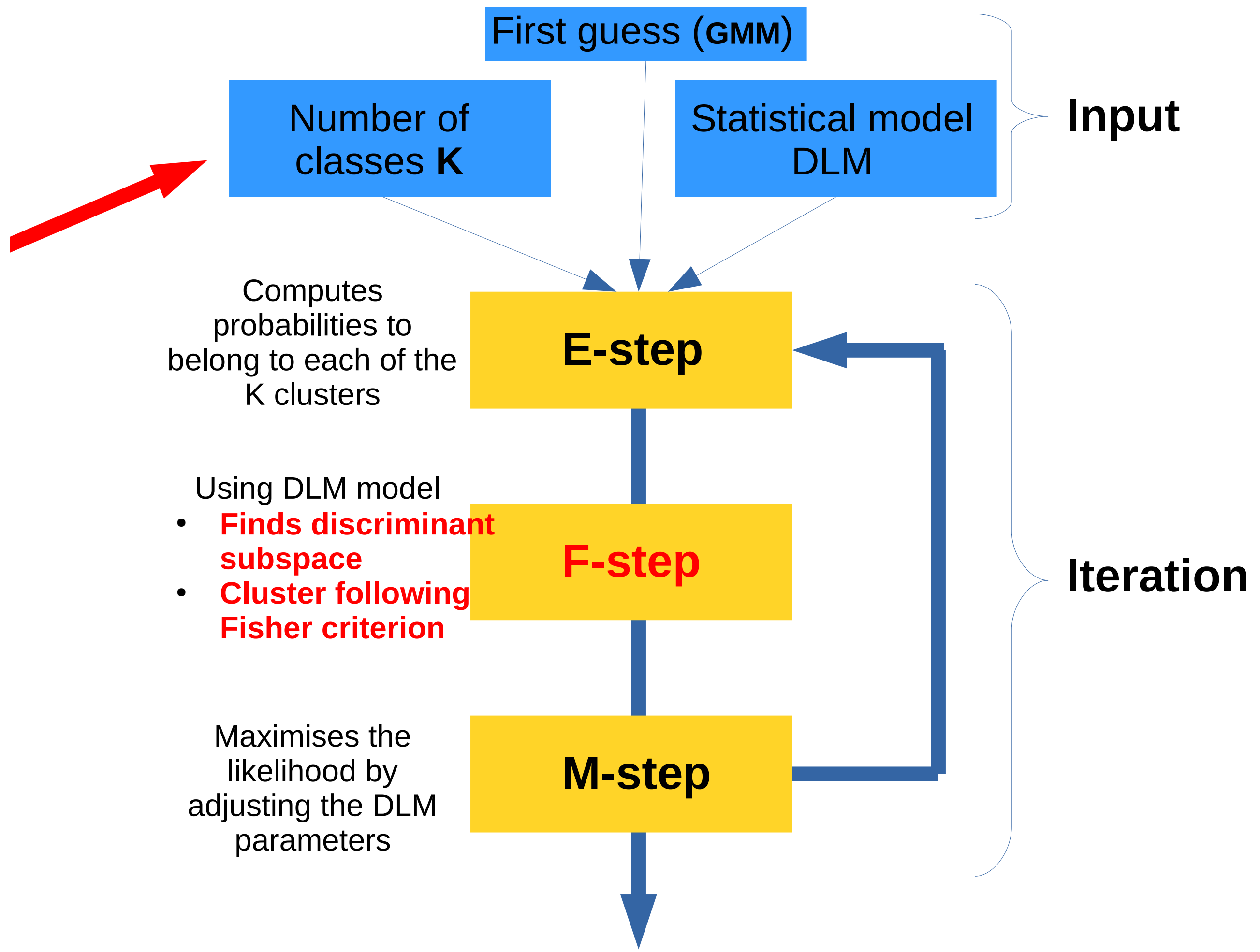


High-dimensional data

Fisher-EM algorithm

Bouveyron, C., & Brunet, C. 2012a, Stat. Comput., 22, 301

The best DLM model and optimal K are those maximising the Integrated Completed Likelihood (ICL) statistical criterion



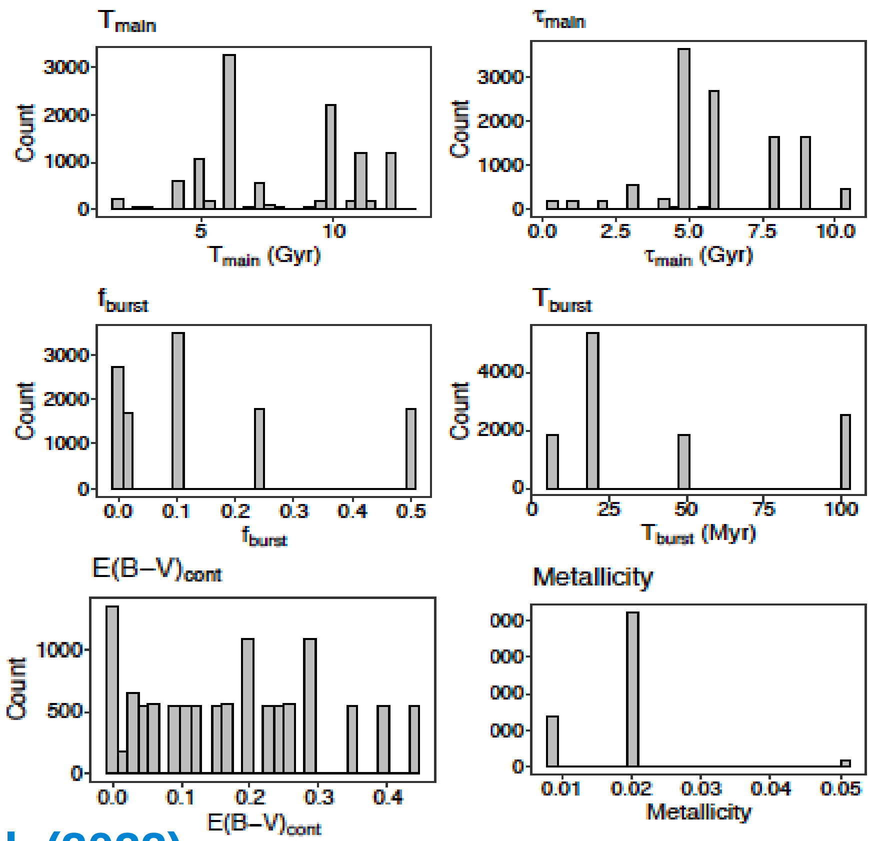
Fisher-EM on simulated spectra

Optical simulated spectra of 11000 galaxies with CIGALE ***Burgarella et al. (2005)***

SFH : Main population and a later burst (same metallicity)

➡ **6 physical parameters :**

- **Age** & **e-folding time** of the main SP (Myr)
- **Age** & stellar **mass fraction** of the burst (Myr)
- **Metallicity**
- **Reddening**



Dubois et al. (2022)

Distribution of parameters

Fisher-EM on simulated spectra

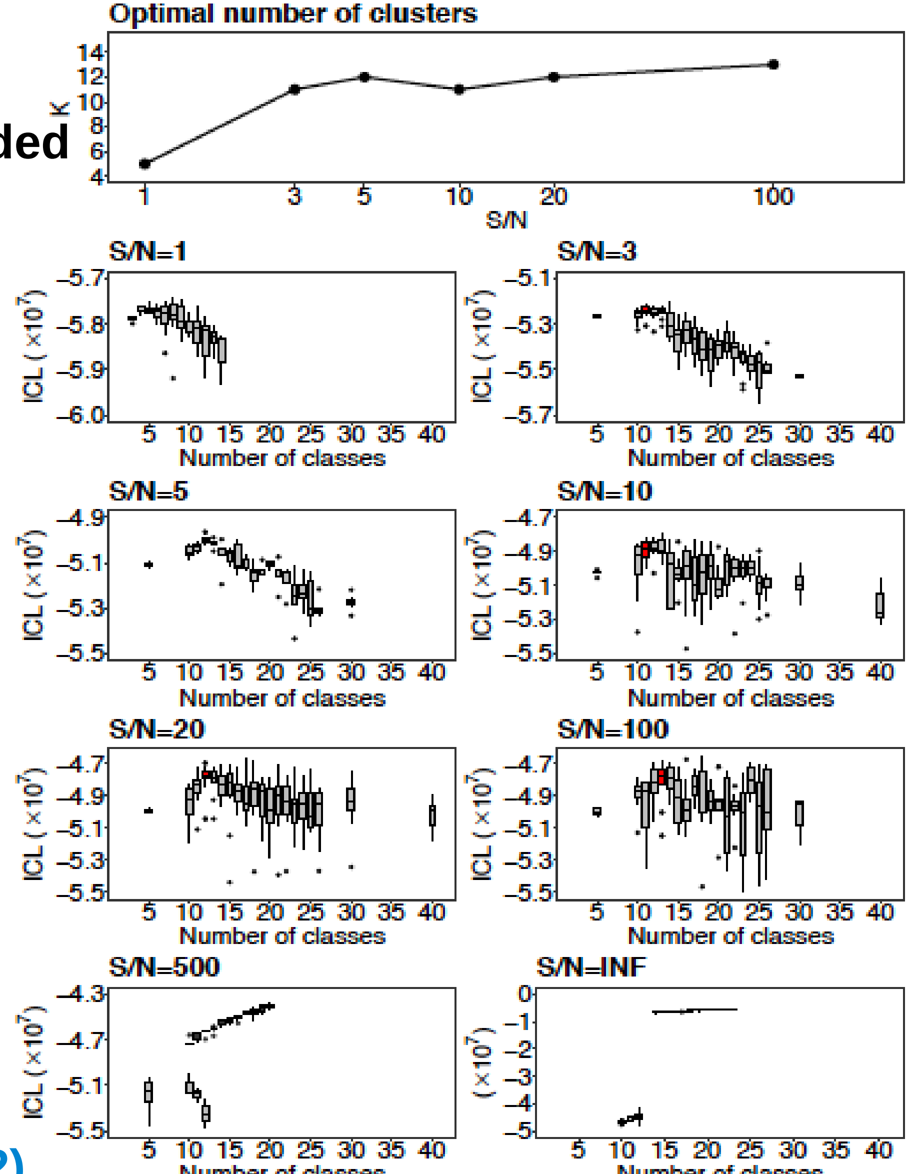
- **Classification of sample without noise**
- **Classifications of samples with different added noise levels (S/N = 1, 3, 5, 10, 20, 100, 500)**

For SNR < 500 :

→ Best solution for K=12 (**Robust** results)
→ Well **discriminated phys. parameters** (but less discrimination at higher noise)

Convergence issues for SNR >= 500

→ Due to coverage of parameter space, sparsity
→ Noise insufficient to blur data sparsity



Dubois et al. (2022)

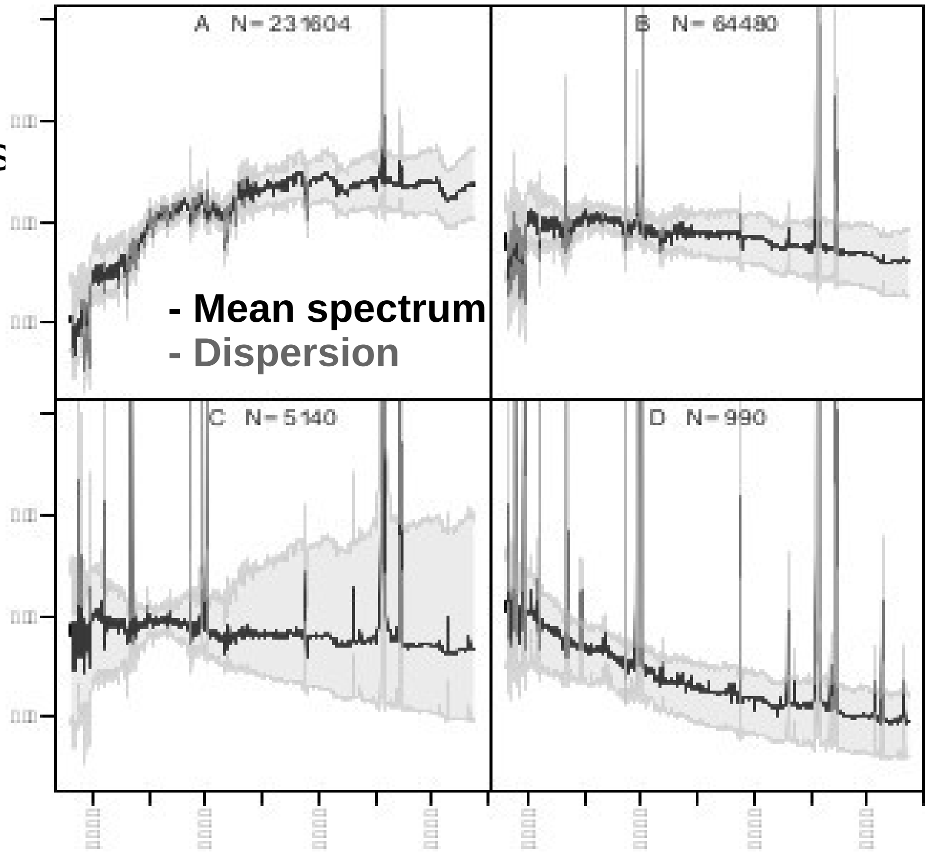
Fisher-EM on SDSS spectra

About 700,000 spectra of galaxies and quasars from SDSS DR7 with z < 0.25 (corrected for redshift, resampled, rebinned and normalised)

Different steps : FEM applied on subsamples

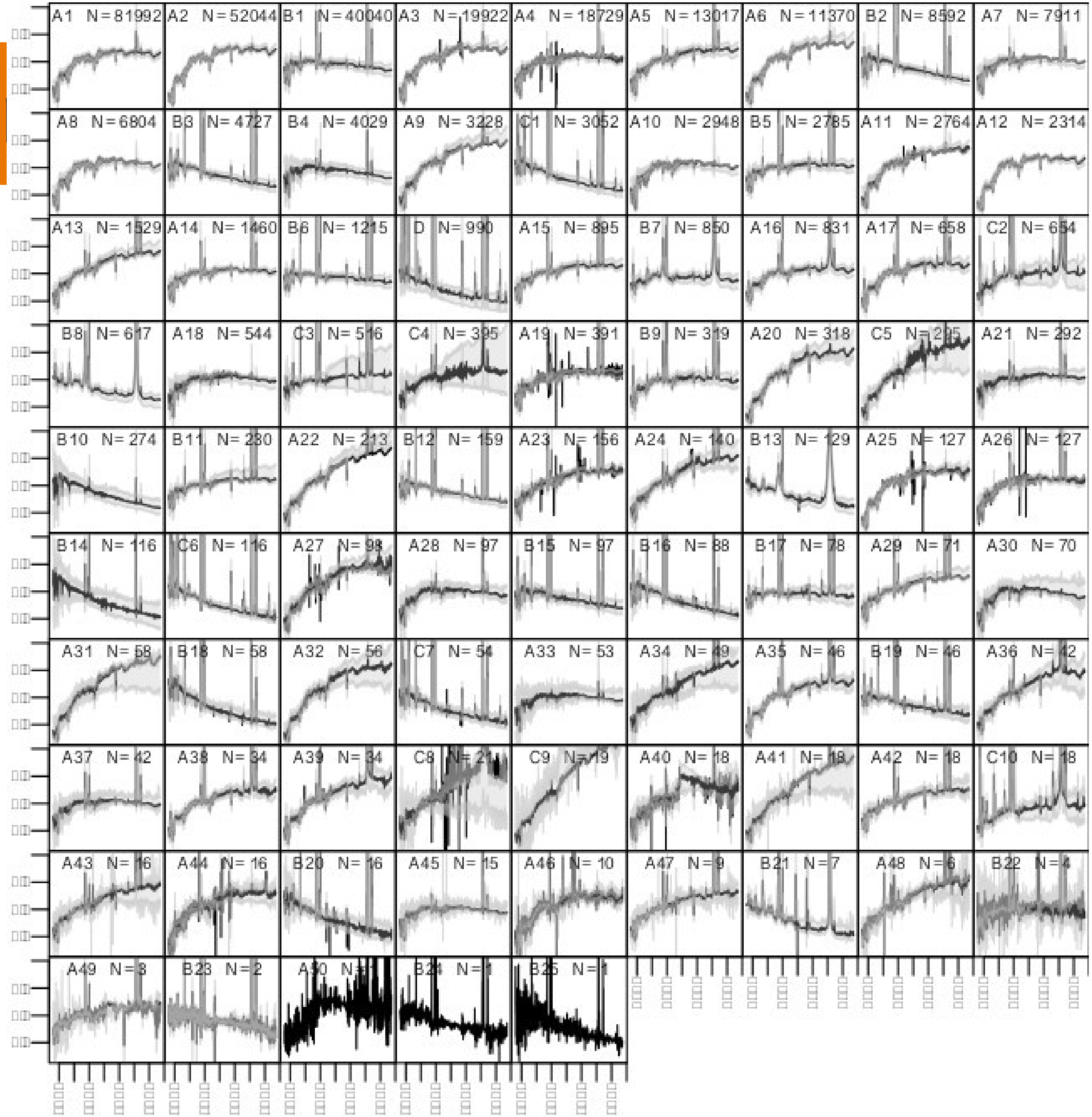
→ 4 main classes : A,B,C,D

→ Subclassification



Fraix-Burnet et al. (2021)

