Galaxy Evolution from Machine Learning

Tsutomu T. TAKEUCHI

Division of Particle and Astrophysical Science, Nagoya University, Japan

6th Galaxy Evolution Workshop, Kashiwa, 5-7 June, 2019

1. Introduction: What is the Galaxy Evolution?

In those days

Galaxies evolve.

Beatrice Tinsley

1. Introduction: What is the Galaxy Evolution?

In those days

Beatrice Tinsley

Galaxies evolve.

In galactic astronomy of 70-80's, the meaning of galaxy evolution seemed to be unambiguous.

Galaxy evolution

= evolution of stellar population and chemical composition

Today

Galaxies evolve in various aspects:

Today

Galaxies evolve in various aspects:

$$
\begin{aligned}
\mathrm{SFR}(t) &= f_1(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_*(t) &= f_2(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{mol}}(t) &= f_3(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{HI}}(t) &= f_4(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{dust}}(t) &= f_5(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{halo}}(t) &= f_6(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
\delta_{\mathrm{gal}}(t) &= f_7(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
&\vdots
\end{aligned}
$$

$$\boxed{x = x(T|T > t)}$$

Today

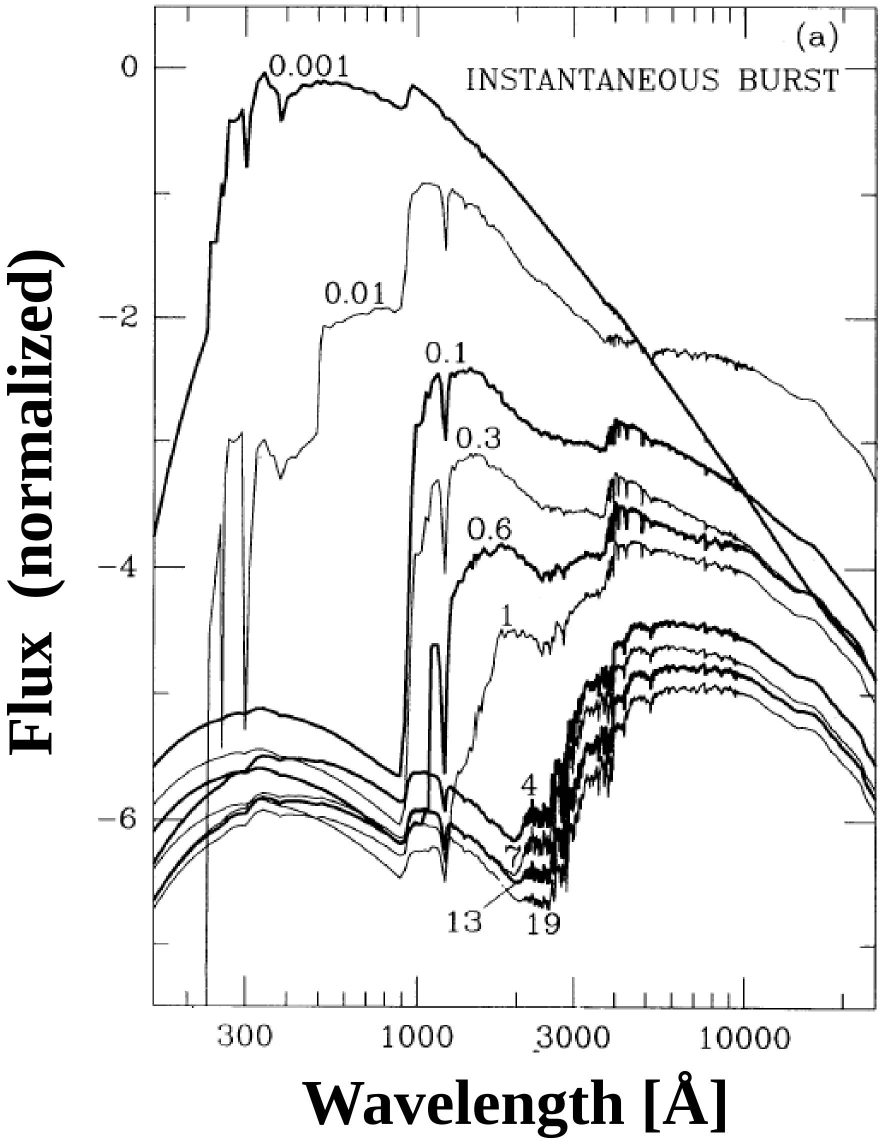
Galaxies evolve in various aspects:

$$
\begin{aligned}
\mathrm{SFR}(t) &= f_1(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_*(t) &= f_2(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{mol}}(t) &= f_3(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{HI}}(t) &= f_4(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{dust}}(t) &= f_5(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{halo}}(t) &= f_6(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
\delta_{\mathrm{gal}}(t) &= f_7(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
&\vdots
\end{aligned}
$$

This is the formal and ultimate goal of the studies on galaxy evolution, but clearly it is a substantially complicated problem. It is time to define the evolution of galaxies with more objective point of view.

2. Galaxy Evolution from Luminosities

2.1 Galaxy evolution in multiband luminosity space



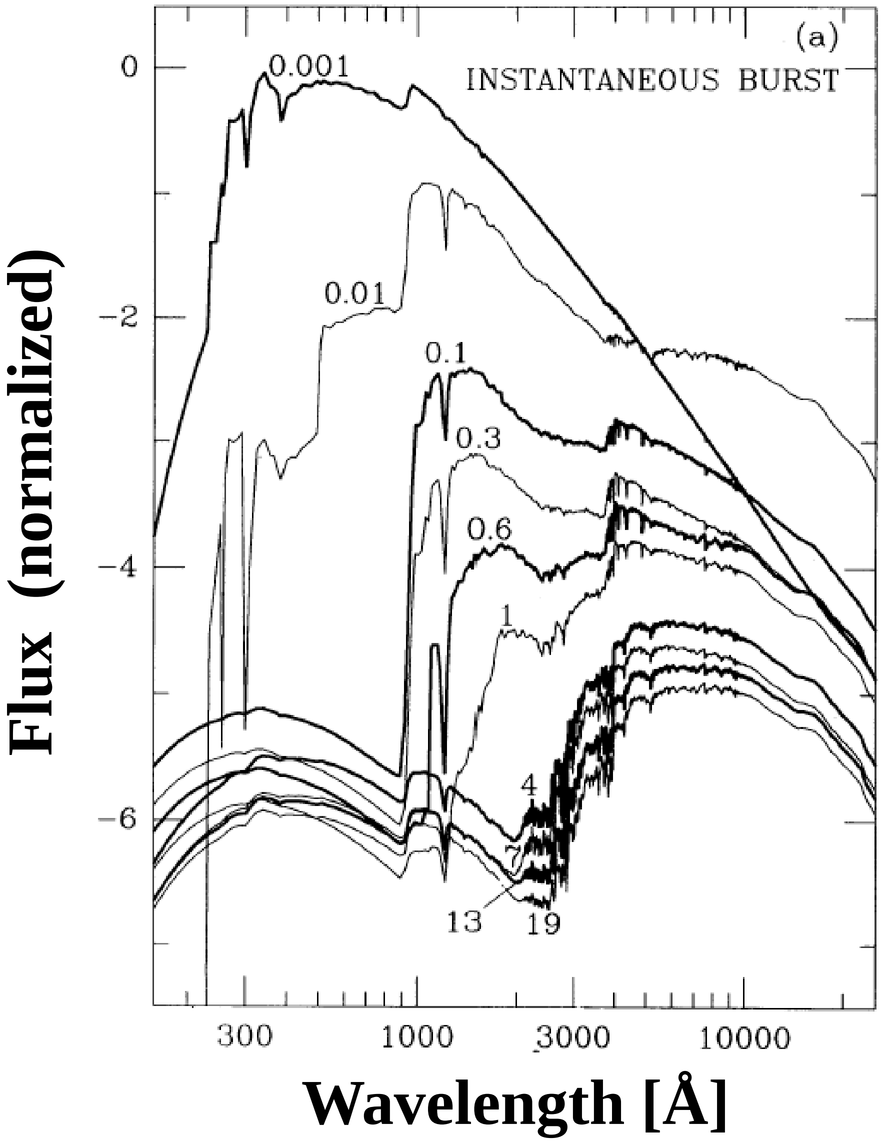
Star formation history (SFH) is one of the key factors of galaxy evolution.

SFH is directly reflected to the spectral luminosity of galaxies.

Bruzual & Charlot (1993)

2. Galaxy Evolution from Luminosities

2.1 Galaxy evolution in multiband luminosity space



Star formation history (SFH) is one of the key factors of galaxy evolution.