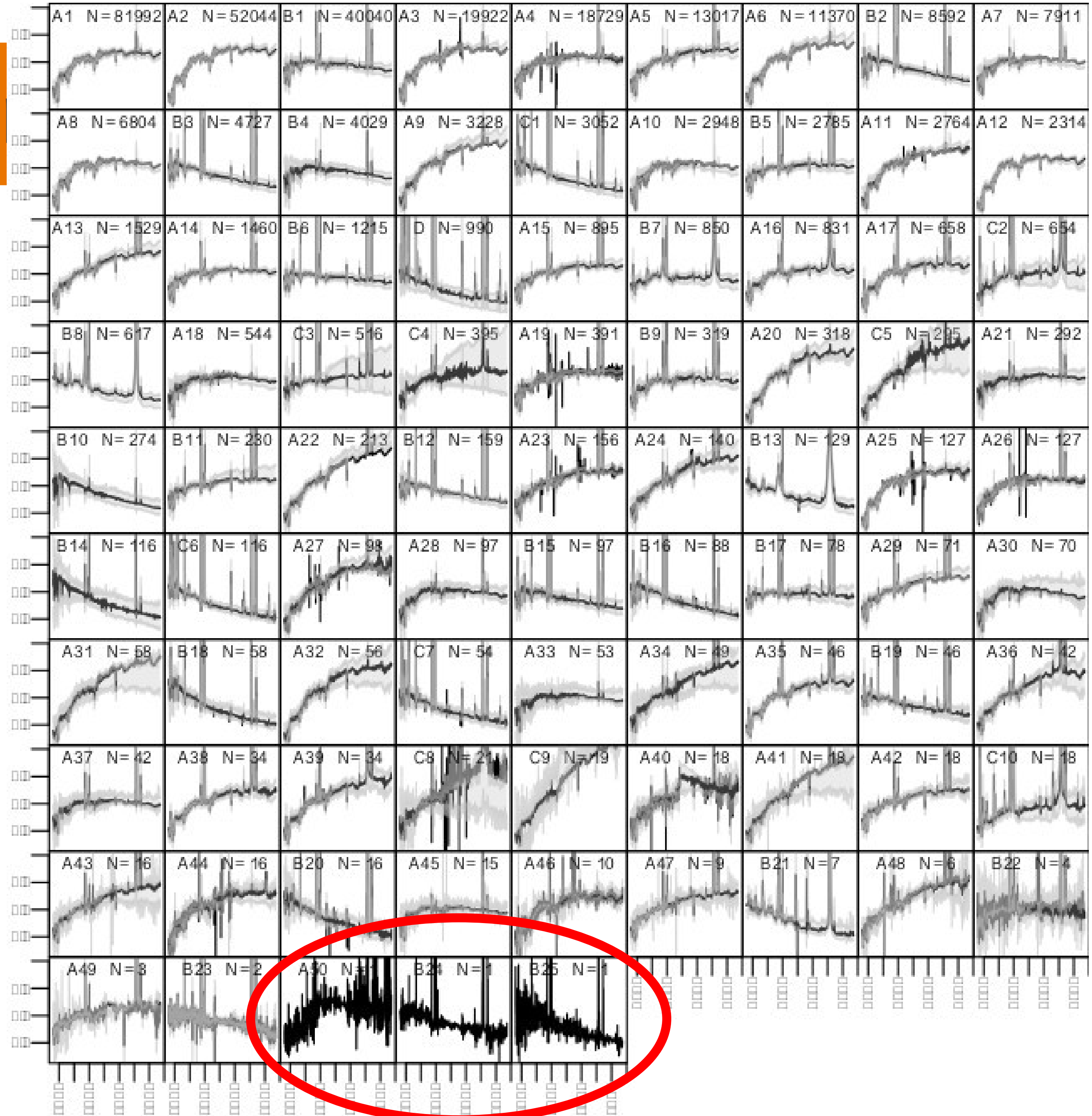
Fishe



86 final classes

Fraix-Burnet et al. (2021)

Fishe

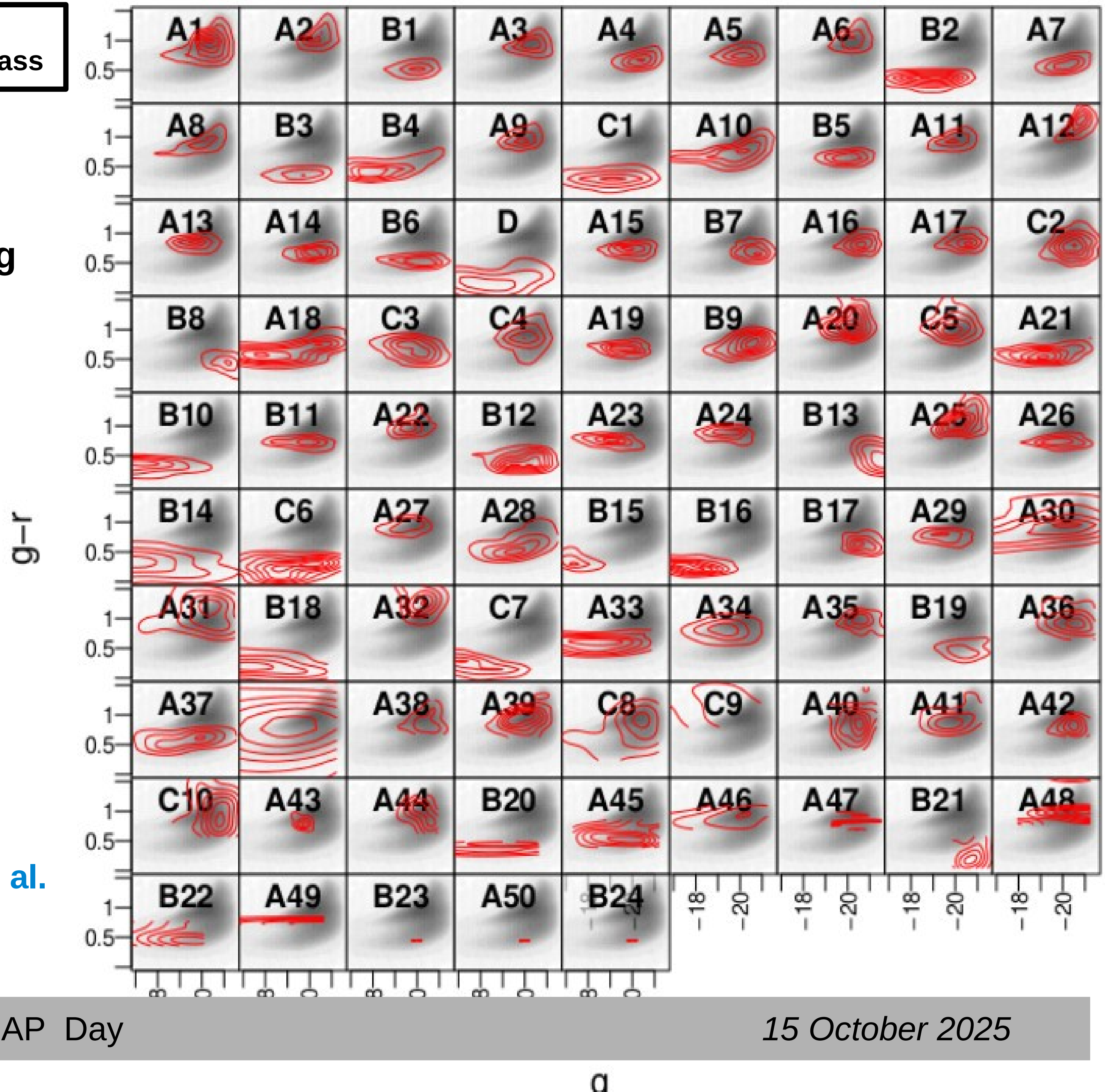


86 final classes

Fraix-Burnet et al. (2021)

Grey: whole sample
Red: spectra of the class

Color-mag diagram

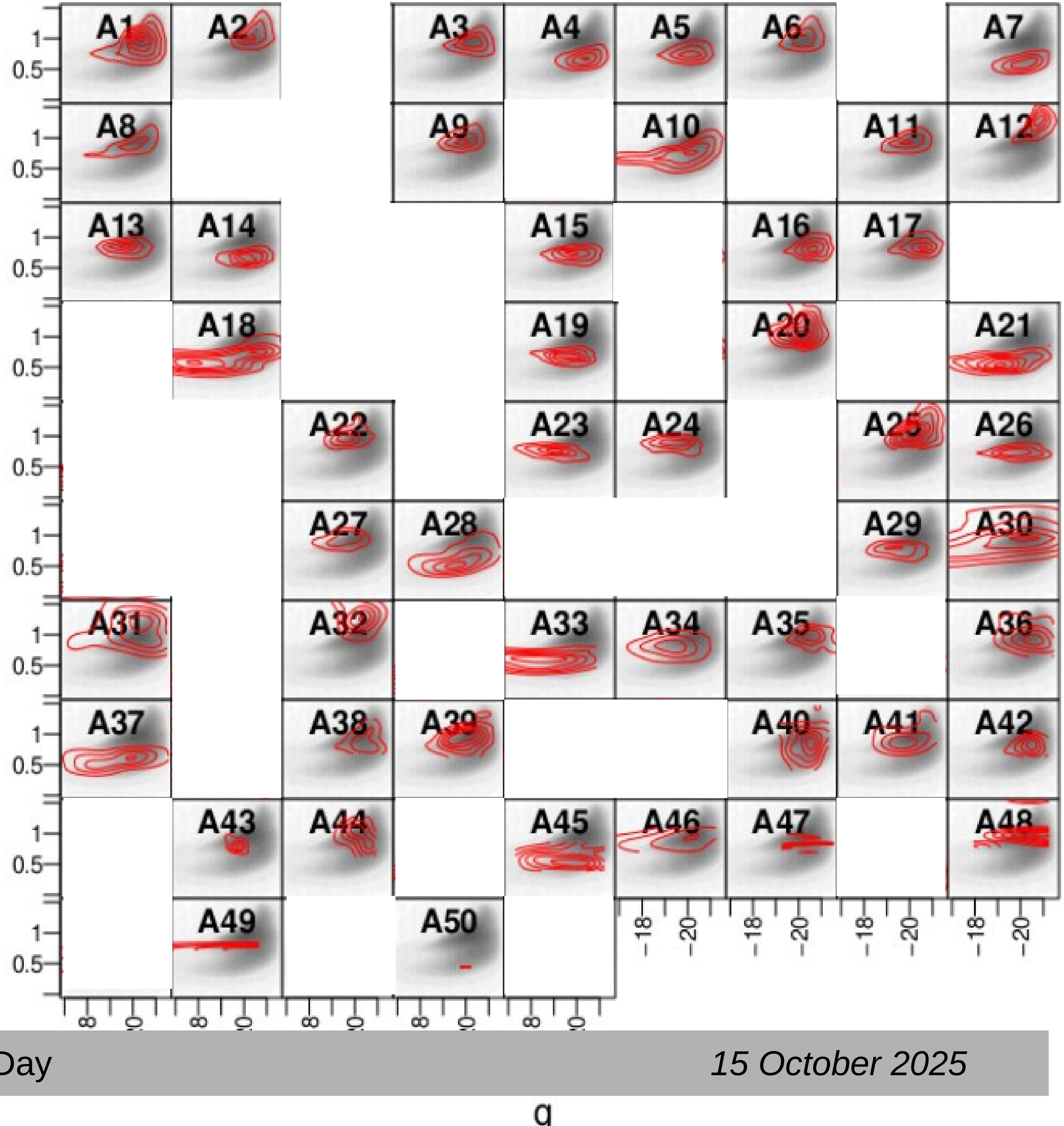


Fraix-Burnet et al. (2021)

Grey: whole sample
Red: spectra of the class

Class A

Red sequence

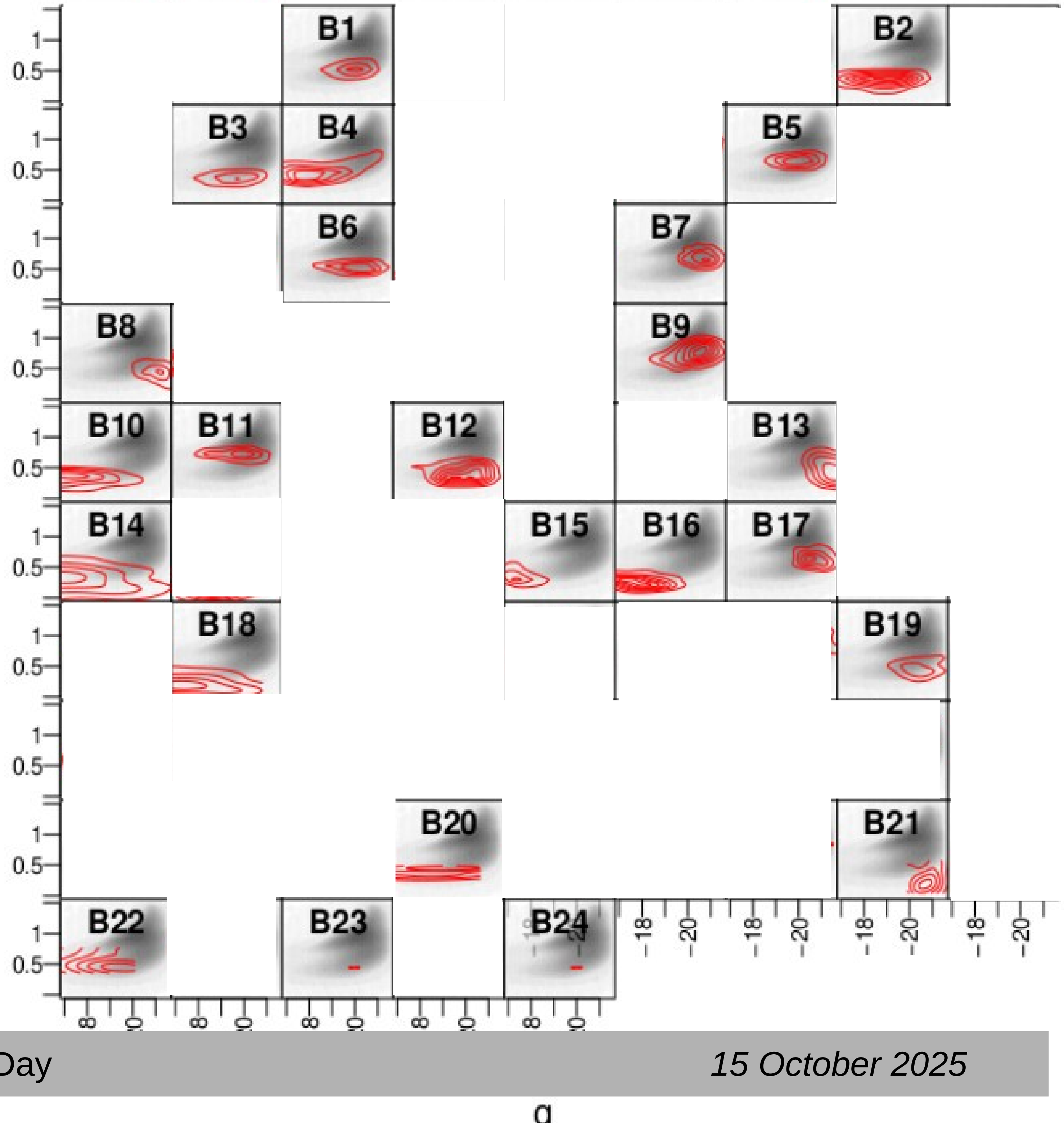


Fraix-Burnet et al. (2021)

Grey: whole sample
Red: spectra of the class

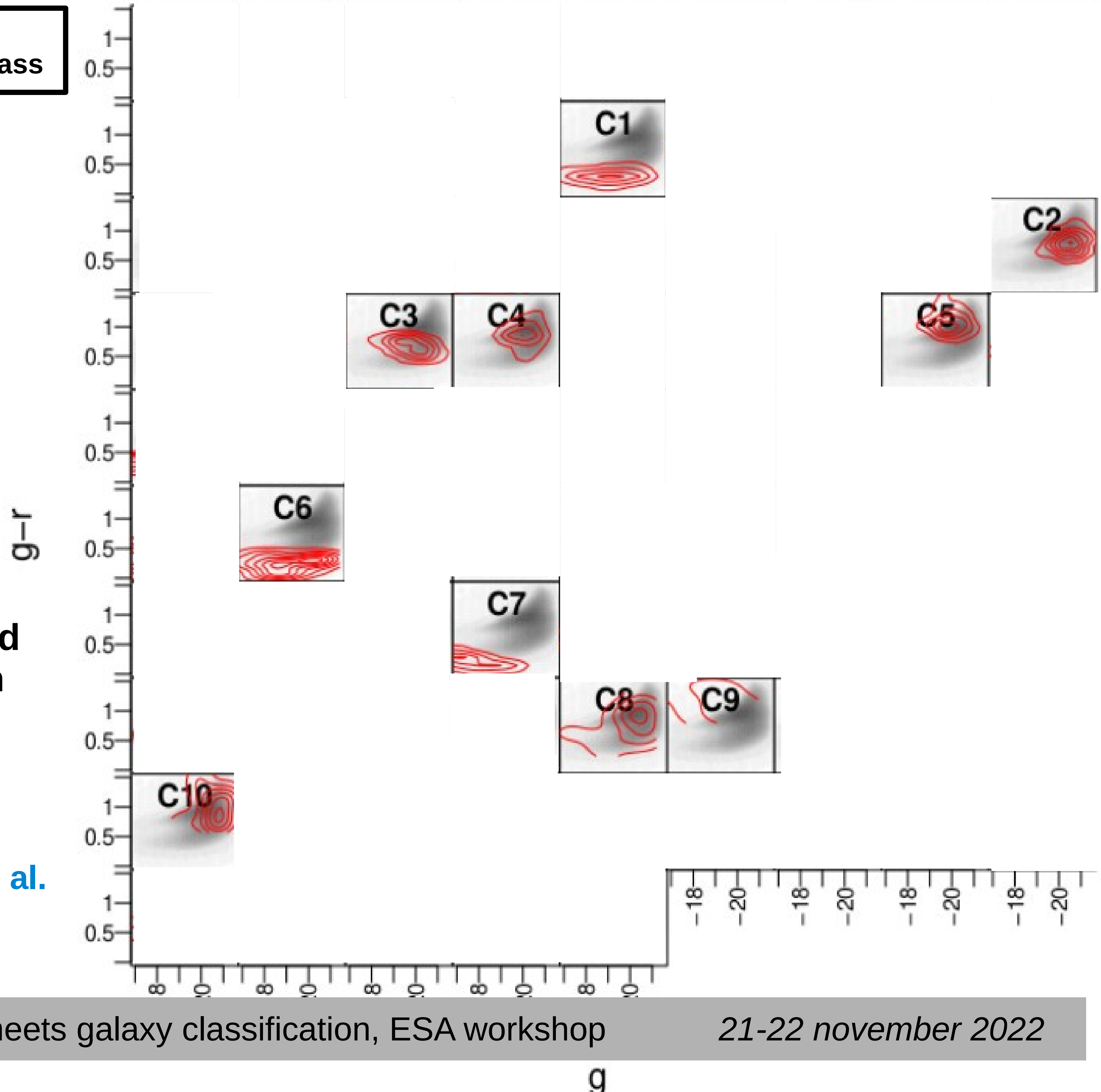
Class B

Blue sequence



Fraix-Burnet et al. (2021)