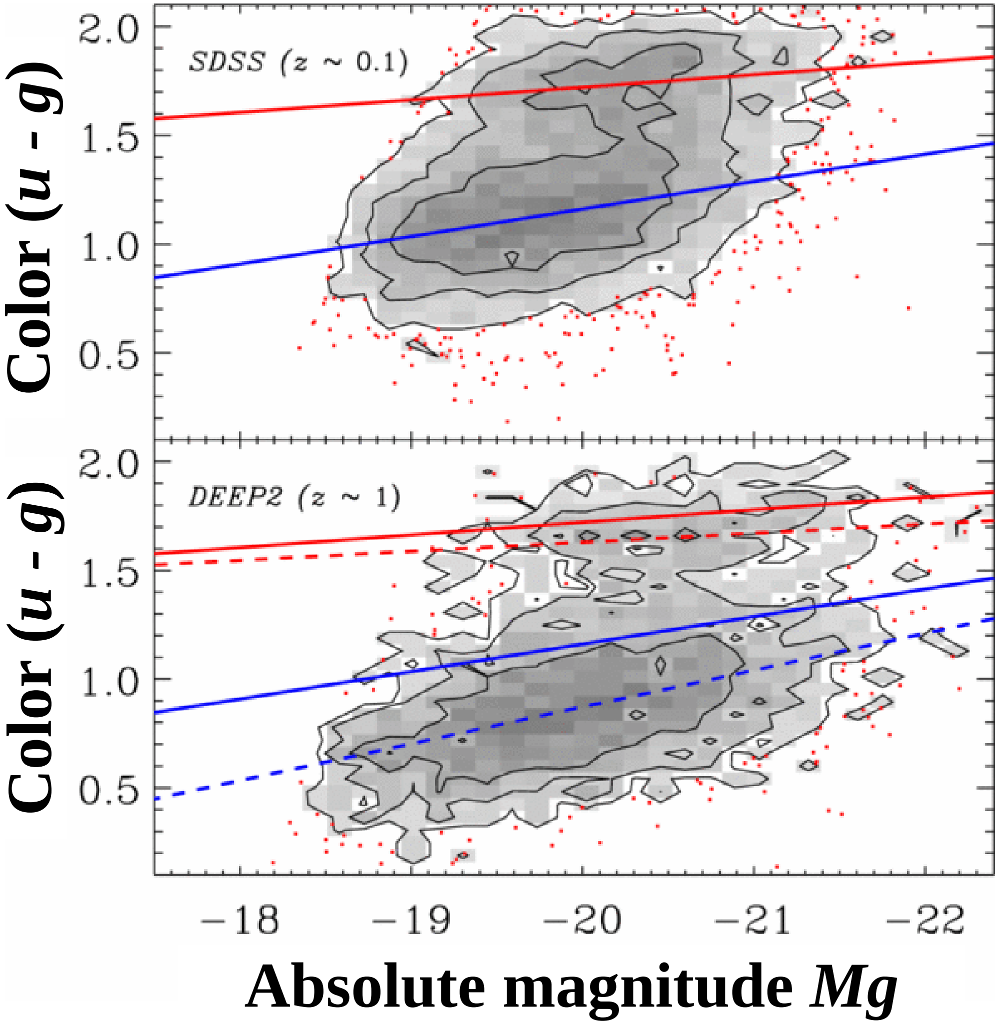
SFH is directly reflected to the spectral luminosity of galaxies.

Galaxy evolution related to the SFH will be well represented in the multiwavelength (band) luminosity space.

Bruzual & Charlot (1993)

2.2 Traditional methods and its limitation

Color-magnitude relation



Blanton (2006)

If we plot galaxy luminosity (absolute magnitude) vs. color, a clear dichotomy is found: the color bimodality.

Redder galaxies:

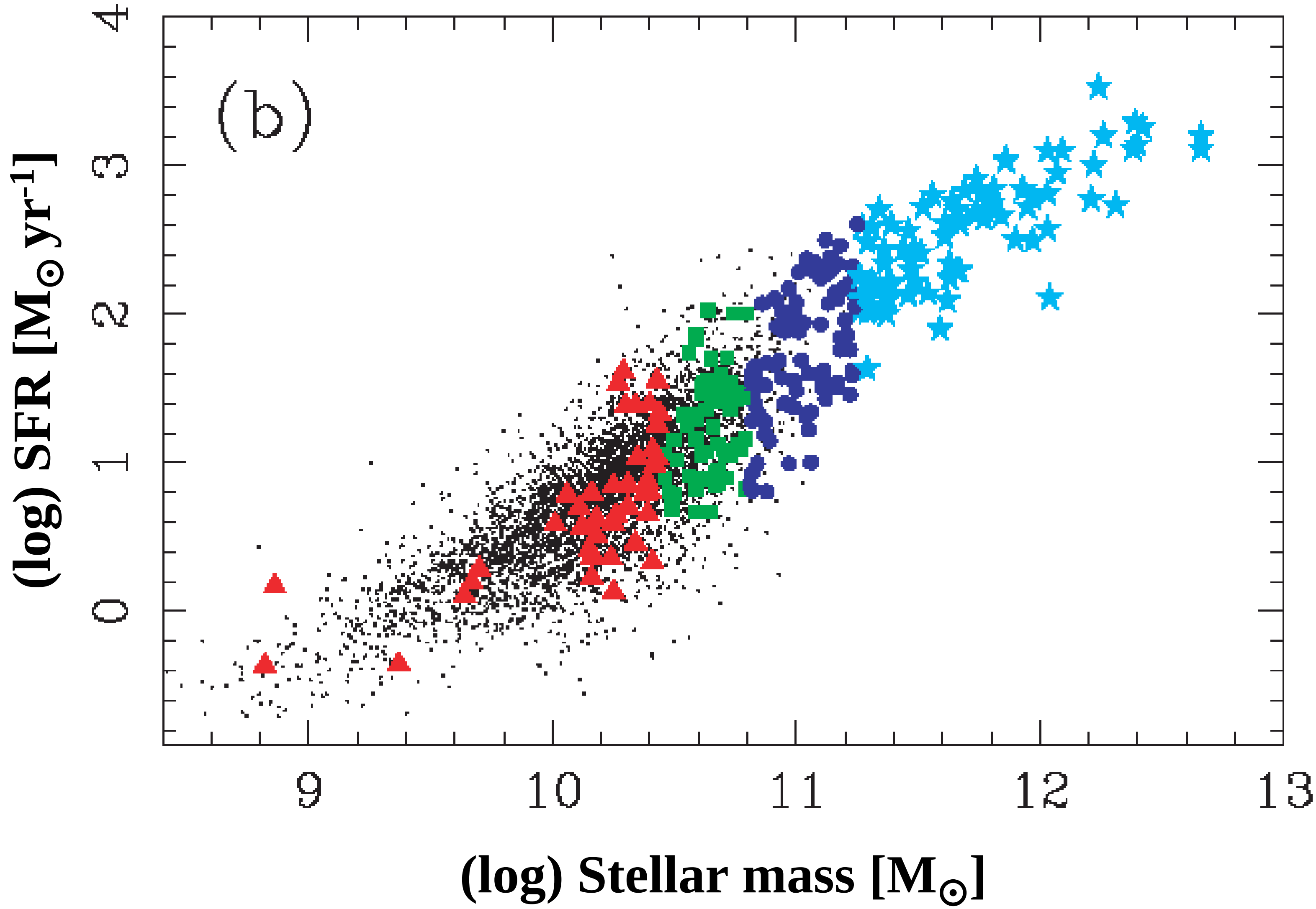
red sequence

Bluer galaxies:

blue cloud

Boundary: green valley

Star formation main sequence of galaxies (Local)



Seymour et al. (2008)

Potential problem in the traditional color-based methods

Colors are basically ratios of two luminosities.

$\Rightarrow$ Selection effect is always too entangled and messy.

Potential problem in the traditional color-based methods

Colors are basically ratios of two luminosities.

⇒ Selection effect is always too entangled and messy.

⇒ Completeness test is almost impossible in a simple way.

Potential problem in the traditional color-based methods

Colors are basically ratios of two luminosities.

⇒ Selection effect is always too entangled and messy.

⇒ Completeness test is almost impossible in a simple way.

What can we do, then?

Potential problem in the traditional color-based methods

Colors are basically ratios of two luminosities.

$\Rightarrow$ Selection effect is always too entangled and messy.

$\Rightarrow$ Completeness test is almost impossible in a simple way.

$\Downarrow$

Suggestion: forget about colors!

Potential problem in the traditional color-based methods

Colors are basically ratios of two luminosities.

⇒ Selection effect is always too entangled and messy.

⇒ Completeness test is almost impossible in a simple way.

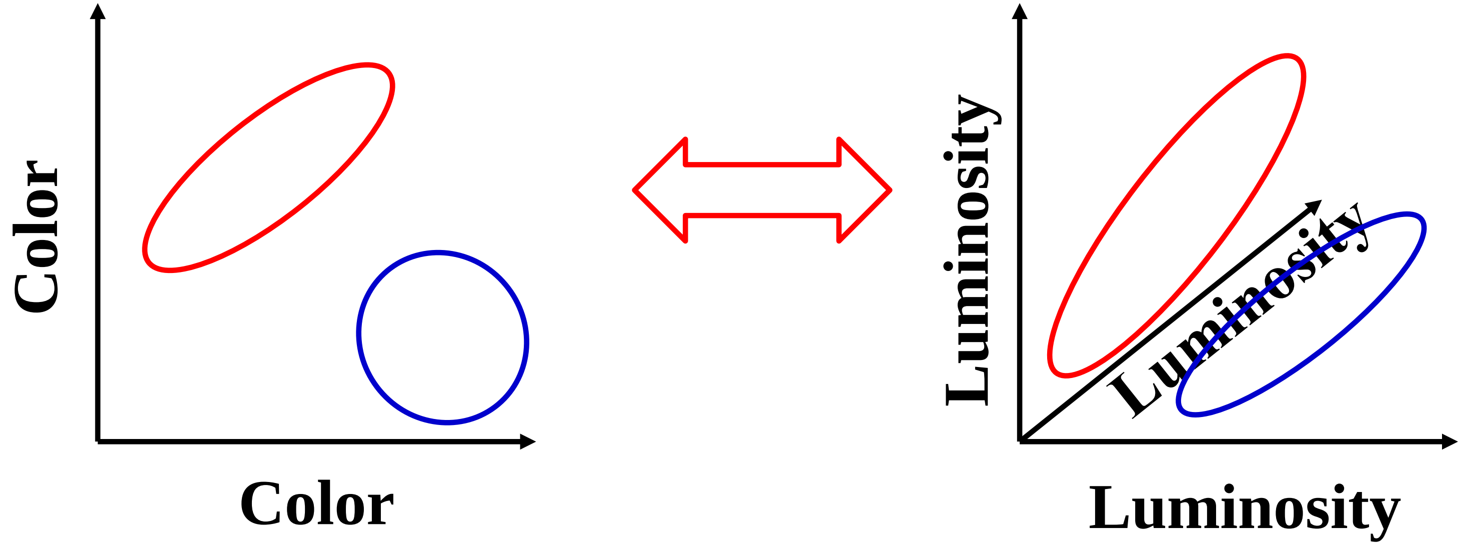
Suggestion: forget about colors!

Instead, we can simply use the distribution of galaxies in a multidimensional luminosity (absolute magnitude) space.

Potential problem in the traditional color-based methods

Instead, we can simply use the distribution of galaxies in a multidimensional luminosity (absolute magnitude) space.

Since we have a bimodality in color-color space, we must have an equivalent peaks in the multidimensional luminosity space. Color-color plots only show reduced information.



Potential problem in the traditional color-based methods

Instead, we can simply use the distribution of galaxies in a multidimensional luminosity (absolute magnitude) space.

Since we have a bimodality in color-color space, we must have an equivalent peaks in the multidimensional luminosity space. Color-color plots only show reduced information.

The boundary can be automatically defined by the machine-learning type method.

Potential problem in the traditional color-based methods

Instead, we can simply use the distribution of galaxies in a multidimensional luminosity (absolute magnitude) space.

Since we have a bimodality in color-color space, we must have an equivalent peaks in the multidimensional luminosity space. Color-color plots only show reduced information.

The boundary can be automatically defined by the machine-learning type method.

Advantage of this idea is that we should simply deal with the selection at each band, not as an entangled multiple selection.

3. Data: VIPERS

VIMOS Public Extragalactic Redshift Survey (VIPERS)

- Want a volume and density comparable to a survey like 2dFGRS, but at z = [0.5-1]: cosmology driven, but with broader legacy return.

- Means Volume ~ $5 \times 10^7\ h^{-3}$ Mpc^3, ~ 90,000 redshifts, close to full sampling.

- $I_{AB} < 22.5$

http://vipers.inaf.it/

Final public release of complete VIPERS galaxy catalogue of ~90,000 redshifts (PDR-2)

VIPERS products

- **Redshifts**
- **Spectra**
- **Photometry from CFHTLS *ugriz* (optical), GALEX FUV and NUV (ultraviolet), and additional *ZYJHK* and *K*s (near-infrared) bands (and other properties derived from SED fitting).**

- **Stellar mass from Hyperzmass code**
- **Line-estimated metallicity for subsamples with high-S/N spectra**

- **Density field reconstructed from galaxy distribution**
- **Cluster/group membership**

- **Ancillary data (Herschel, WISE, etc.)**

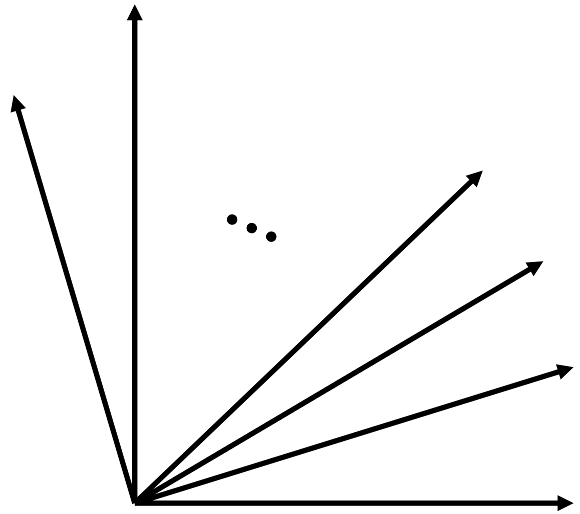
4. Classification in Luminosity Space

4.1 Classification in multiwavelength luminosity space

Multiwavelength luminosity space

Construct a subsample with high-S/N data: 52,114
Redshift range: 0.4 < *z* < 1.3
Twelve rest-frame magnitudes: FUV, NUV, *u*, *g*, *r*, *i*, *z*, *B*, *V*, *J*, *H*, *K*s, normalized around unity.

Perhaps impossible to classify by intuition.



Unsupervised machine learning in luminosity space

Fisher Expectation-Maximization (FEM) algorithm (Bouveryron & Brunet 2012)

1. **Assign initial cluster (class) centers by *k*-means++**
2. **Execute FEM**
   1. **E-step: calculate the complete log-likelihood under the current value of the Gaussian mixture model**
   2. **F-step: boundary is chosen to maximize the distances between groups, and to minimize the internal scatters**
   3. **M-step: parameters of Gaussian functions are optimized by maximizing the conditional expectations of the complete log-likelihood**
   4. **Back to 2.i (E-step) until the result converges.**

Unsupervised machine learning in luminosity space

Fisher Expectation-Maximization (FEM) algorithm (Bouveryron & Brunet 2012)

1. Assign initial cluster (class) centers by *k*-means++
2. Execute FEM
   i. E-step: calculate the complete log-likelihood under the current value of the Gaussian mixture model
   ii. F-step: boundary is chosen to maximize the distances between groups, and to minimize the internal scatters
   iii. M-step: parameters of Gaussian functions are optimized by maximizing the conditional expectations of the complete log-likelihood
   iv. Back to 2.i (E-step) until the result converges.

Unsupervised machine learning in luminosity space

Fisher Expectation-Maximization (FEM) algorithm (Bouveryron & Brunet 2012)

1. Assign initial cluster (class) centers by *k*-means++
2. **Execute FEM**
   i. E-step: calculate the complete log-likelihood under the current value of the Gaussian mixture model
   ii. **F-step: boundary is chosen to maximize the distances between groups, and to minimize the internal scatters**
   iii. M-step: parameters of Gaussian functions are optimized by maximizing the conditional expectations of the complete log-likelihood
   iv. Back to 2.i (E-step) until the result converges.

Unsupervised machine learning in luminosity space

Fisher Expectation-Maximization (FEM) algorithm (Bouveryron & Brunet 2012)

1. **Assign initial cluster (class) centers by *k*-means++**
2. **Execute FEM**
   i. **E-step: calculate the complete log-likelihood under the current value of the Gaussian mixture model**
   ii. **F-step: boundary is chosen to maximize the distances between groups, and to minimize the internal scatters**
   iii. **M-step: parameters of Gaussian functions are optimized by maximizing the conditional expectations of the complete log-likelihood**
   iv. **Back to 2.i (E-step) until the result converges.**

Unsupervised machine learning in luminosity space

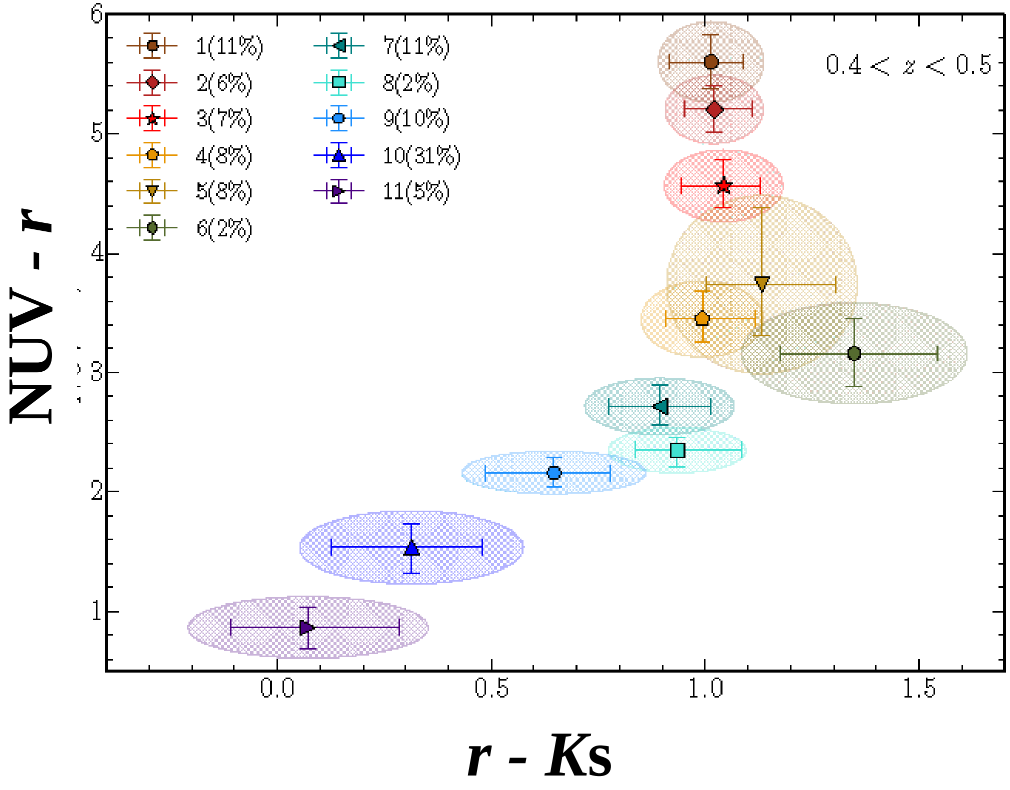
Fisher Expectation-Maximization (FEM) algorithm (Bouveryron & Brunet 2012)

1. Assign initial cluster (class) centers by *k*-means++
2. Execute FEM
   i. E-step: calculate the complete log-likelihood under the current value of the Gaussian mixture model
   ii. F-step: boundary is chosen to maximize the distances between groups, and to minimize the internal scatters
   iii. M-step: parameters of Gaussian functions are optimized by maximizing the conditional expectations of the complete log-likelihood
   iv. Back to 2.i (E-step) until the result converges.