Grey: whole sample
Red: spectra of the class



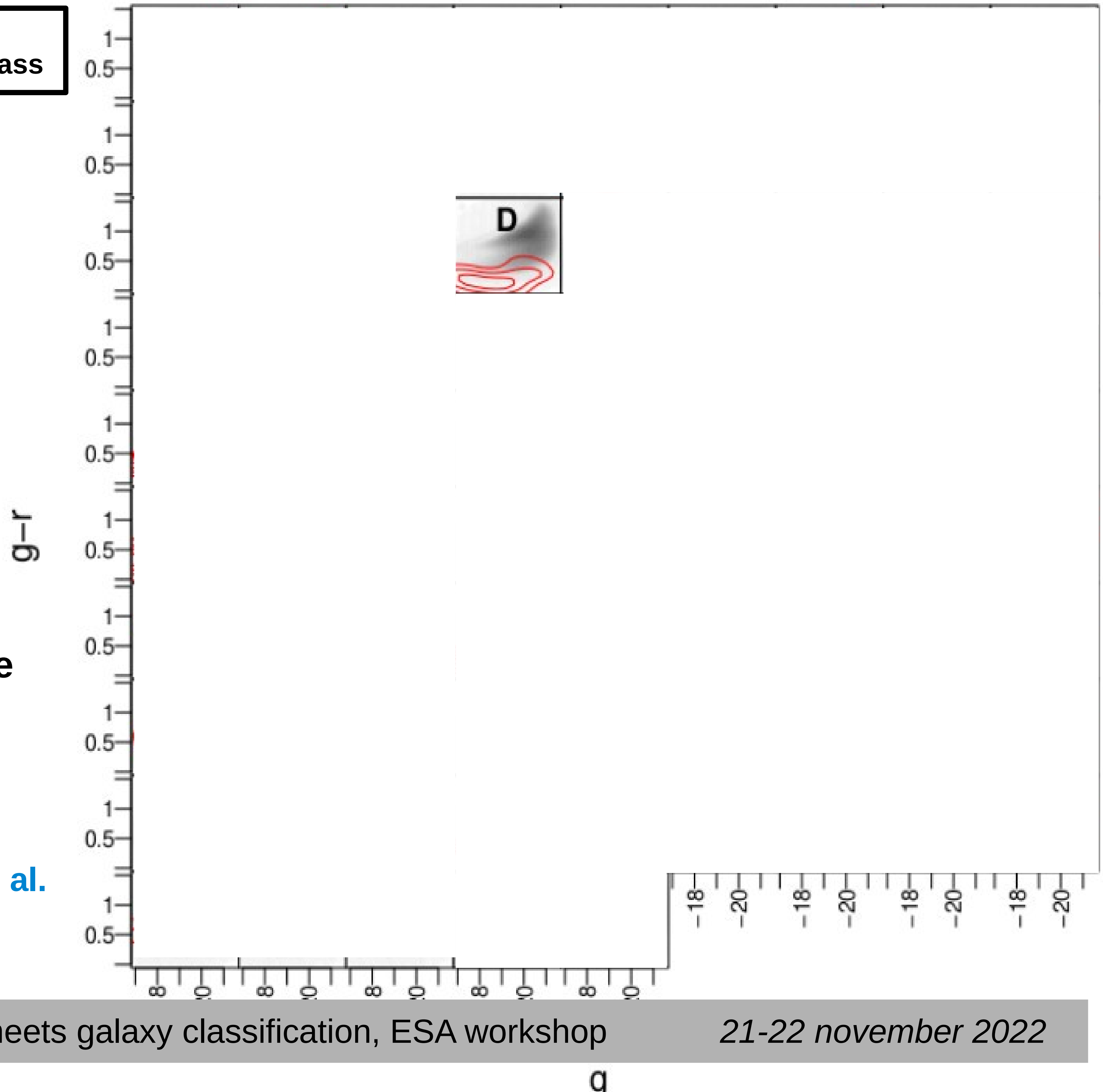
Class C

Diverse

Characterized by emission lines

Fraix-Burnet et al. (2021)

Grey: whole sample
Red: spectra of the class



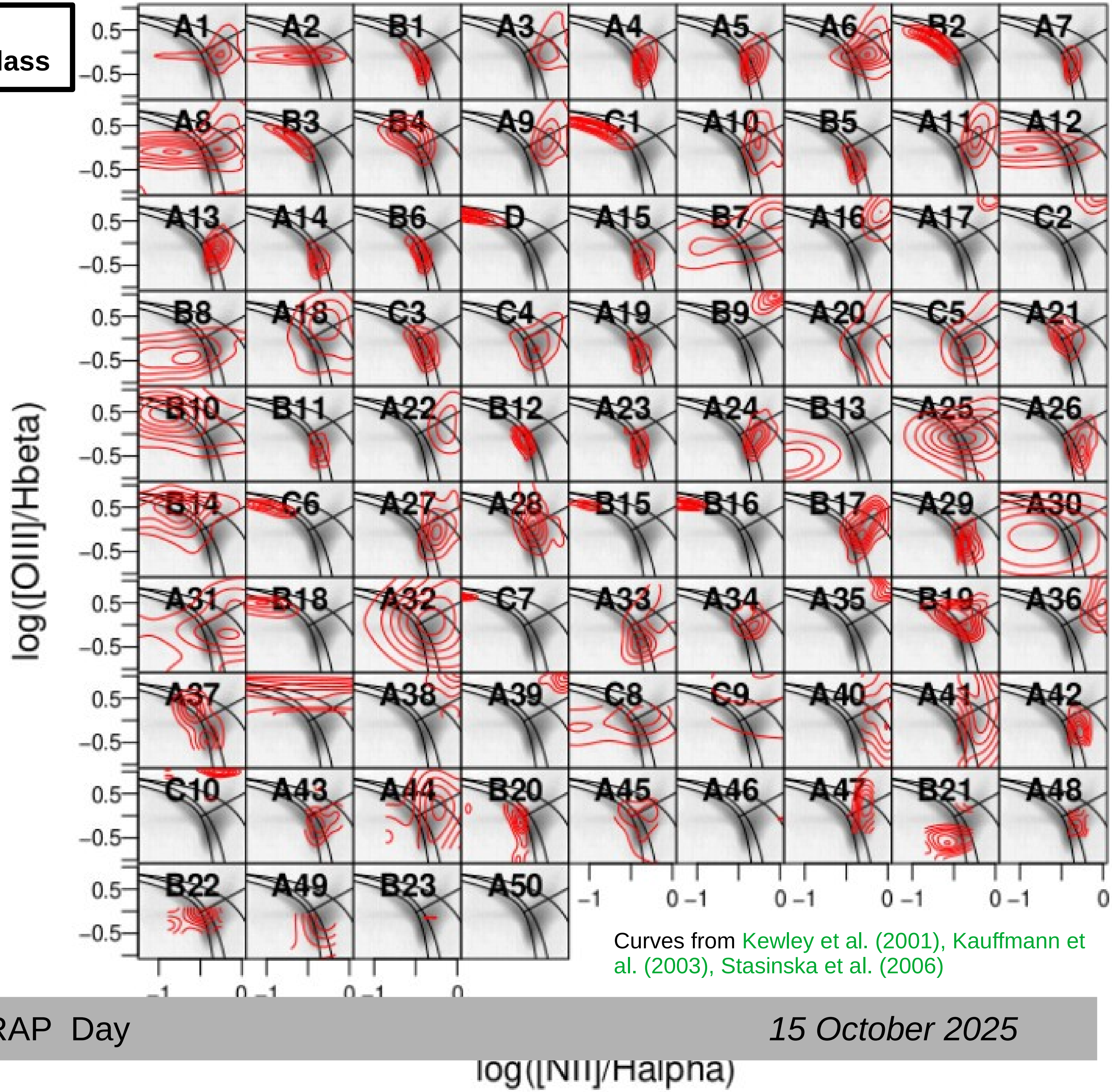
Class D

Extremely blue

Fraix-Burnet et al. (2021)

Grey: whole sample
Red: spectra of the class

BPT diagrams



Curves from Kewley et al. (2001), Kauffmann et al. (2003), Stasinska et al. (2006)

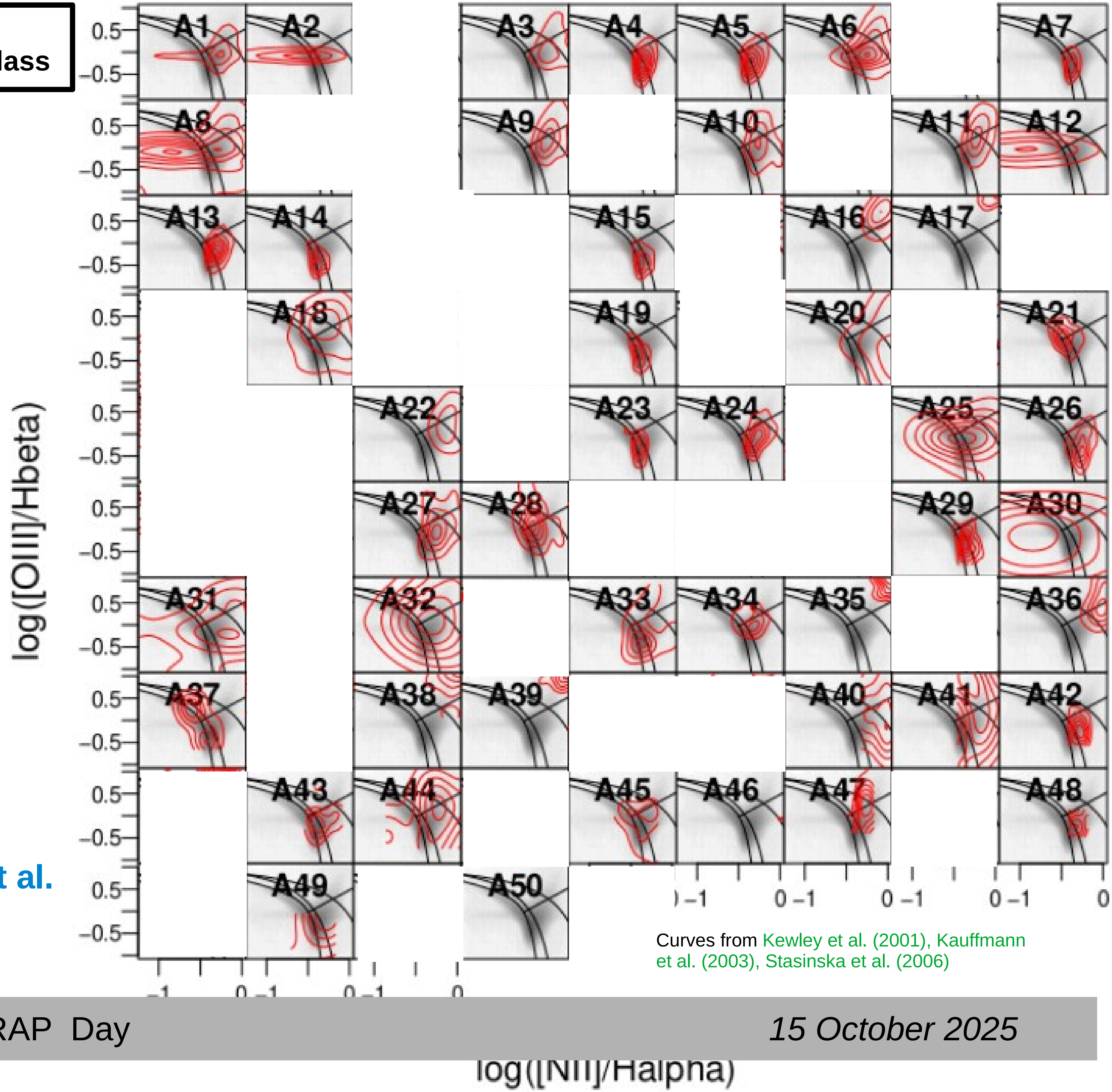
Grey: whole sample
Red: spectra of the class

BPT diagrams

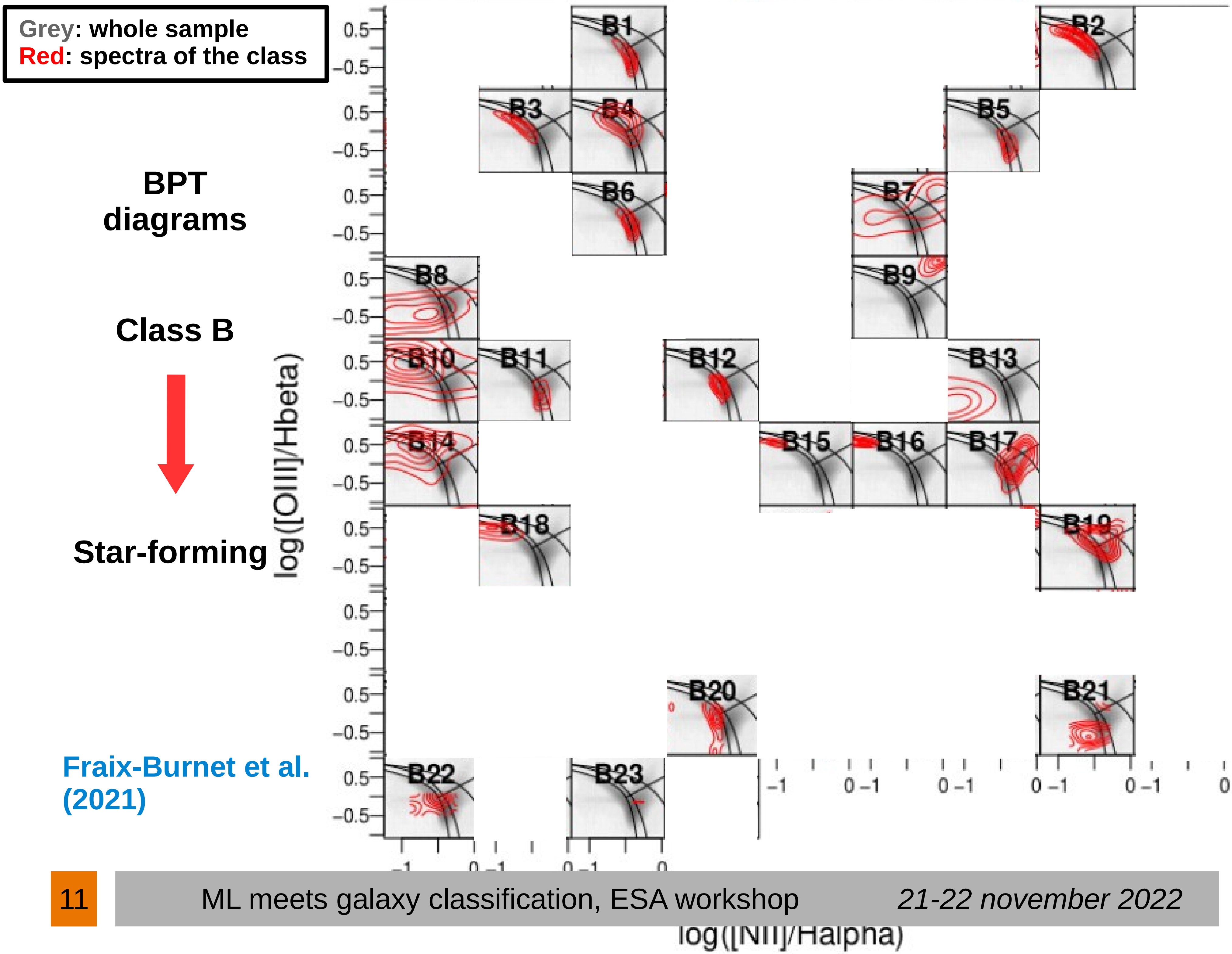
Class A

Seyferts or LINER galaxies

Fraix-Burnet et al. (2021)



Curves from Kewley et al. (2001), Kauffmann et al. (2003), Stasinska et al. (2006)



BPT diagrams

Class B

Star-forming

Fraix-Burnet et al. (2021)

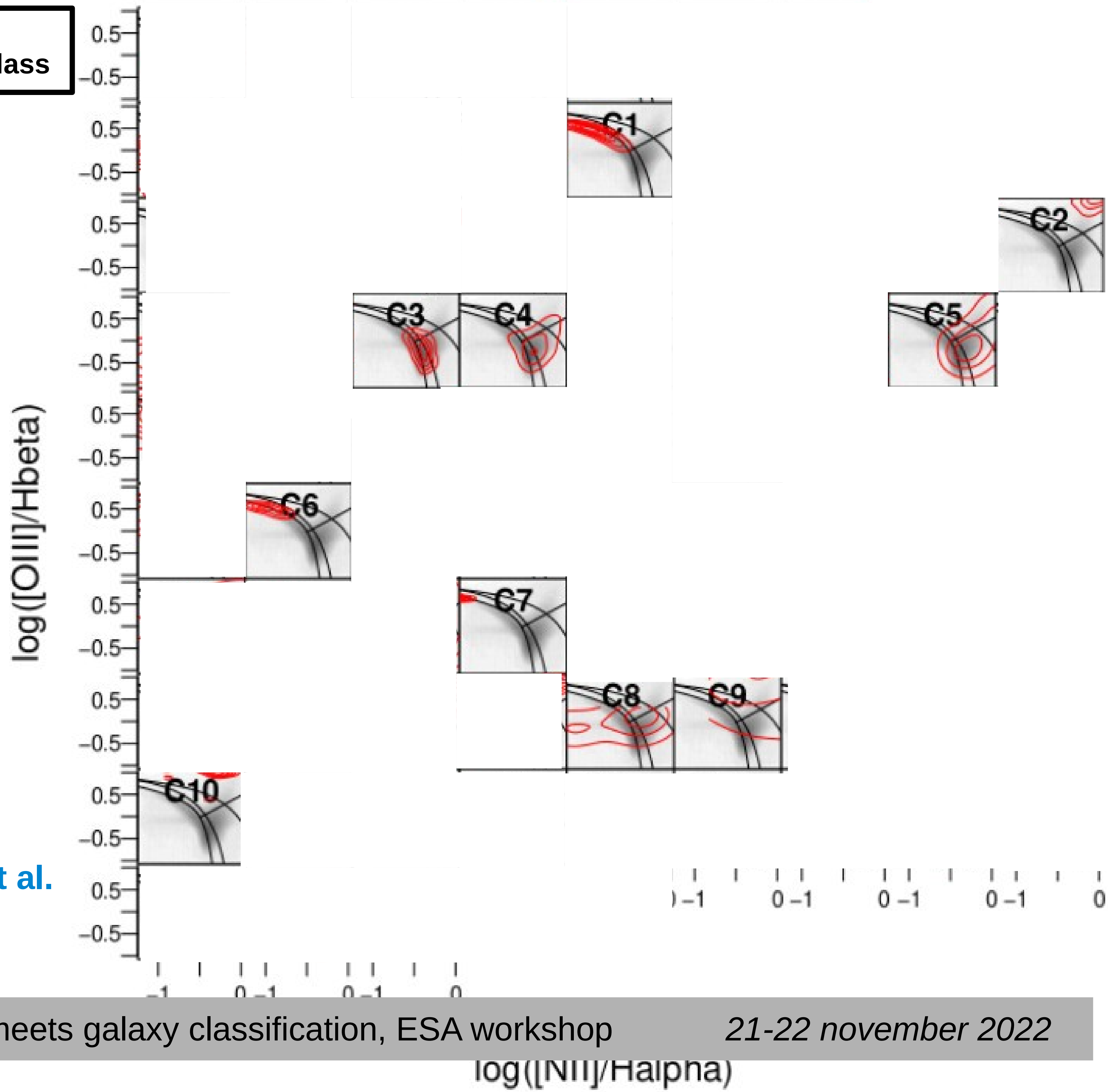
Grey: whole sample
Red: spectra of the class