4.2 Classification result

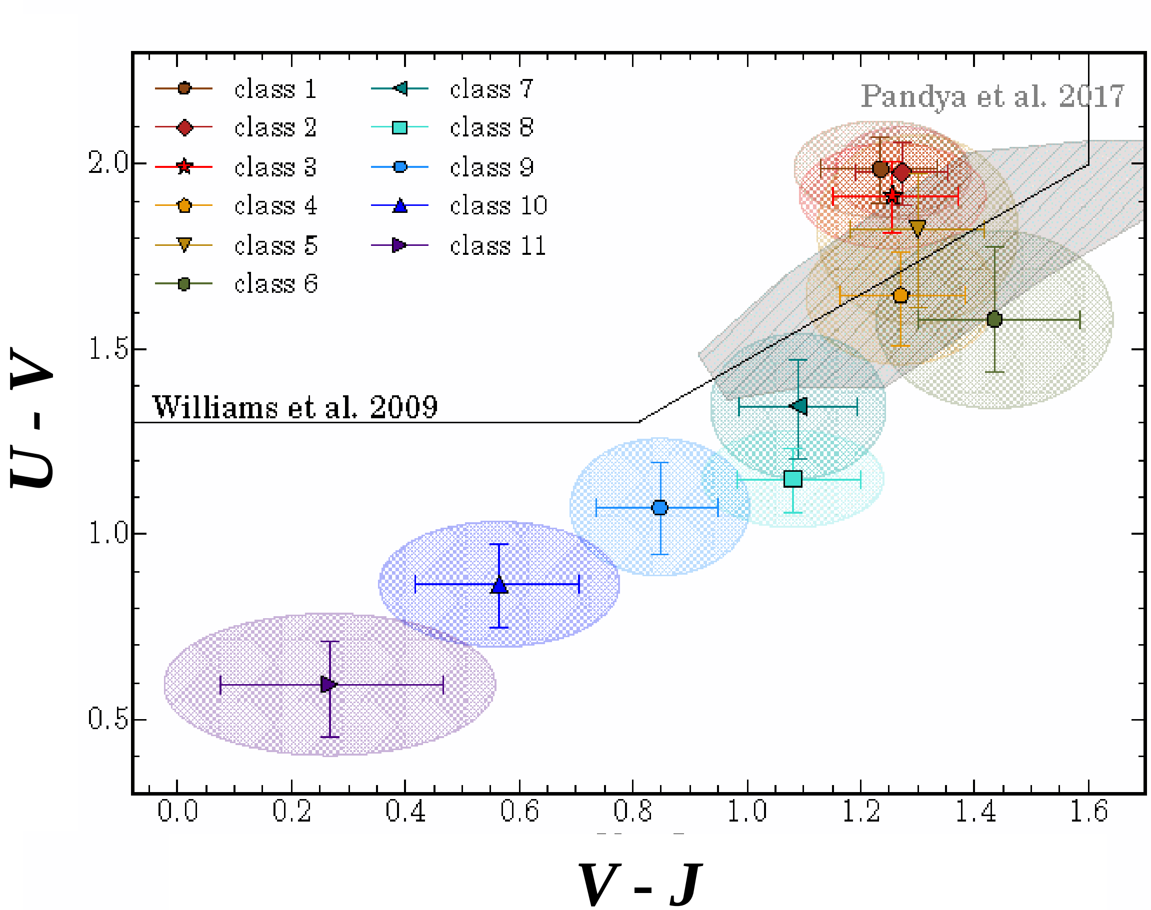
The FEM classification separates VIPERS galaxies into twelve classes.

N.B. **Twelfth class is the AGN and not included in the further discussion.**



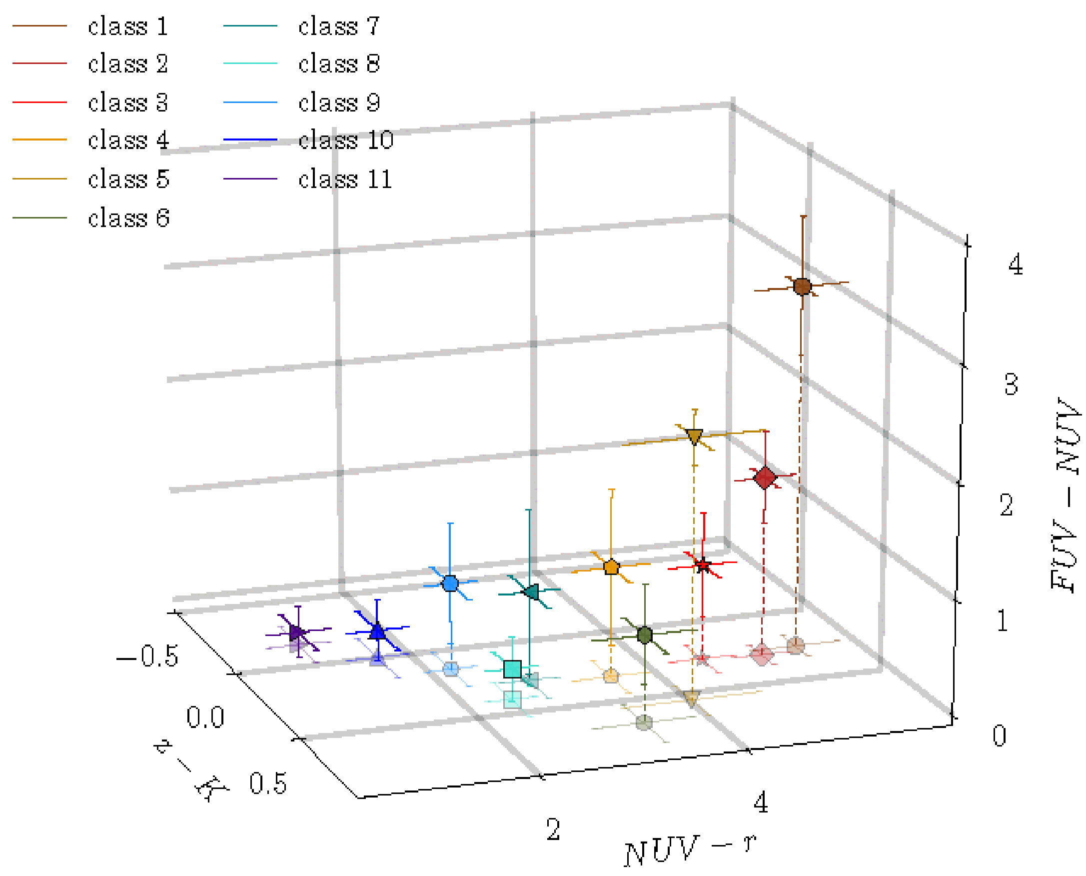
The *UVJ* diagram

Williams et al. (2009) and subsequent authors proposed that the *UVJ* diagram provide a separation between passive and star-forming galaxies.



The NUV*rK* 3D diagram

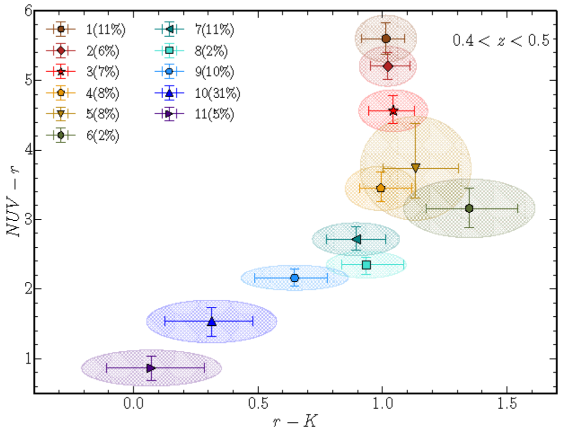
The red galaxies are divided into three subclasses, separated by the FUV - NUV color.



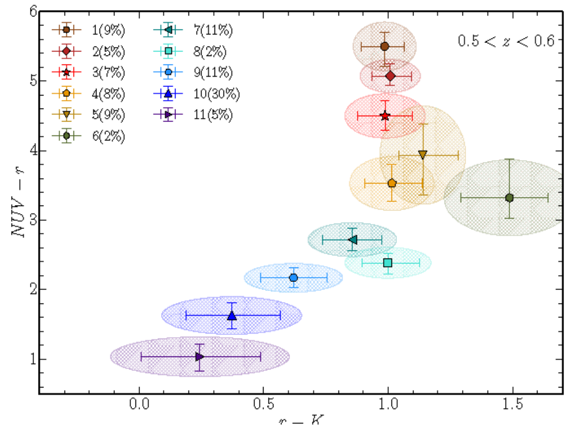
The redshift (in)dependence at $0.4 < z < 1.0$

The color evolution of galaxies is clearly visible, but the classification is almost unchanged with redshifts.