BPT diagrams

Class C

Star formation & Nuclear activity

Fraix-Burnet et al. (2021)



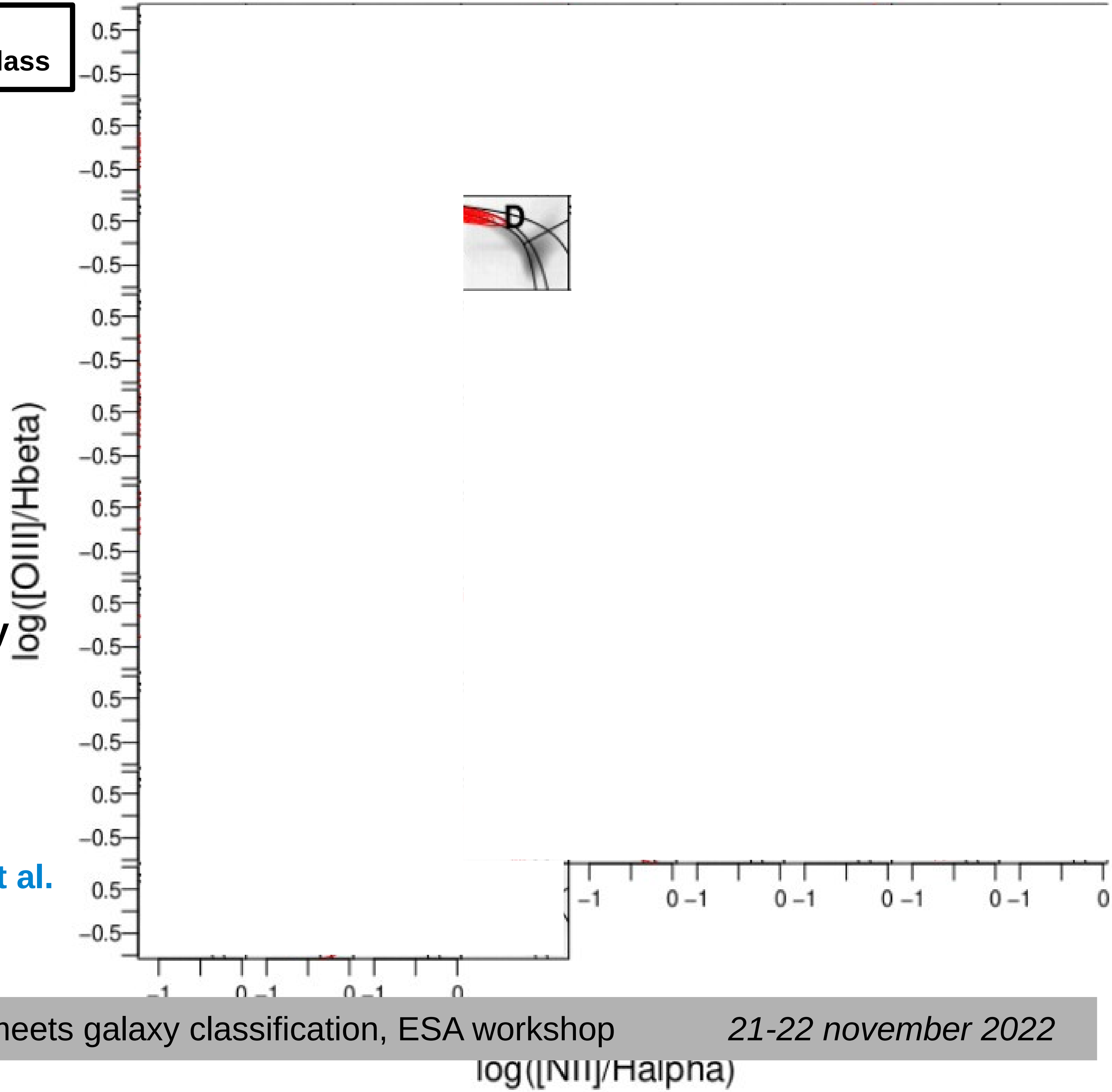
Grey: whole sample
Red: spectra of the class

BPT diagrams

Class D

High activity of Star formation

Fraix-Burnet et al. (2021)



Fisher-EM on SDSS spectra

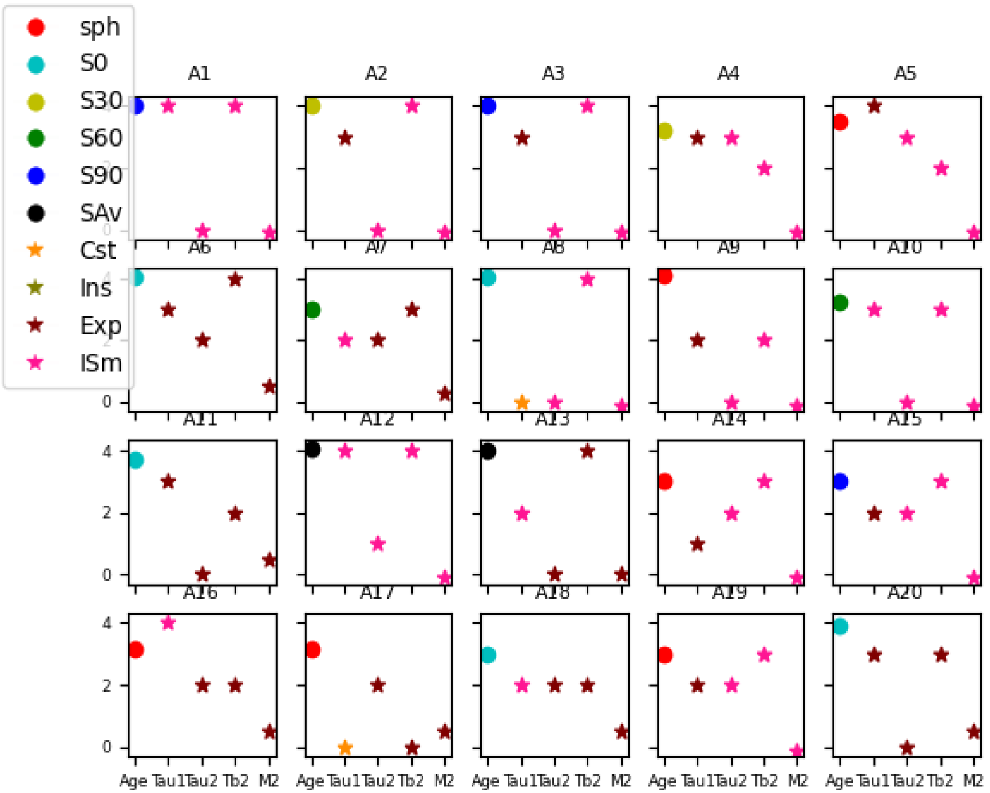
Next step → **Label** the mean spectra of the 86 classes

Mean spectra fitted by modeled stellar population models with PEGASE3 (***Fioc & Rocca-Volmerange 2019***)

Parameters :

- 2 stell. pop. (ages, e-folding times, burst mass)
- 4 SFR shapes
- 6 morphological types

→ **Each class has a distinct SP scenario**



Labeling example of the first A classes

Moultaka et al. (in prep)

Fisher-EM on GAMA survey

Two samples of galaxies from GAMA survey (***Driver et al. 2009***) :

- Sample of **5 physical parameters** of **7338 galaxies** from ***Turner et al. 2019 (T19)*** :

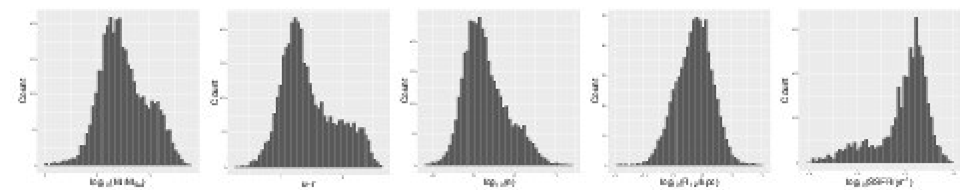
    -Stellar mass (**M***)

    -Color (**u-r**)

    -Sérsic index (**n**)

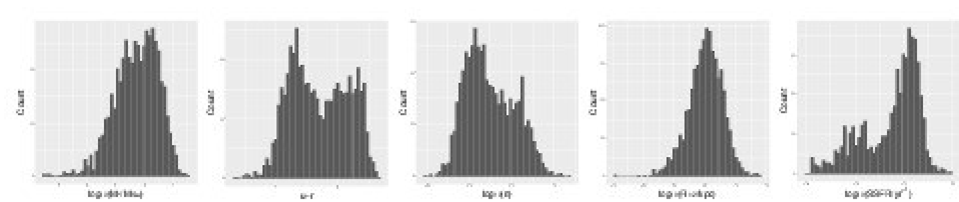
    -Effective radius (**$R_{1/2}$**)

    -Specific star formation rate (**sSFR**)



M* **u-r** **n** **$R_{1/2}$** **sSFR**

- Subsample of **spectra** of **2550 galaxies** taken from the previous sample for which **SDSS optical spectra** are available.



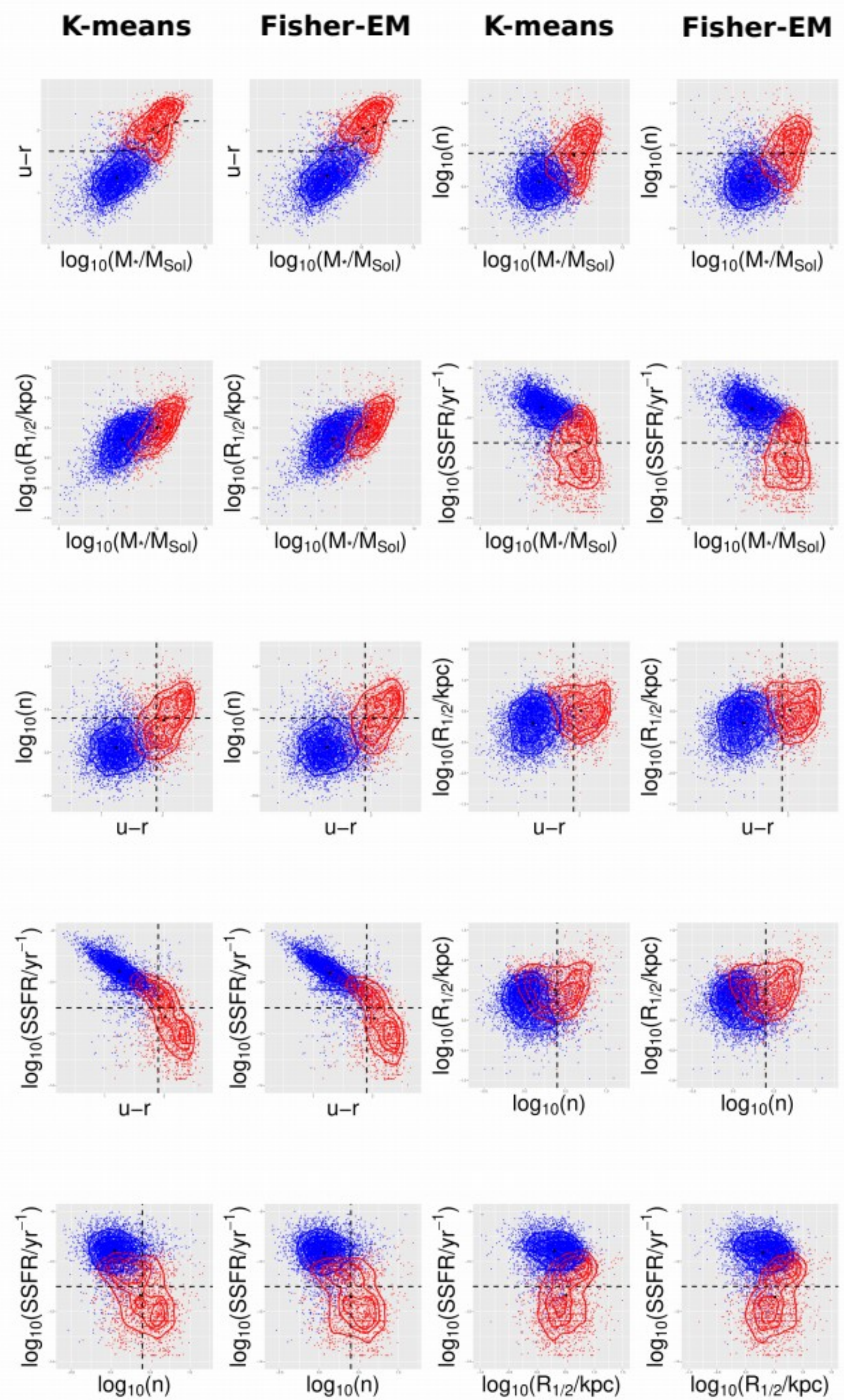
Distribution of the samples in the 5-physical parameter space

Beissière & Moultaka (in prep)

Fisher-EM

Comparison of the classifications obtained with K-means and Fisher-EM of the sample of physical parameters for K = 2

Equivalent solutions

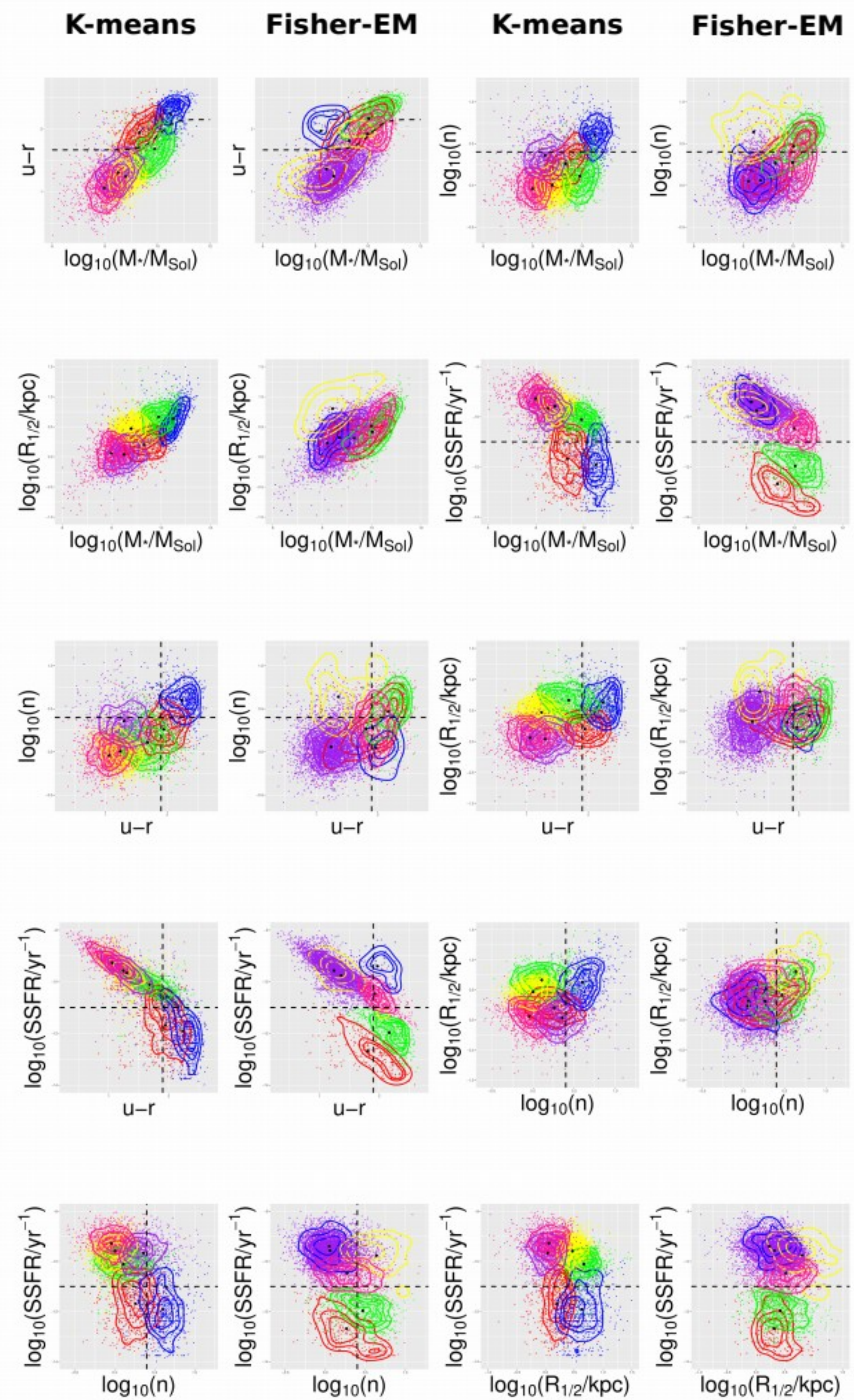


Beissière & Moultaka (in prep)

Fisher-EM

Comparison of the classifications obtained with K-means and Fisher-EM of the sample of physical parameters for K=6

Fisher-EM solutions are more physical

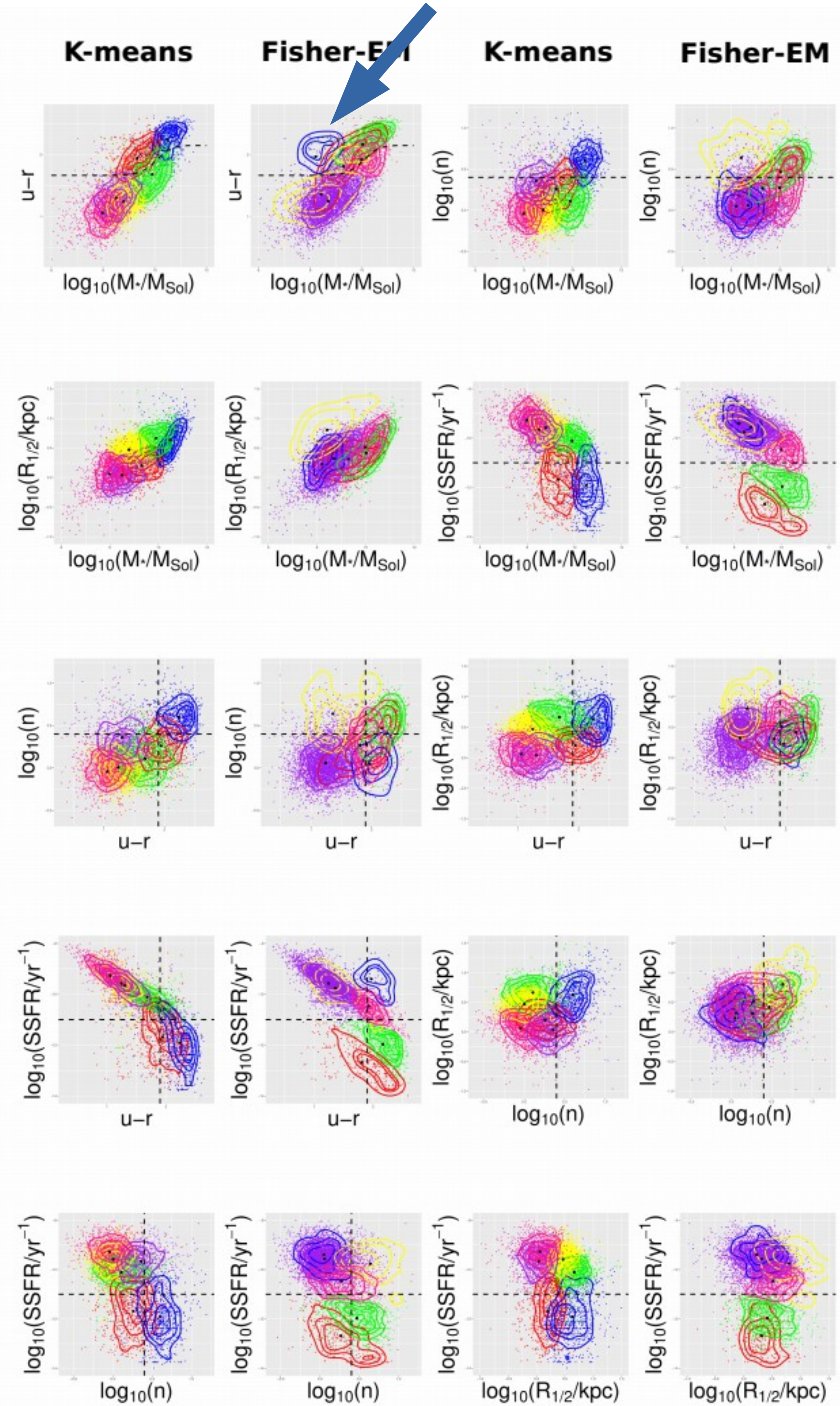


Beissière & Moultaka (in prep)

Fisher-EM

Comparison of the classifications obtained with K-means and Fisher-EM of the sample of physical parameters for K=6

Fisher-EM solutions are more physical

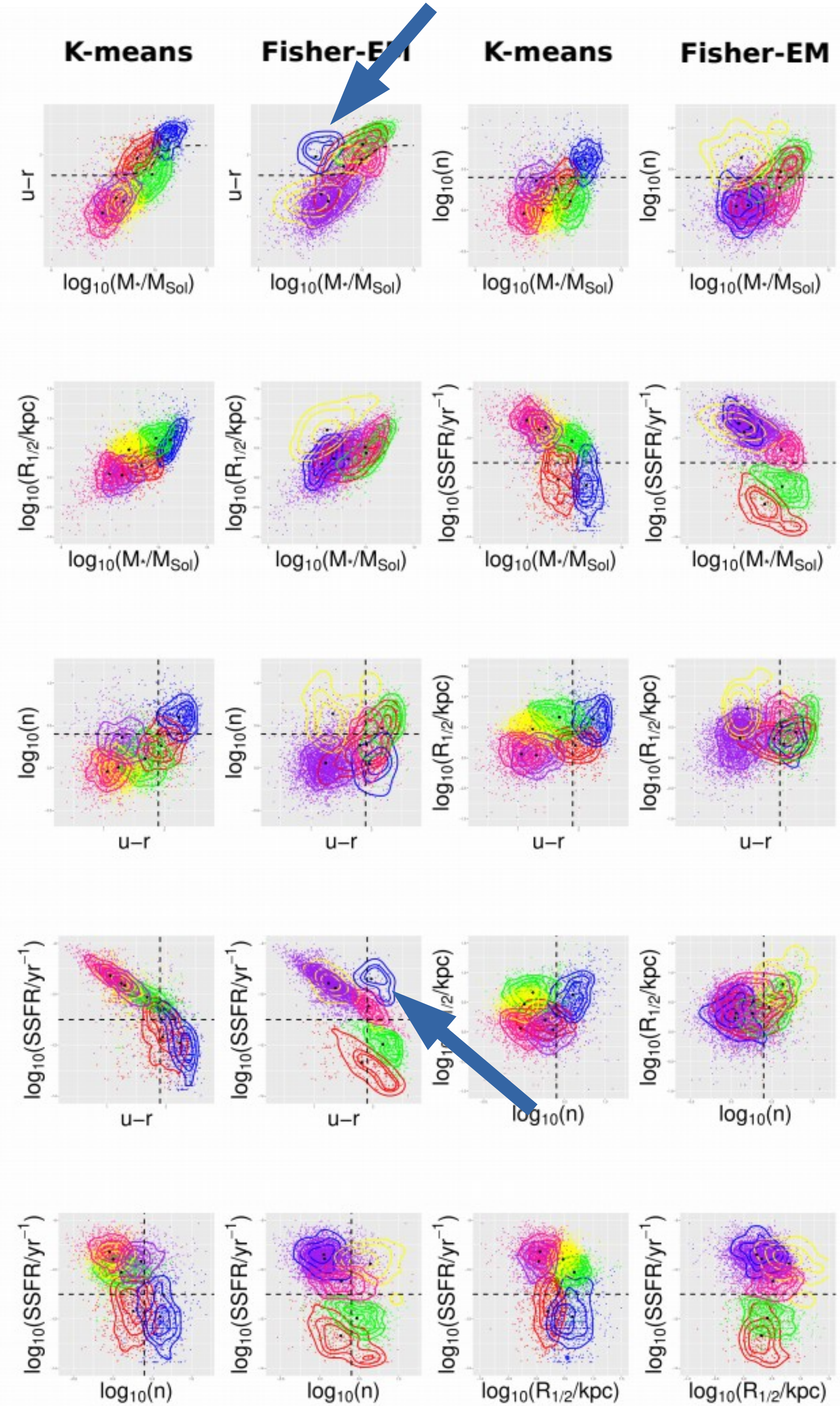


Beissière & Moultaka (in prep)

Fisher-EM

Comparison of the classifications obtained with K-means and Fisher-EM of the sample of physical parameters for K=6

Fisher-EM solutions are more physical

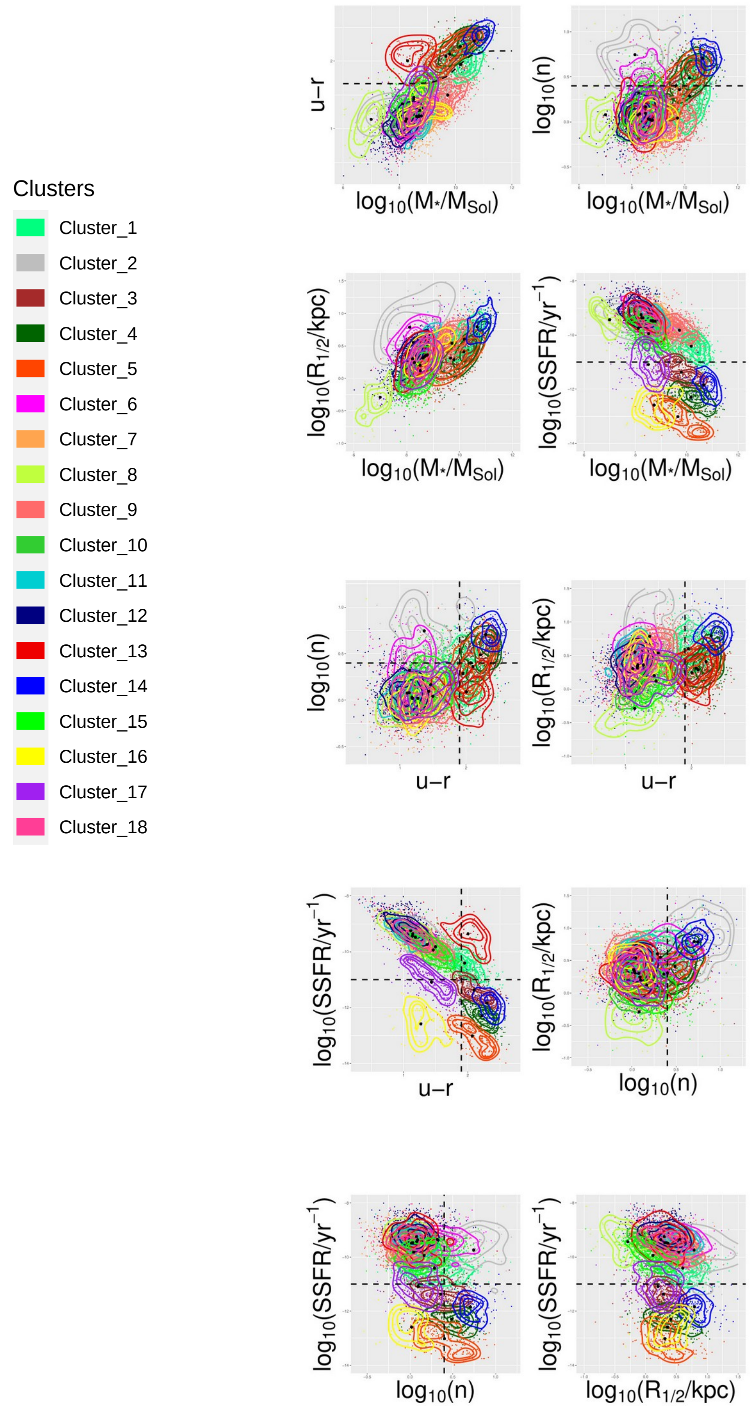


Beissière & Moultaka (in prep)

Fisher-EM

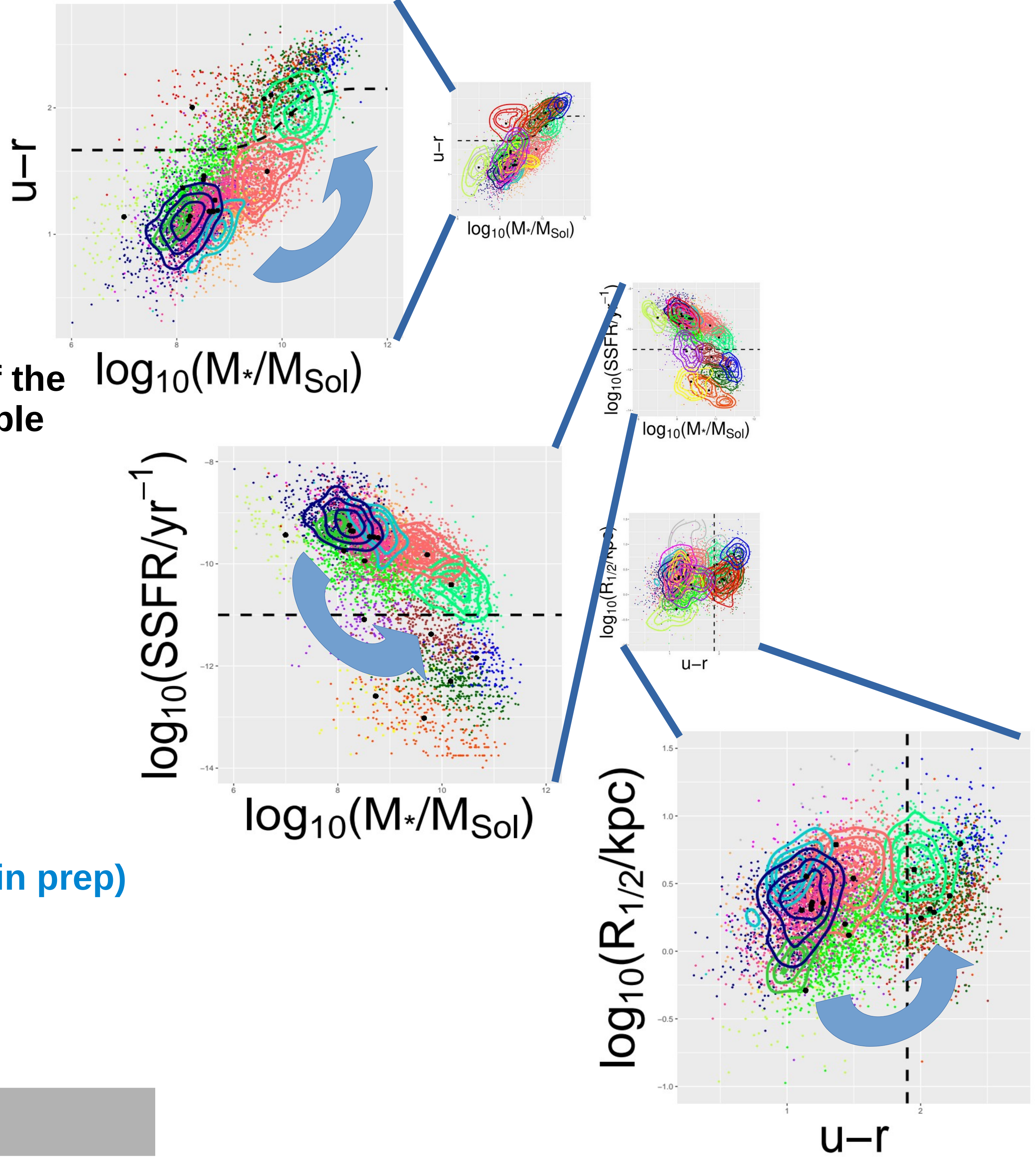
Optimal solution of the 5-parameter sample

K= 18



Beissière & Moultaka (in prep)

Fisher



Optimal solution of the 5-parameter sample

K= 18

Pathways from blue to red galaxies

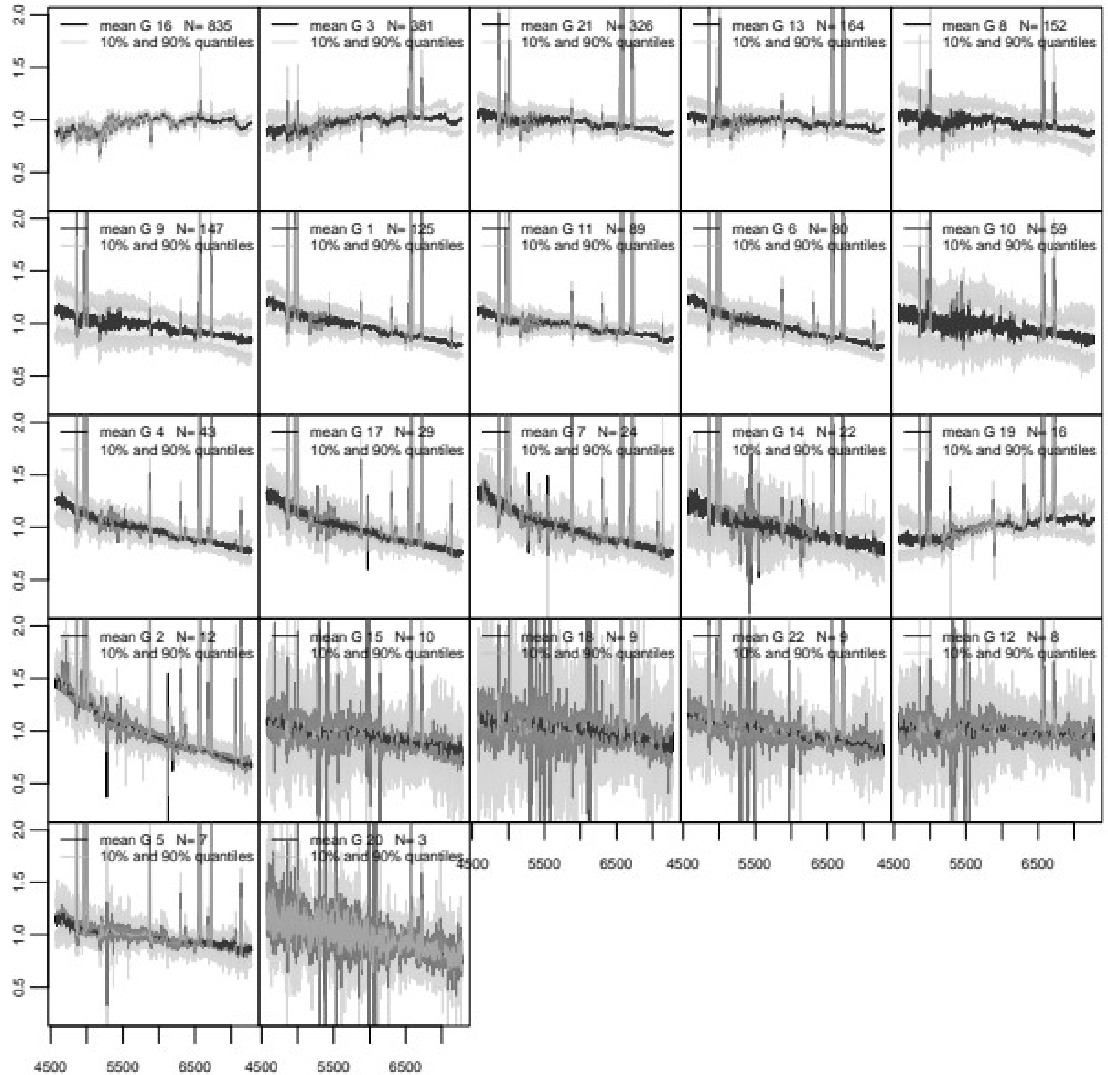
Beissière & Moultaka (in prep)