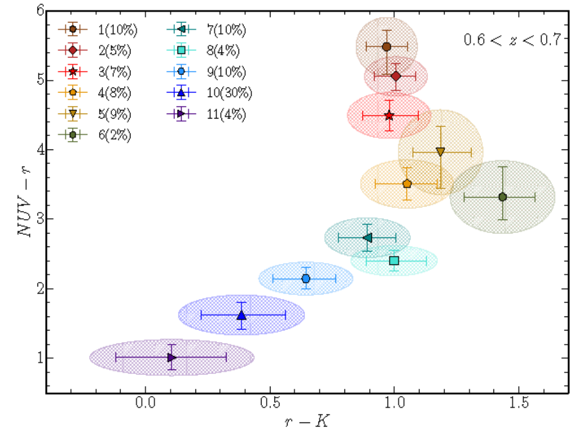
$0.4 < z < 0.5$



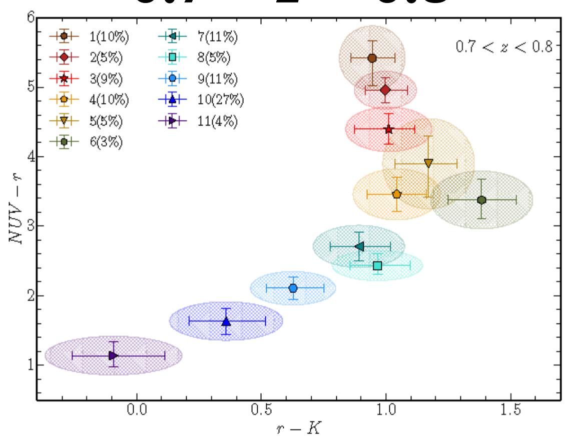
$0.5 < z < 0.6$



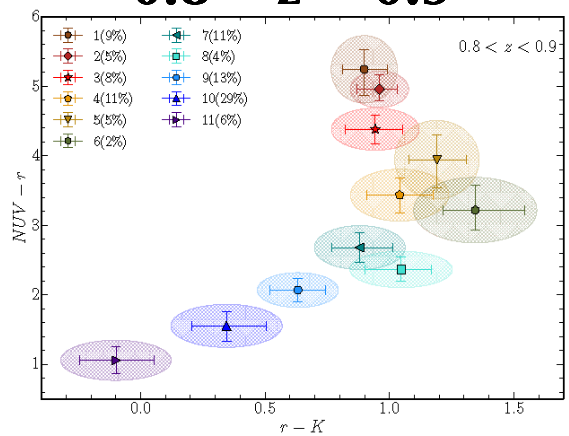
$0.6 < z < 0.7$



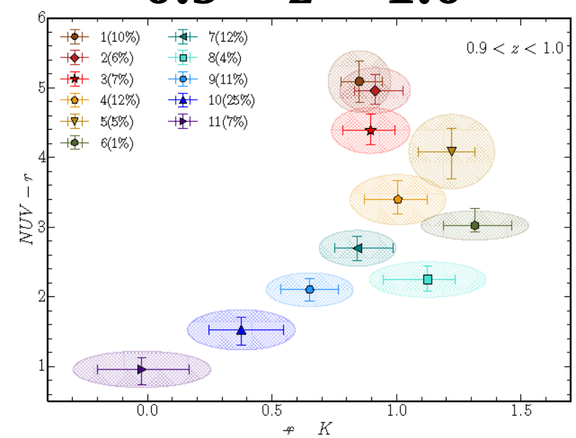
$0.7 < z < 0.8$



$0.8 < z < 0.9$

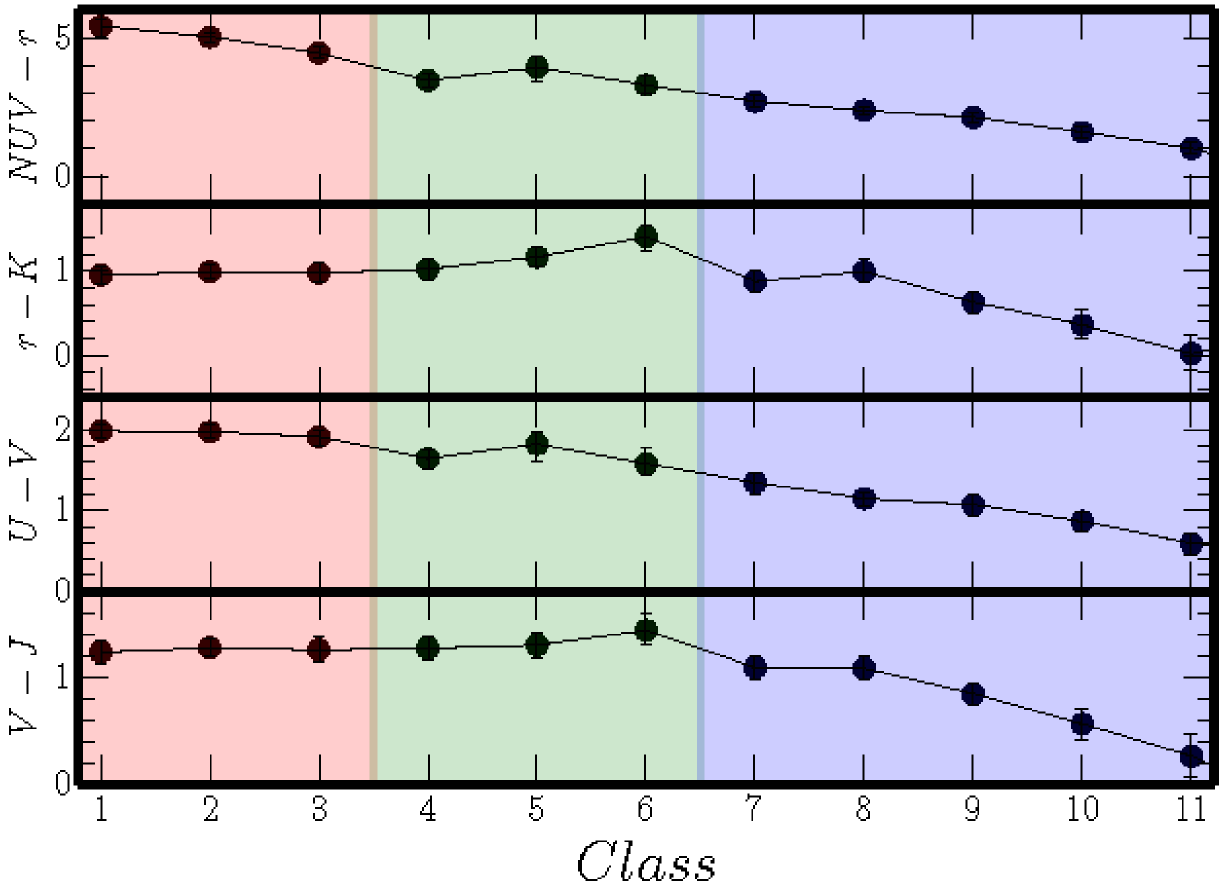


$0.9 < z < 1.0$



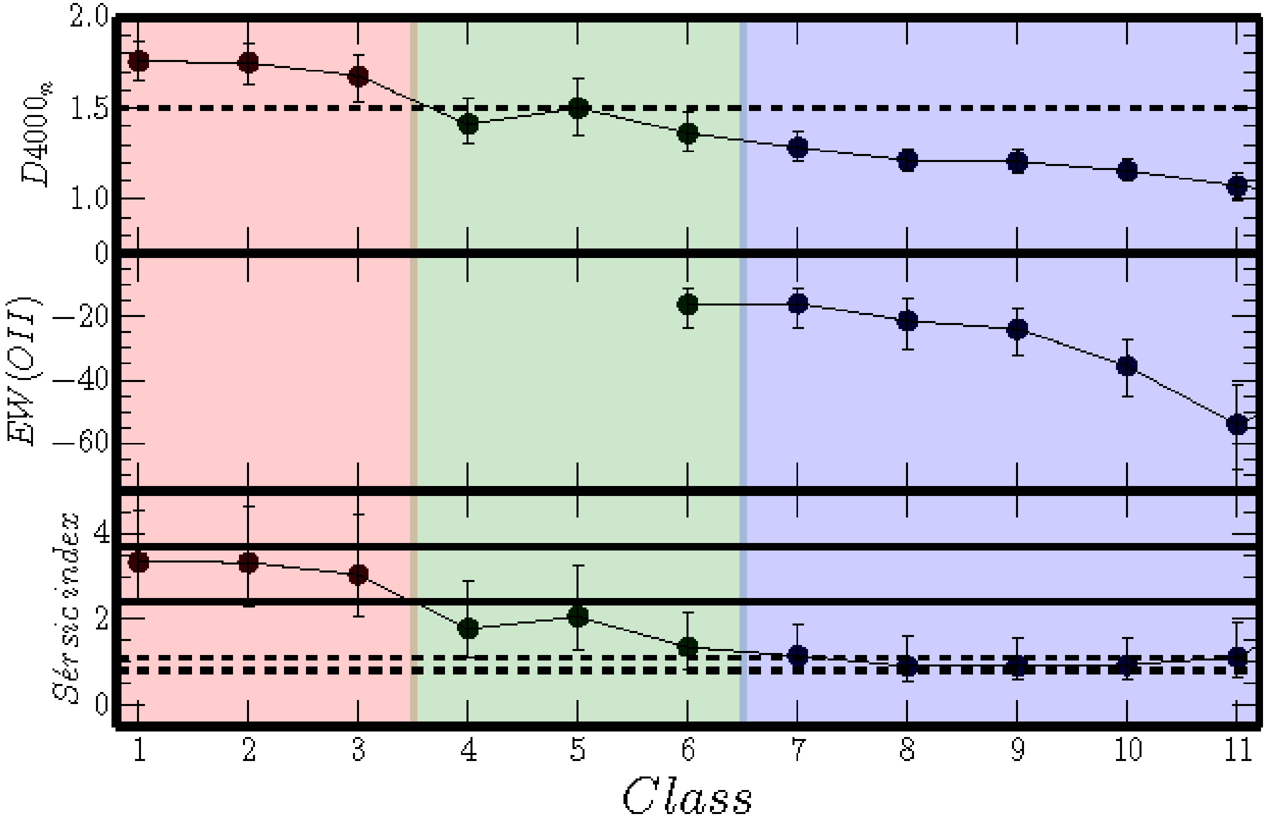
4.3 Global properties of classified galaxies