Fisher-EM on GAMA survey

Optimal solution of the sample of spectra

↓

K = 22

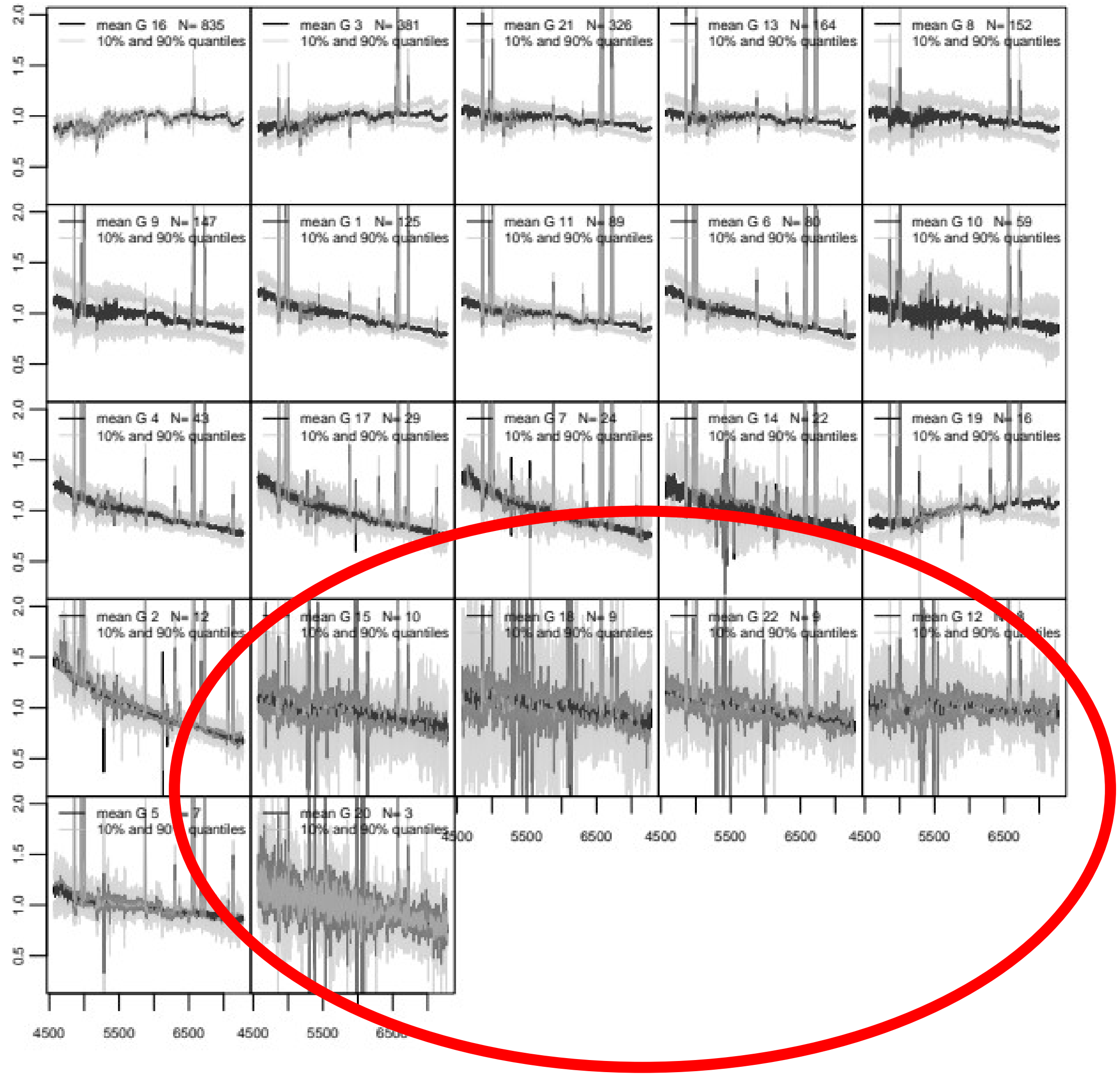


Beissière & Moultaka (in prep)

Fisher-EM on GAMA survey

Optimal solution of the sample of spectra

K = 22



Beissière & Moultaka (in prep)

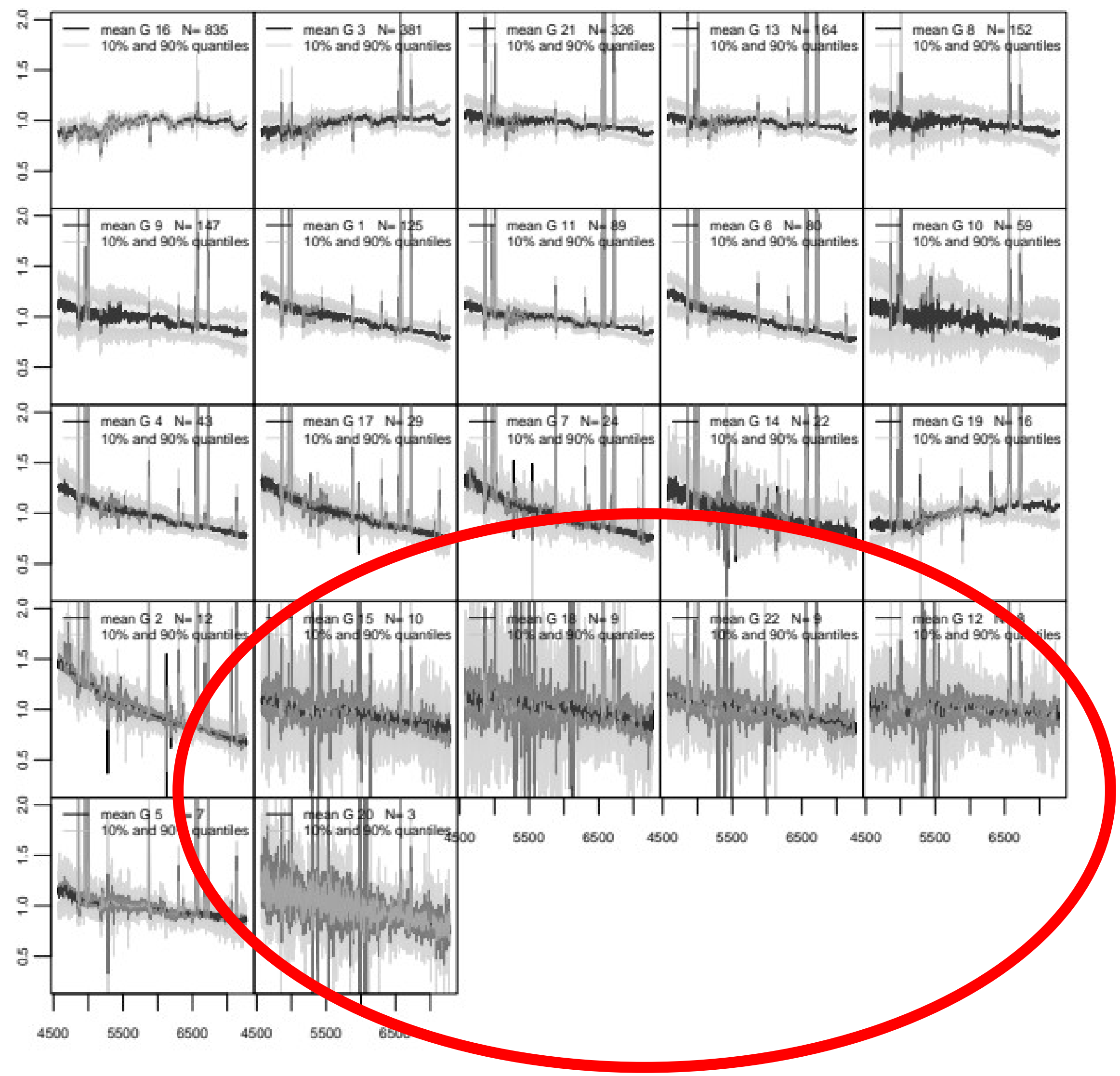
Fisher-EM on GAMA survey

Optimal solution of the sample of spectra

K = 22

K = 17

Beissière & Moultaka (in prep)



Fisher-EM on GAMA survey

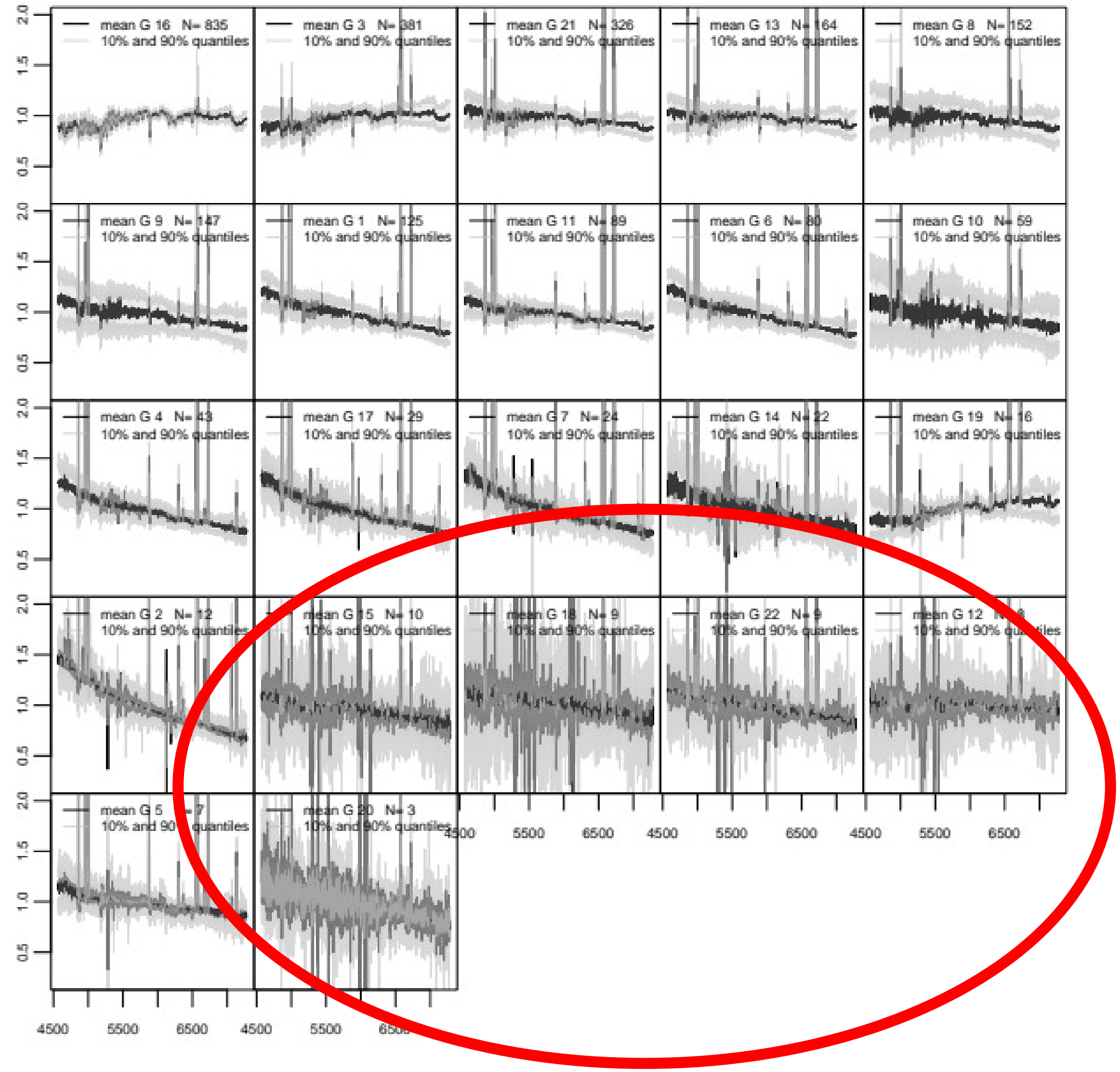
Optimal solution of the sample of spectra

K = 22

K = 17

Beissière & Moultaka (in prep)

To be continued ...

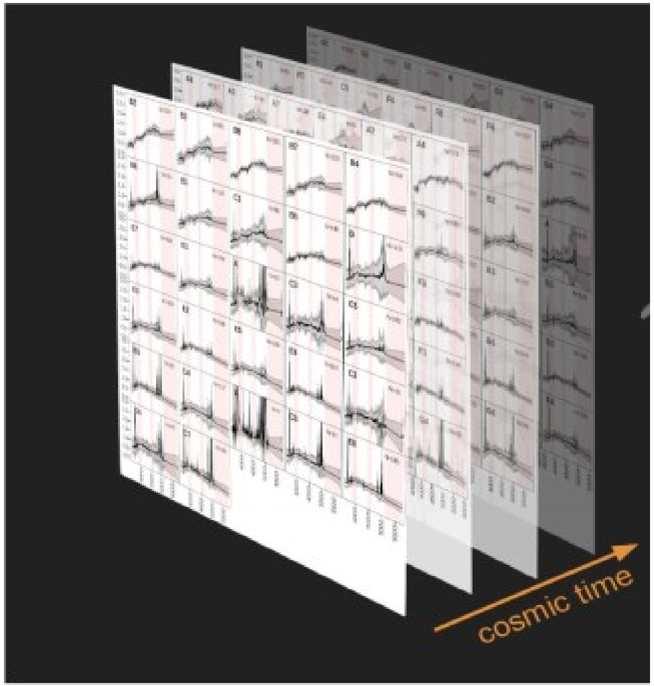


Fisher-EM on VIPERS spectra

79 000 spectra of galaxies from VIPERS survey ($0.4 < z < 1.2$)

(corrected for redshift and masked, homogeneously sampled)

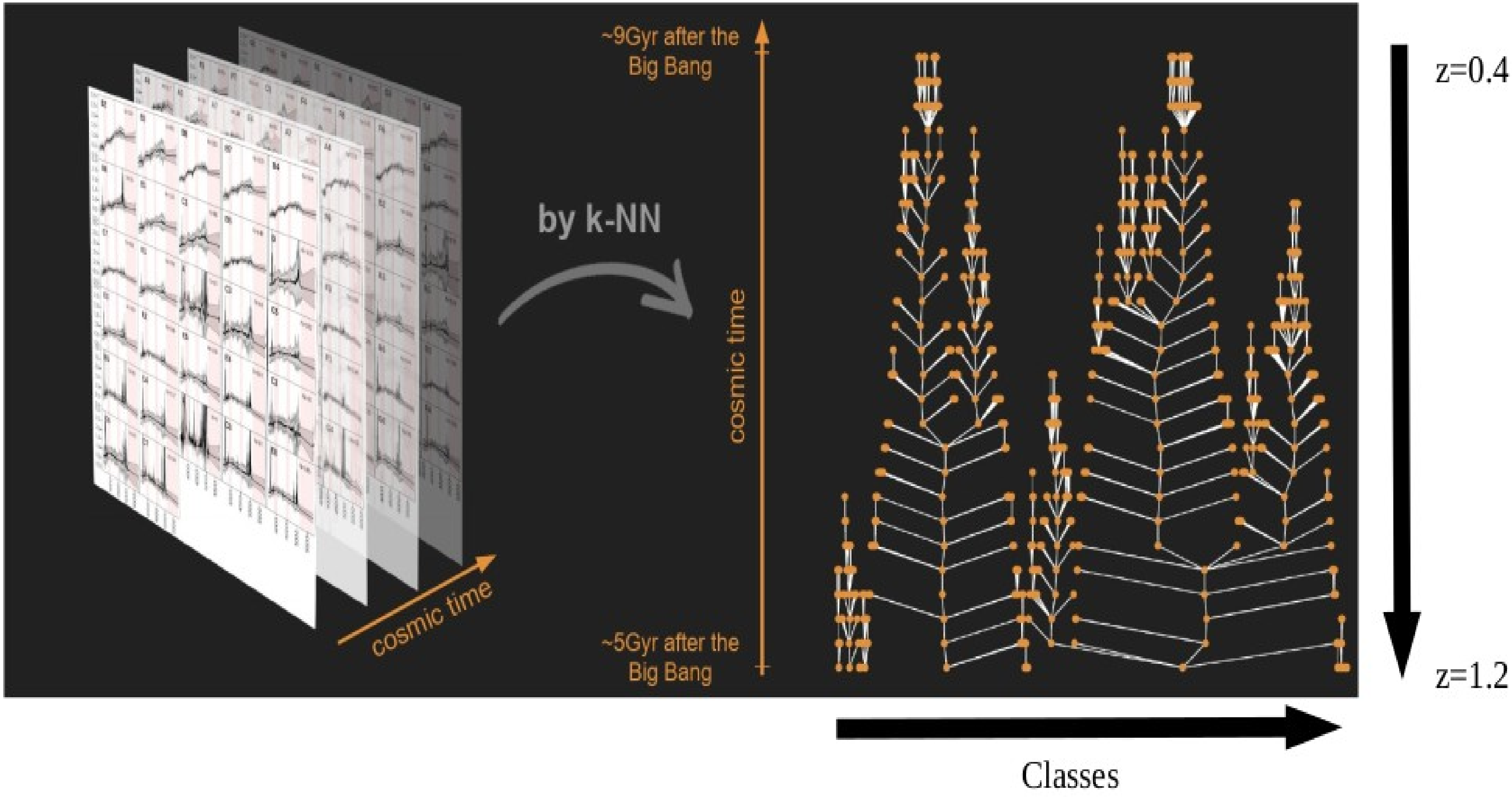
- Divided into 26 bins of redshifts → constant step of cosmic time
- Classification at each redshift bin → 26 classifications



Dubois et al. (2024)

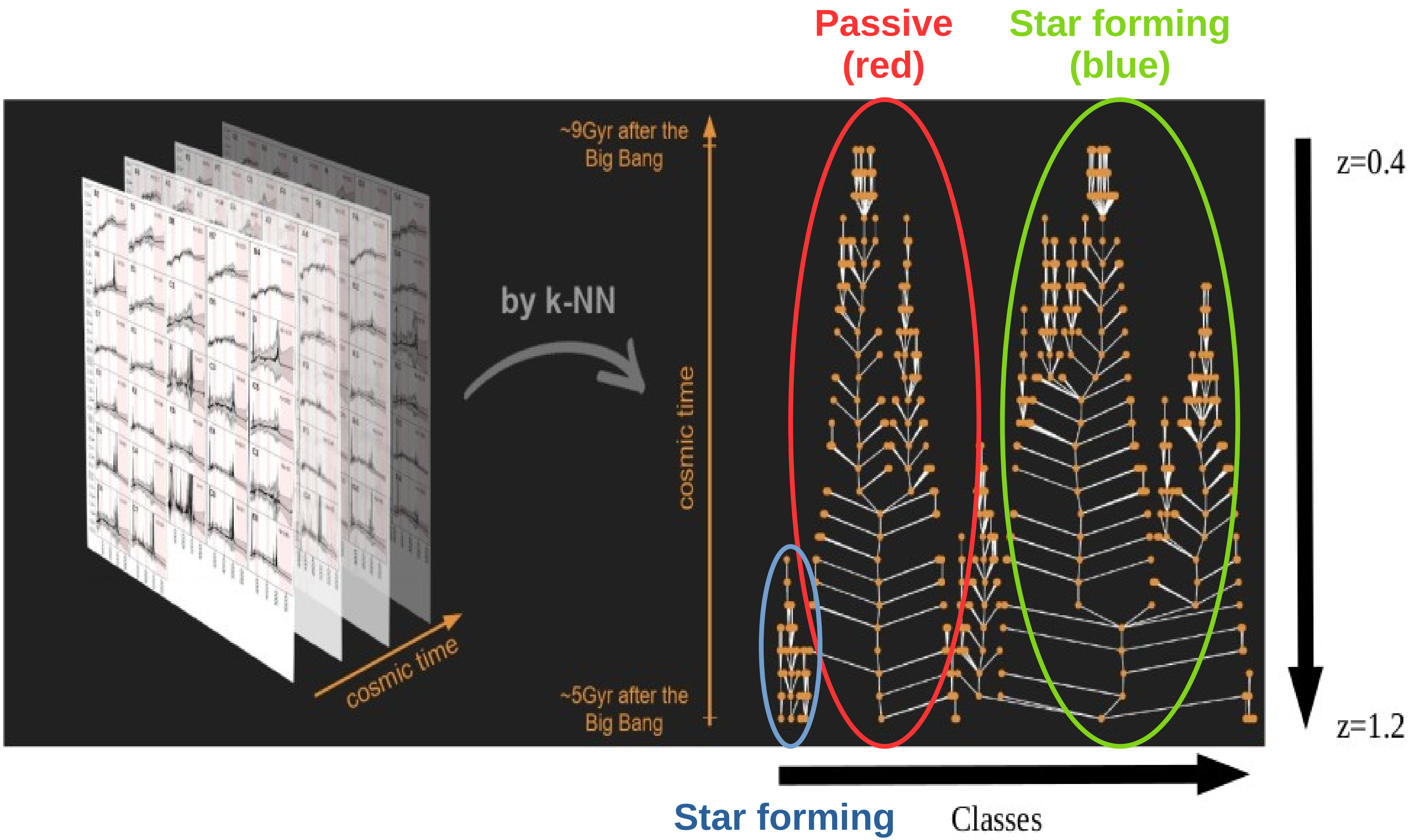
Link classes of successive epochs using k-NN algorithm → **Evolutionary tree**