Colors



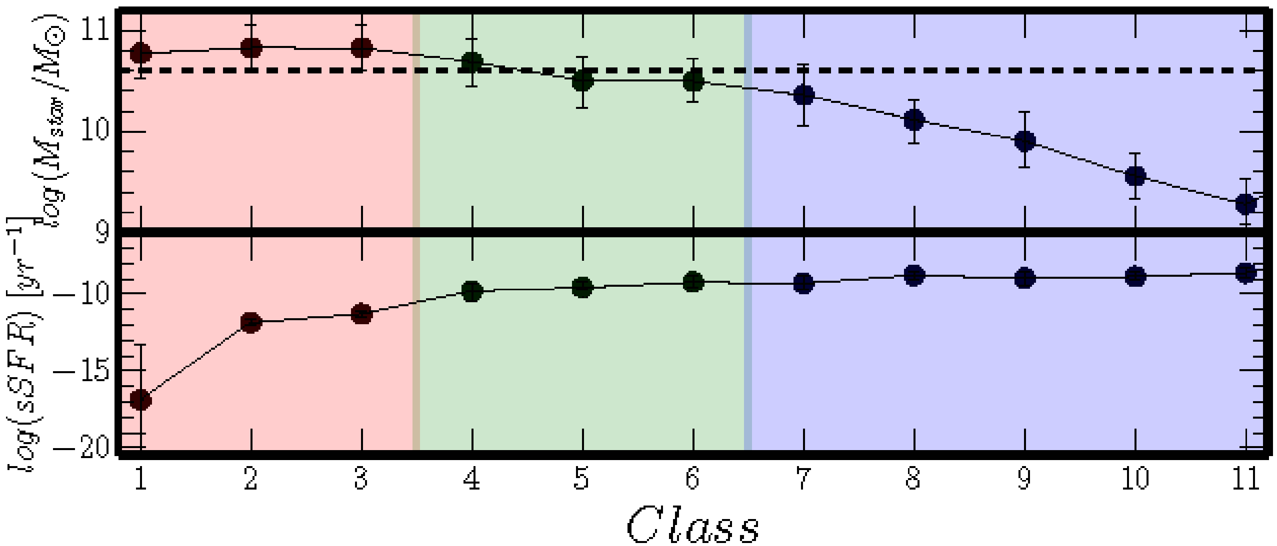
A turnover exists in some colors.

Spectral and morphological features



D4000: indicator of galaxy age (relatively new SFH)
EW[OII]: emission from SF region
Sérsic index: shape parameter indicating early (~ 4) or late (~ 1) type

Stellar mass and specific SFR (sSFR)

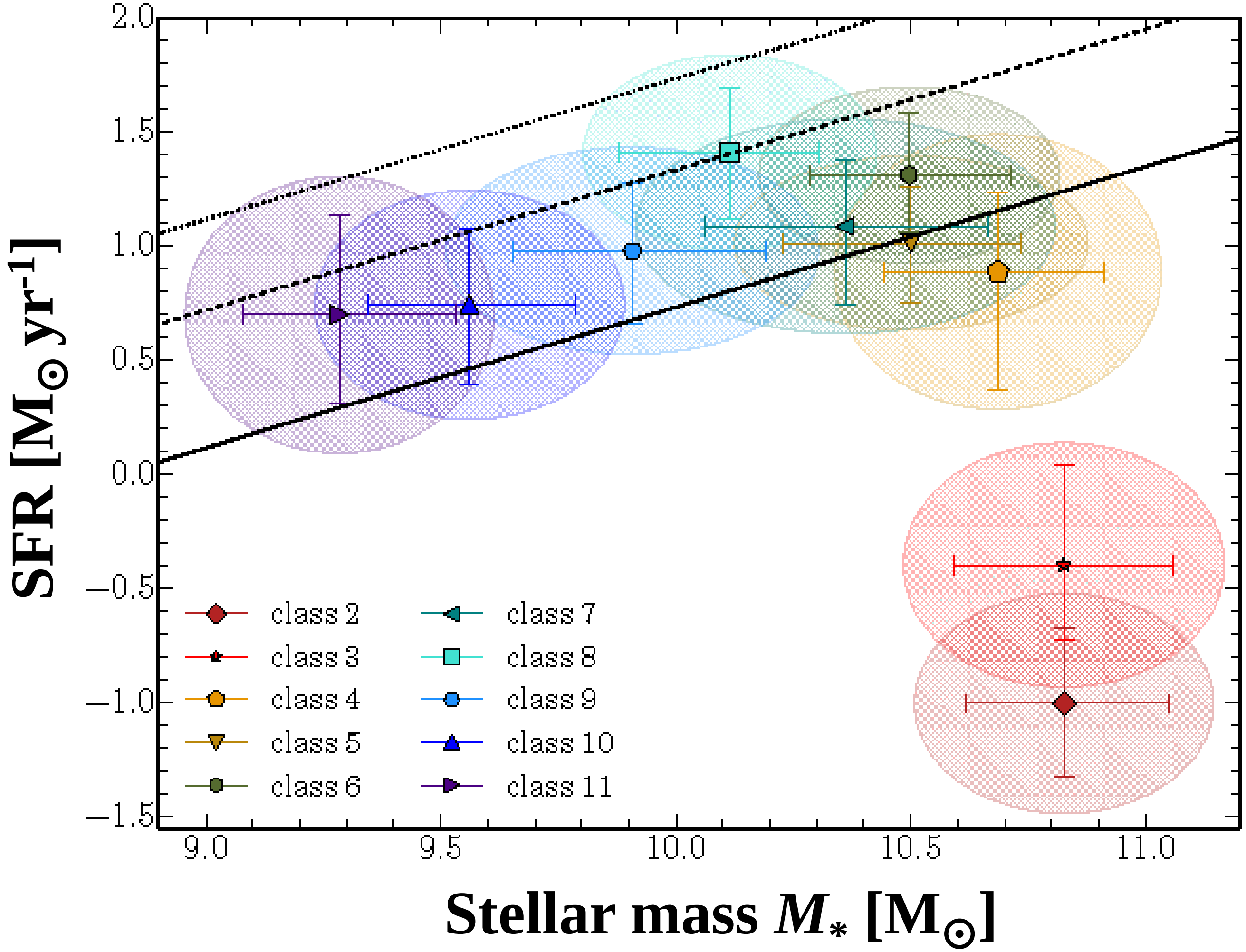


sSFR: SFR per stellar mass [yr^{-1}]

$$\mathrm{sSFR} = \frac{\mathrm{SFR}}{M_*}$$

Indicator of the SF activity normalized with the galaxy size.

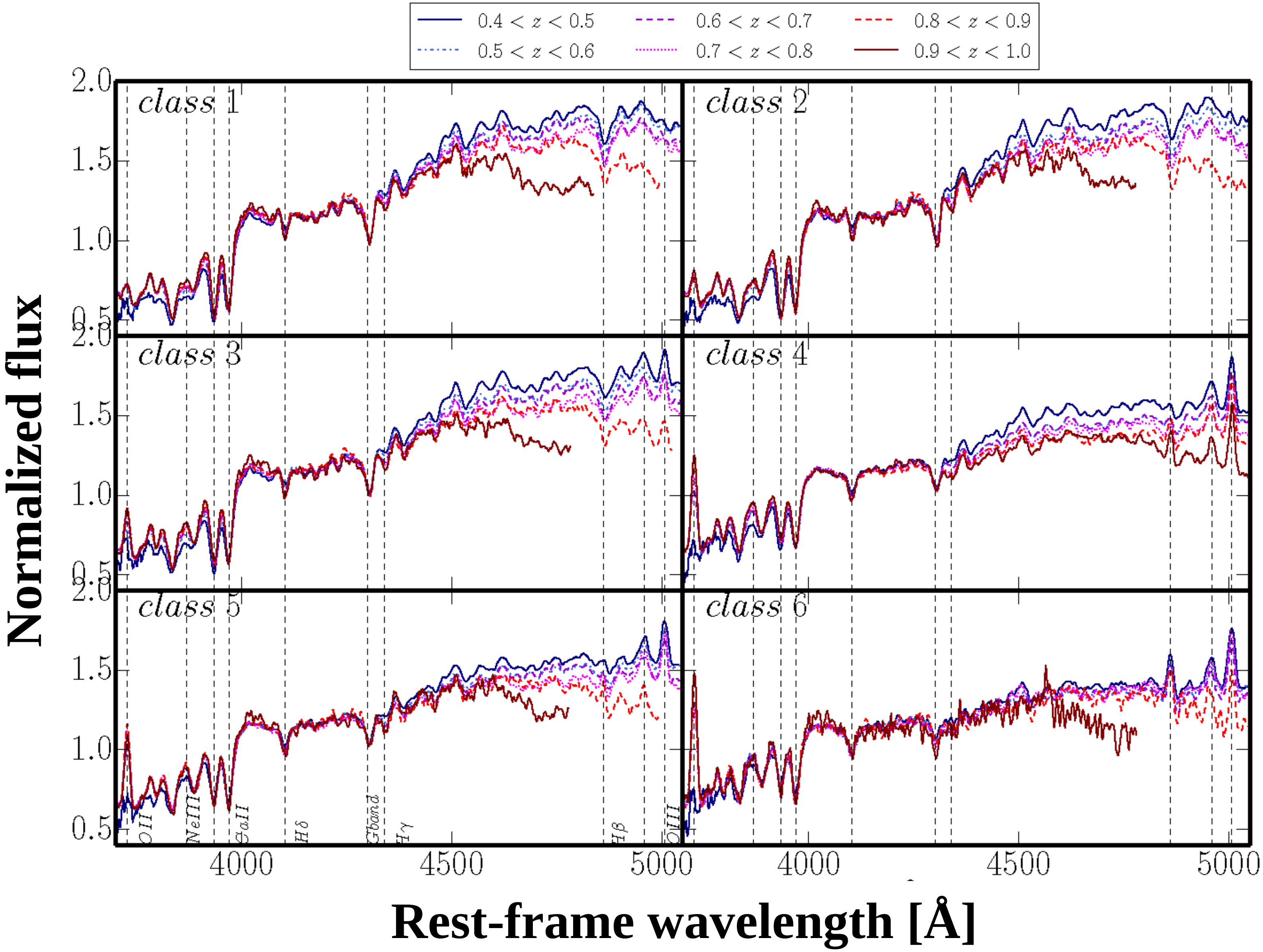
Main sequence of star-forming galaxies



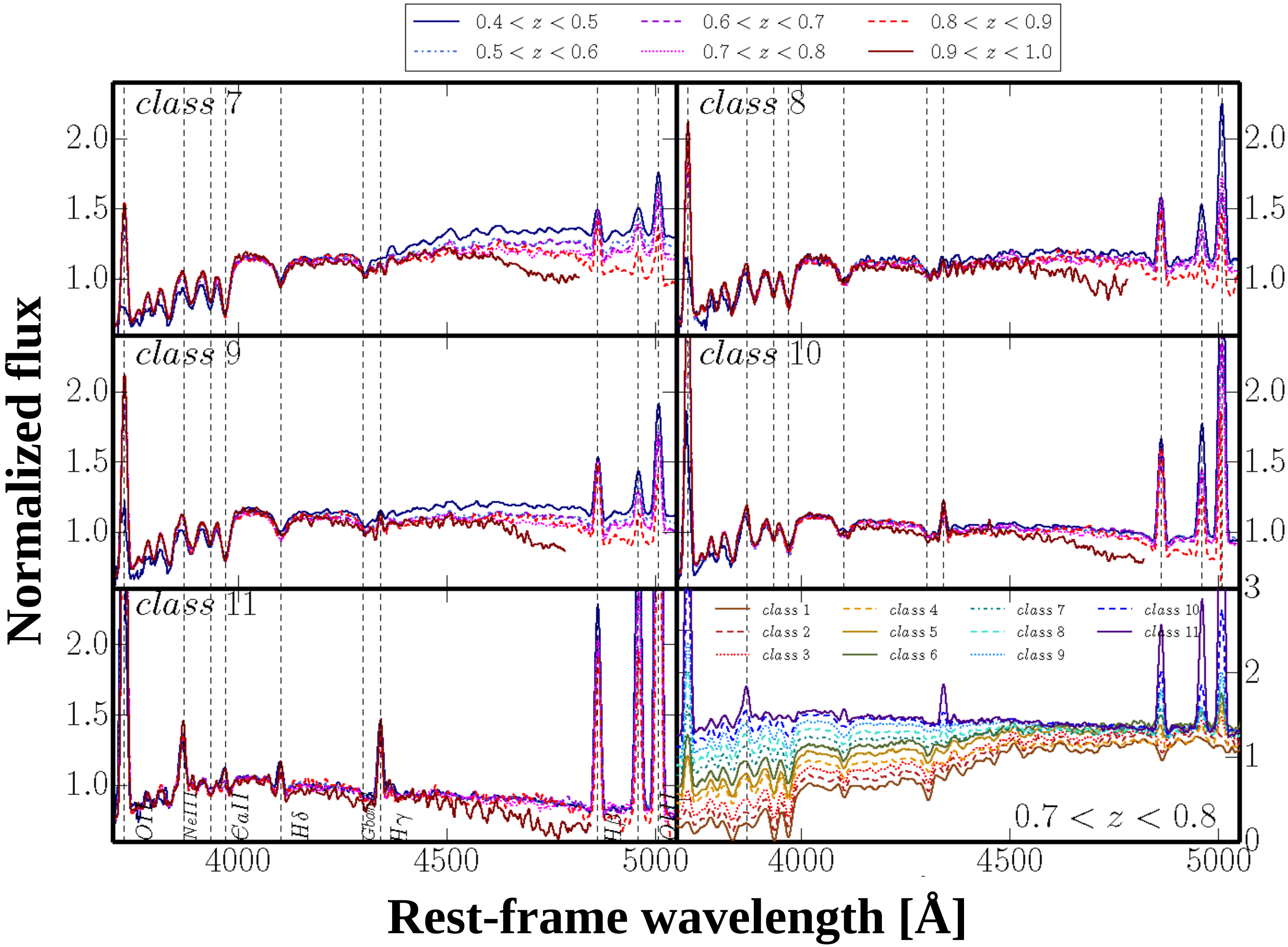
The main sequence is naturally reproduced from this method. Note that this is free from color selection bias, which introduced a significant confusion in this diagram.

4.4 Spectral properties of classified galaxies

Stacked spectra of class 1-6



Stacked spectra of class 7-11



Summary of spectral features in classified galaxies

Class 1-3 (red)

Strong absorption features, D4000 break, red continuum

Class 4-6 (green)

Balmer absorption, gradually stronger emission lines

Class 7-11 (blue)

Strong emission lines, blue continuum

Redshift dependence

In each class, higher-*z* galaxies have gradually bluer continuum, significant at longer wavelengths

5. Conclusions

With VIPERS multicolor galaxy photometric data with redshifts, we explored the performance of a machine learning method to study galaxy evolution.

1. **To avoid complicated selection effects and subjective choice of features, we abandoned the color-based method. Instead, we performed an unsupervised machine learning classification directly to the twelve-dimensional luminosity-redshift space of VIPERS galaxies.**

2. **Our classification yielded twelve galaxy classes (+ one AGN class). Class 1-3 corresponds to red passive galaxies, 4-6 to “green” galaxies, and 7-11 to blue star-forming galaxies, respectively.**

5. Conclusions

3. **In the redshift range of 0.4 < z < 1.0, the classes remain unchanged with a gradual shift to bluer colors.**

4. **Classes 1-3 show stronger absorption lines and D4000 and Sérsic index *n* ~ 4 (spheroid-like), indicating old stellar population. Classes 7-11 show strong [OII] emission and *n* ~ 1 (disk-like), consistent with active SF.**

5. **Galaxies in each class showed clearly distinguishable physical properties, as a sequence of features. Stellar mass decreases from class 1 to 11. In contrast, specific SFR increases with the class labels.**

5. Conclusions

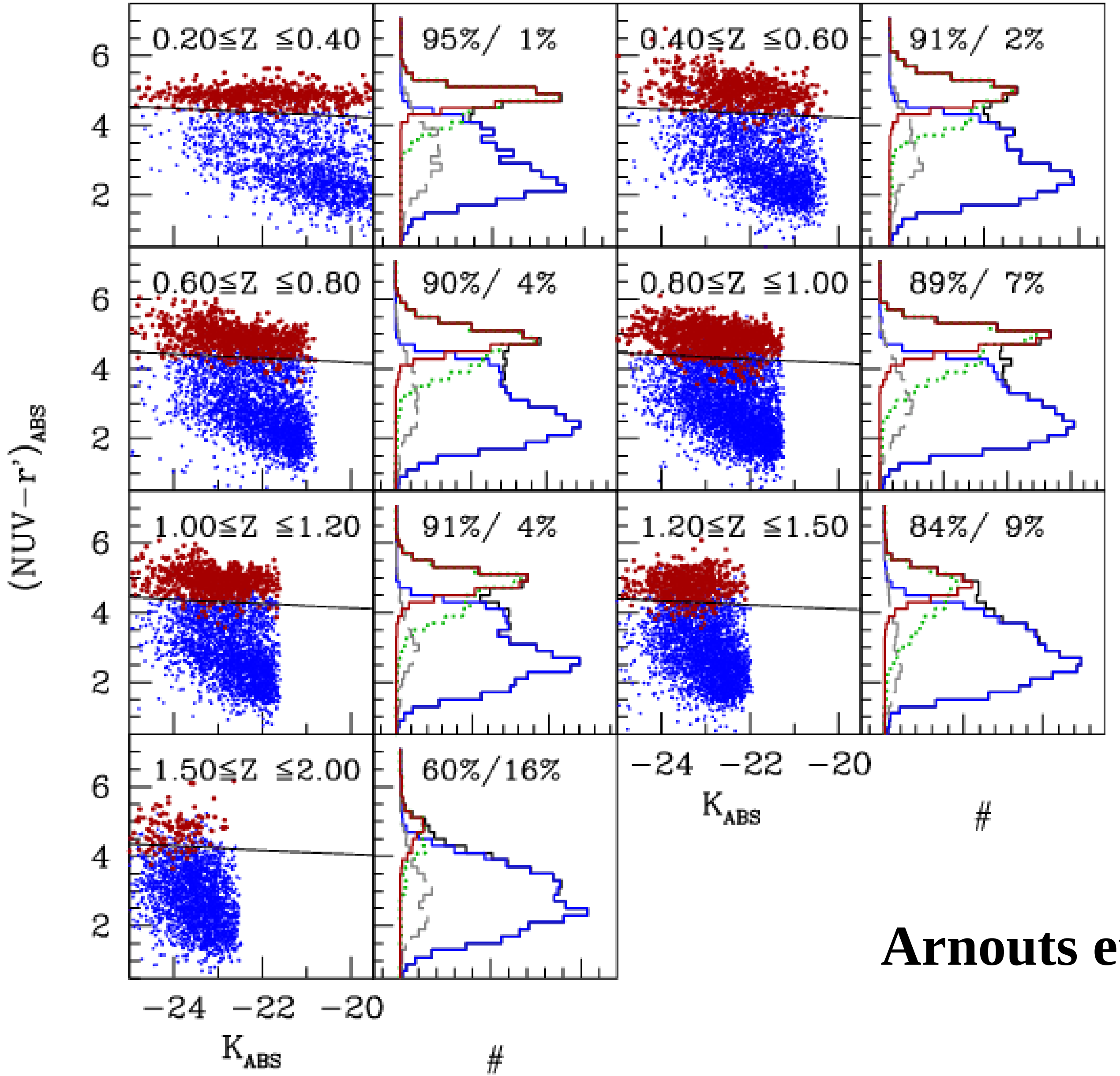
6. **The star-forming galaxy main sequence is clearly reproduced by classes 7-11, with class 1-3 as quiescent non-star forming galaxies. Note that, in contrast to previous studies, our result is not affected by the complicated color selection and robust.**