Fisher-EM on VIPERS spectra



Dubois et al. (2024)

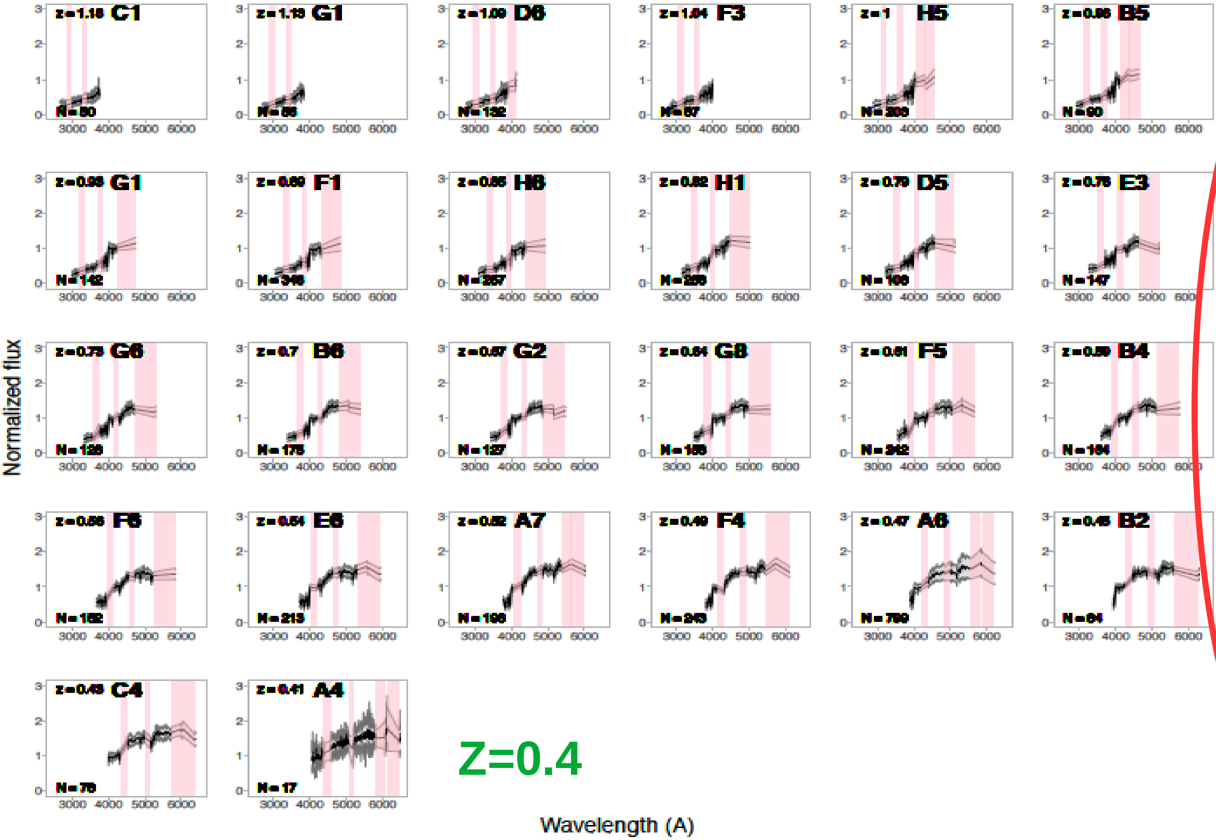
Fisher-EM on VIPERS spectra



Dubois et al. (2024)

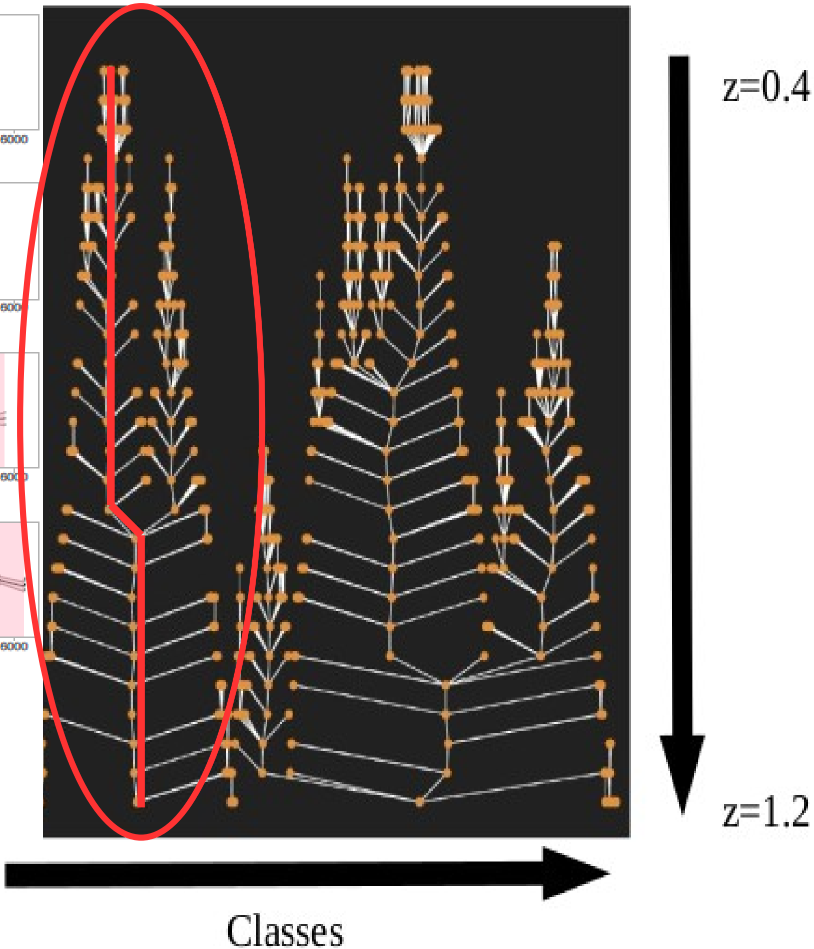
Fisher-EM on VIPERS spectra

Z=1.2



Z=0.4

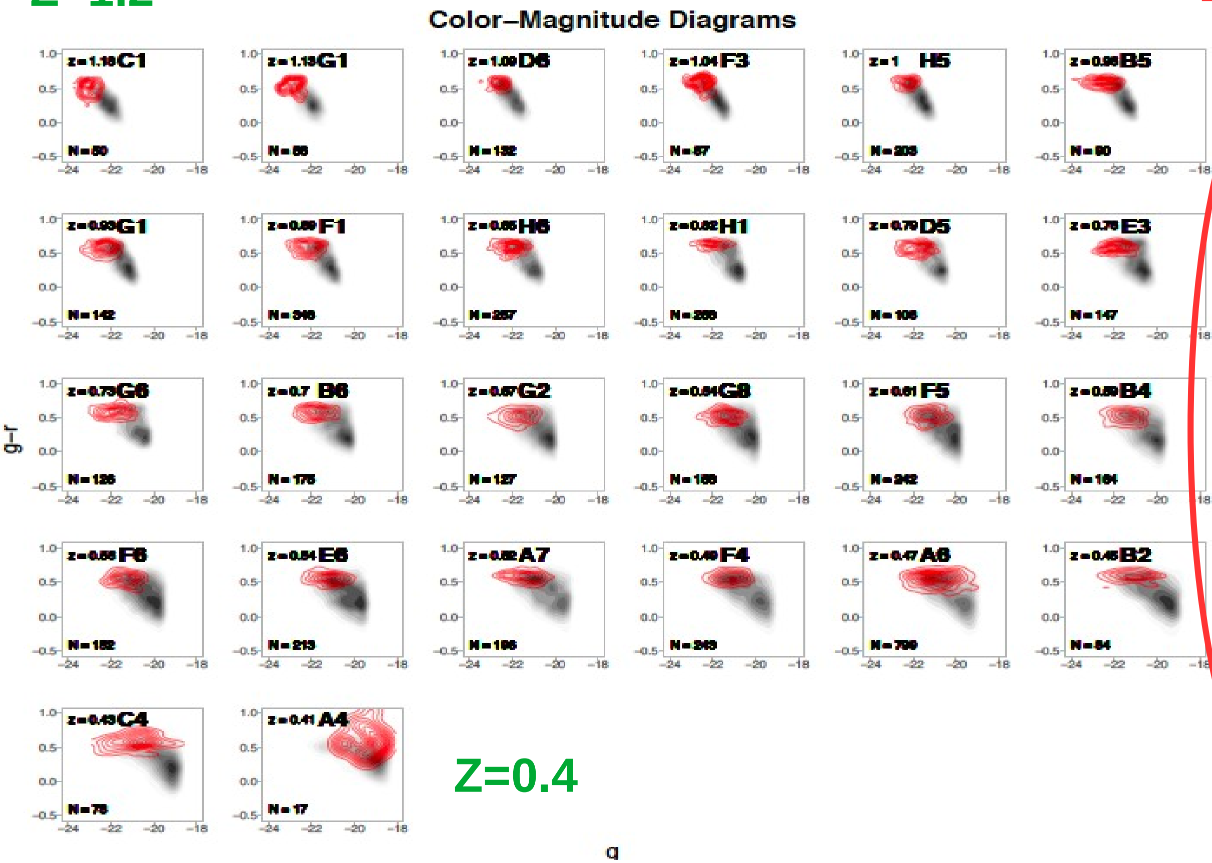
Passive Branch



Dubois et al. (2024)

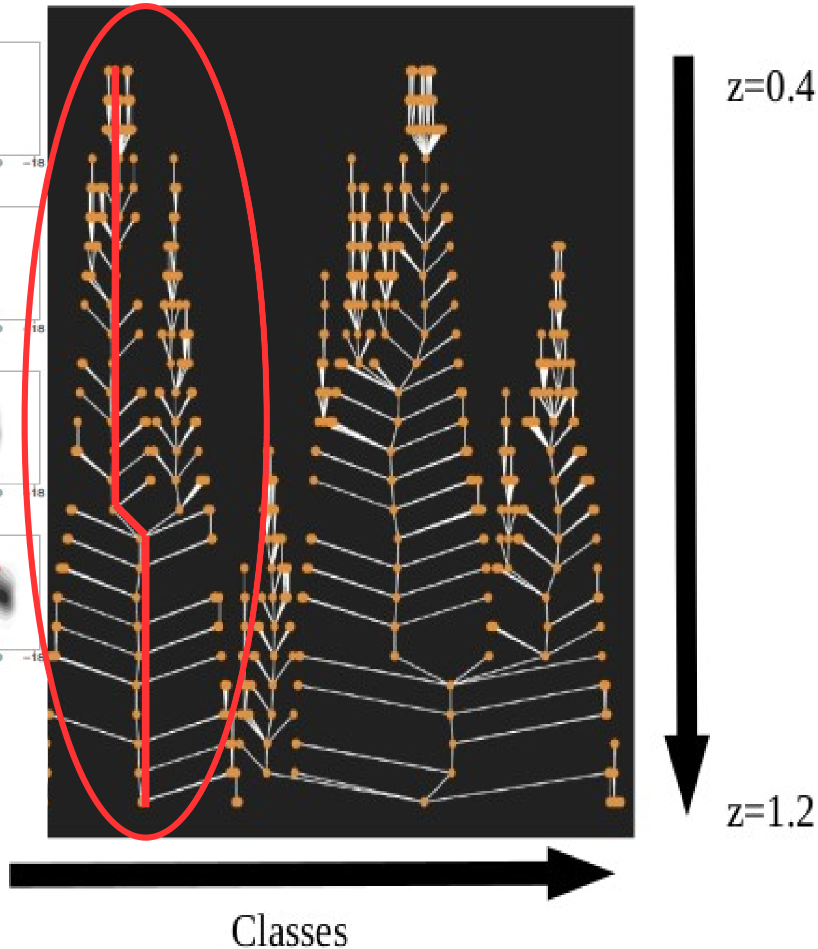
Fisher-EM on VIPERS spectra

Z=1.2



Z=0.4

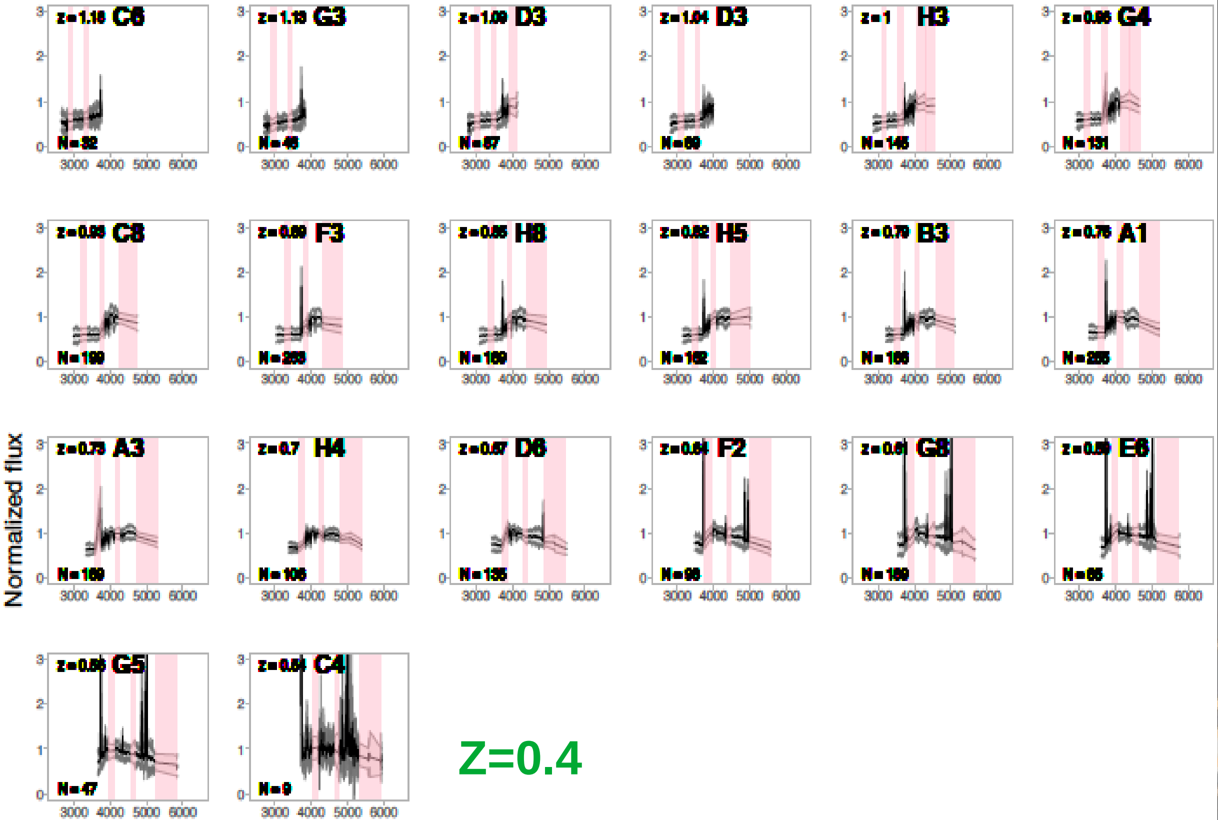
Passive Branch



Dubois et al. (2024)

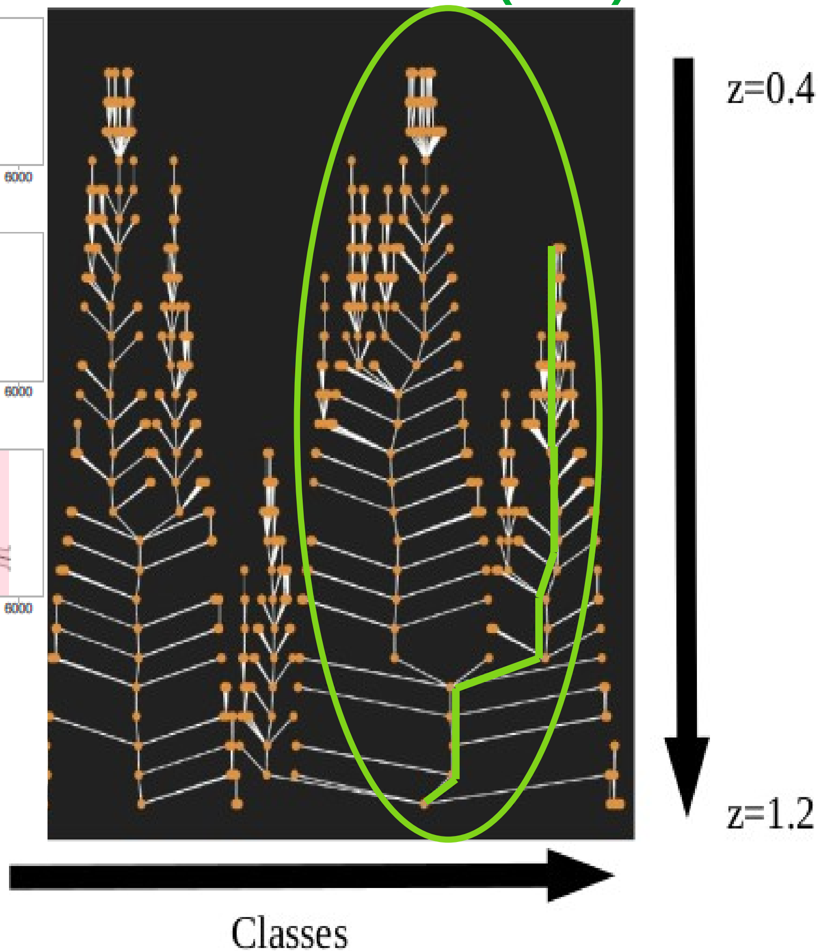
Fisher-EM on VIPERS spectra

Z=1.2



Z=0.4

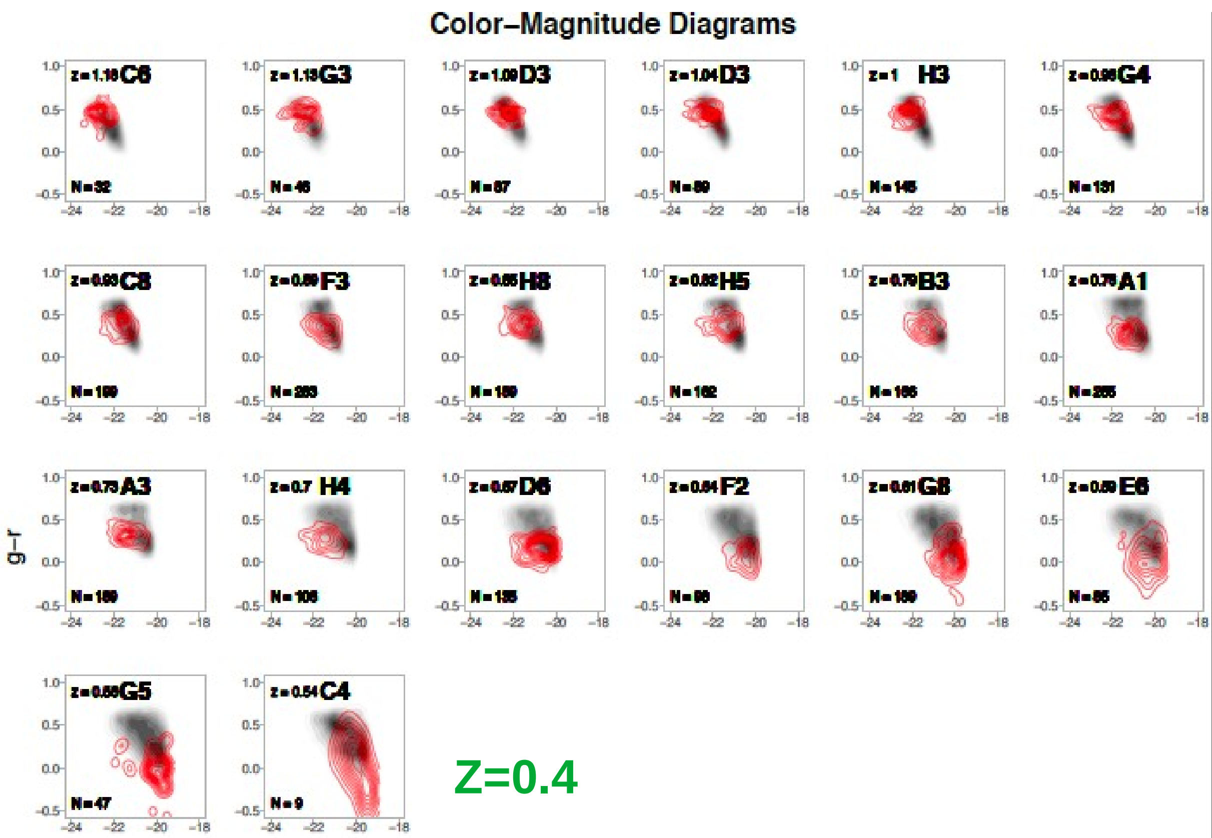
Star forming (blue)



Dubois et al. (2024)

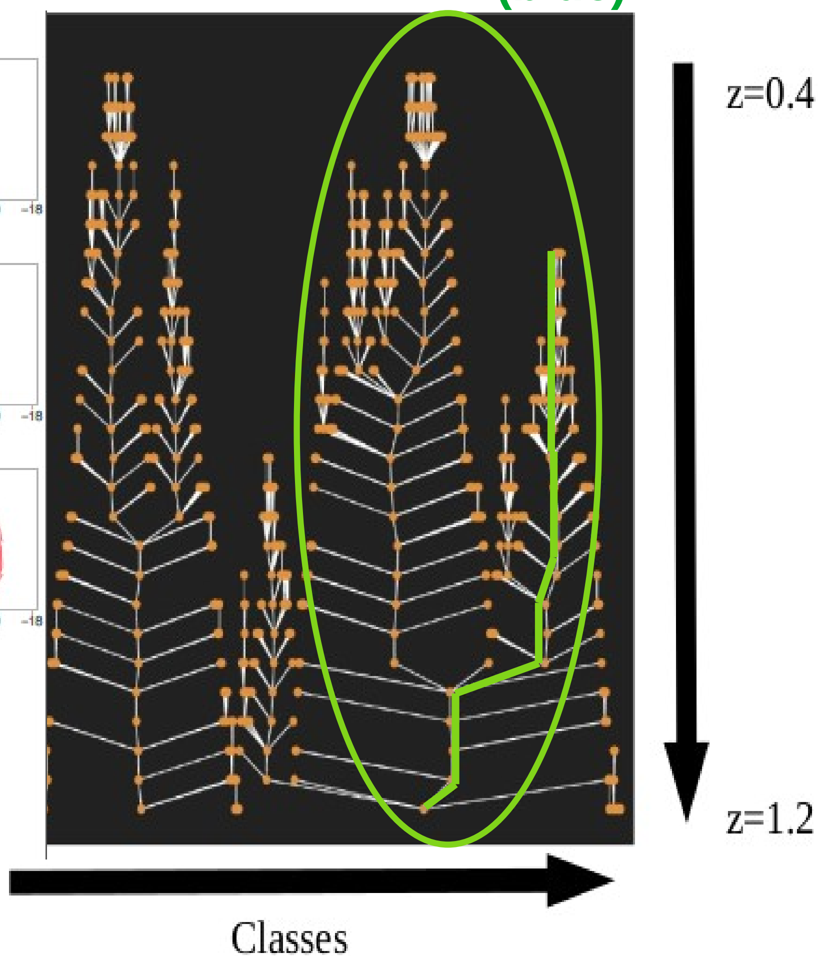
Fisher-EM on VIPERS spectra

Z=1.2



Z=0.4

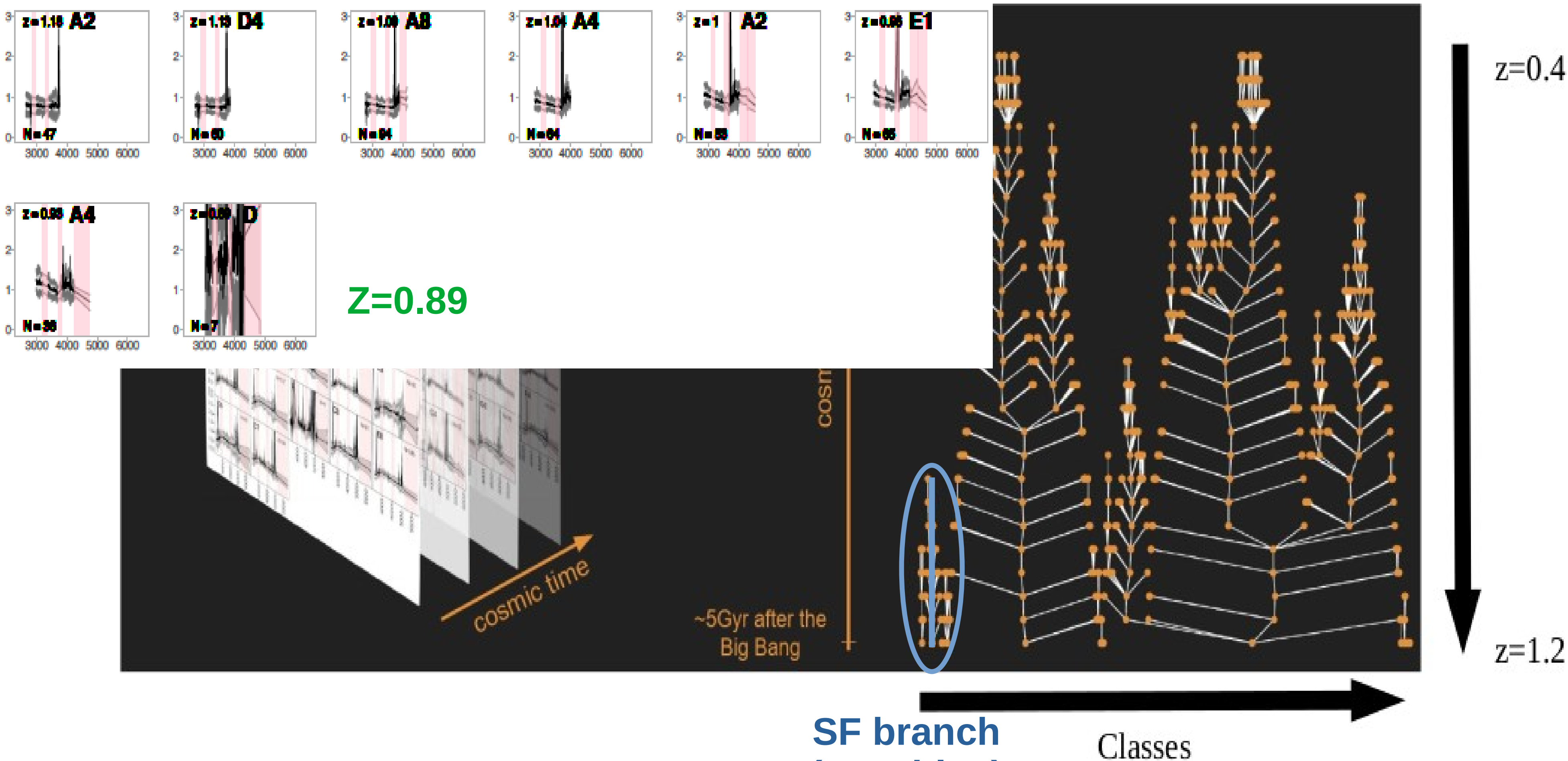
Star forming
(blue)



Dubois et al. (2024)

Fisher-EM on VIPERS spectra

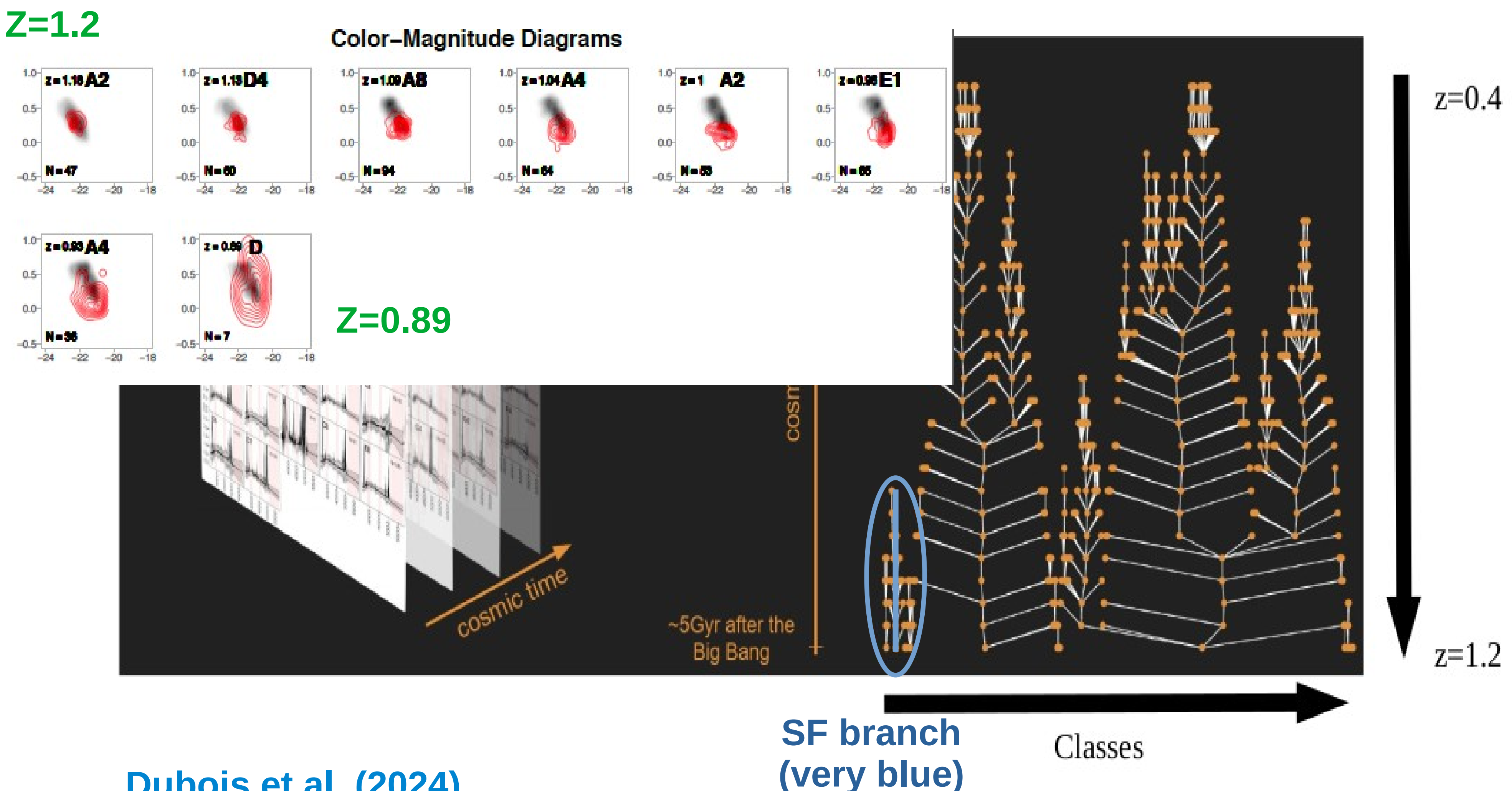
Z=1.2



SF branch
(very blue)

Dubois et al. (2024)

Fisher-EM on VIPERS spectra



Dubois et al. (2024)

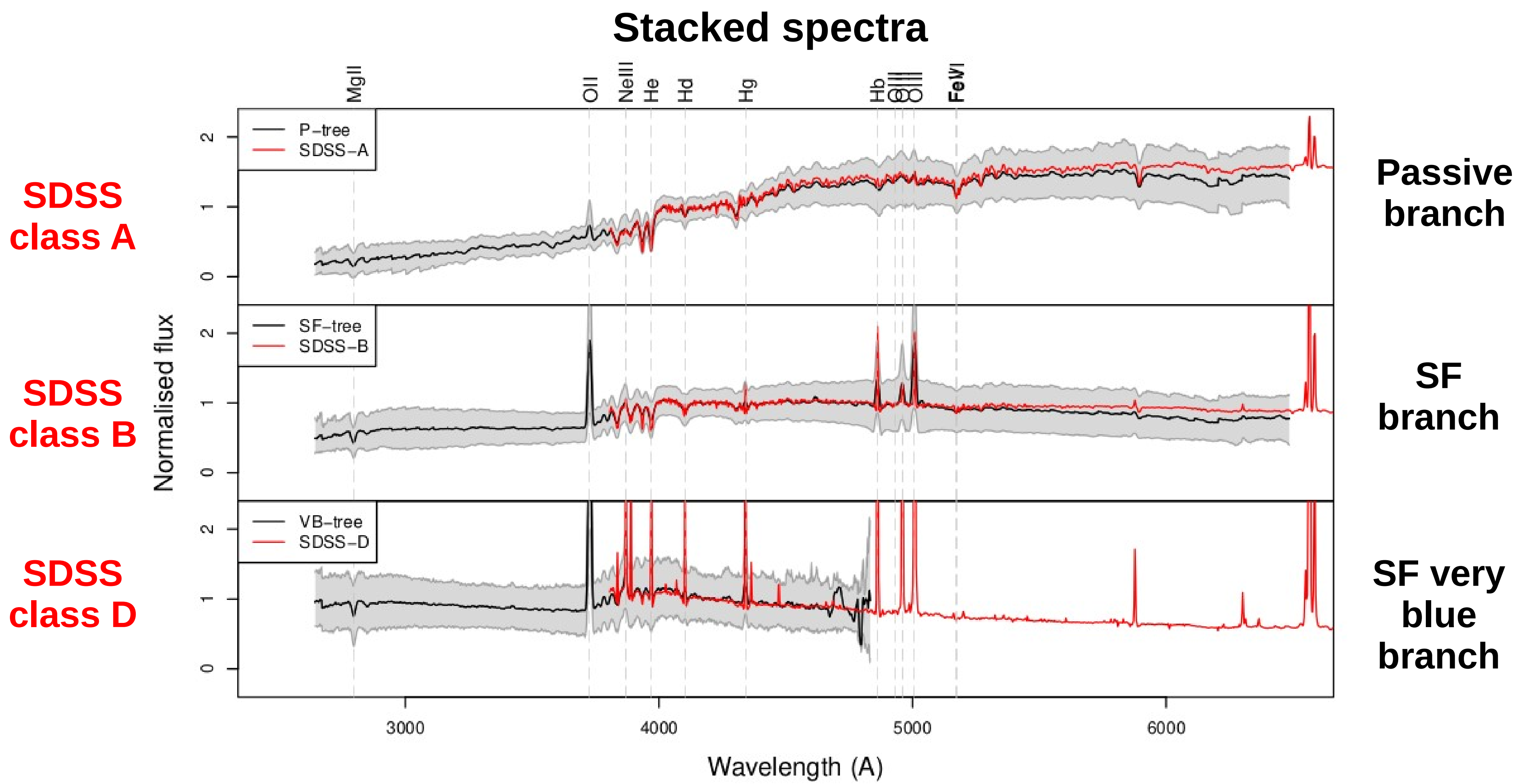
Conclusion

- Demonstrate the **power of unsupervised approach**
- **Fisher-EM** is a reliable method for unsupervised classification at **high dimension**
- **Discriminates** physical parameters of galaxies from their spectra
- Provides a meaningful **physical classification of optical spectra** at low redshift galaxies
- **Distinguishes pathways** in the parameter spaces of the classical **bimodal distribution** of galaxies
- Seems to discriminate **evolutionary sequences** of galaxy spectra through the cosmic ages

What’s next ?

Probe other redshifts, wavelength ranges and physical observables

Fisher-EM on VIPERS spectra



Dubois et al. (2024)