7. **Classified galaxies demonstrated a clear representative spectrum for each class. The spectra are redshift-dependent: higher-*z* galaxies have gradually bluer continuum, corresponding to the evolution.**

Take-home message

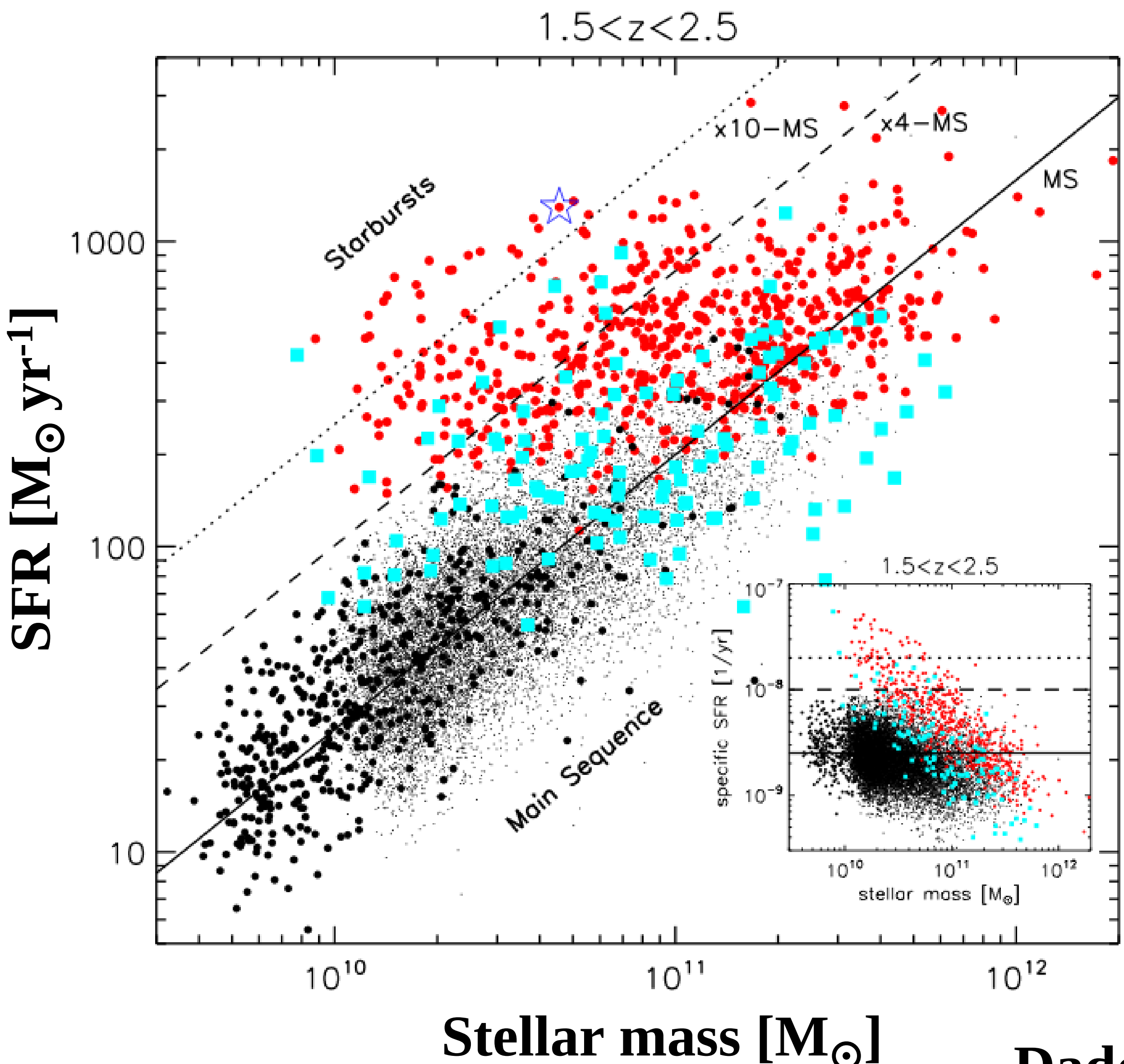
Machine learning will open a new window to redefine galaxy evolution, and provide us with a substantially new discovery.

Evolution of the color-magnitude relation



Arnouts et al. (2007)

Star formation main sequence of galaxies (high-z)



Daddi et al. (2007)

VIPERS Public Data Release

http://vipers.inaf.it/

VIPERS
VIMOS PUBLIC EXTRAGALACTIC REDSHIFT SURVEY

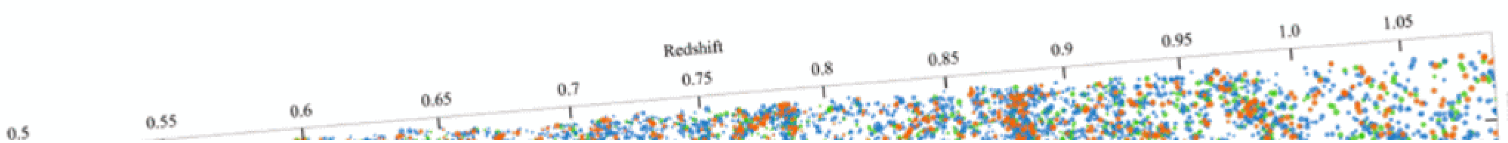
HOME | TEAM | PROJECT | GALLERY | PAPERS | DATA PDR-2 | INTERNAL

Final public release of complete VIPERS galaxy catalogue of ~90,000 redshifts (PDR-2)

- 18 November 2016 -

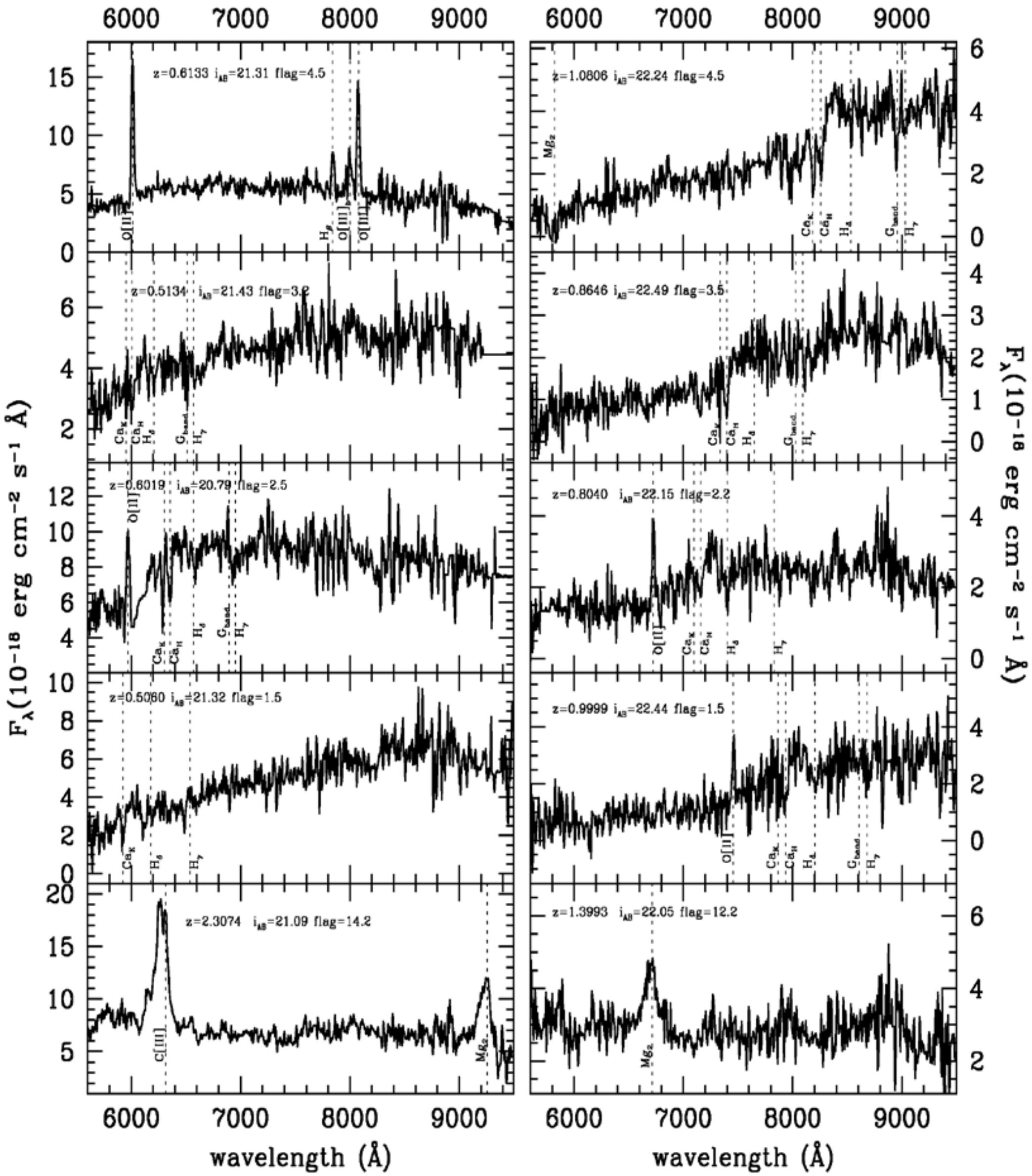
Go to PDR-2 data download page

For the press: final science release information page



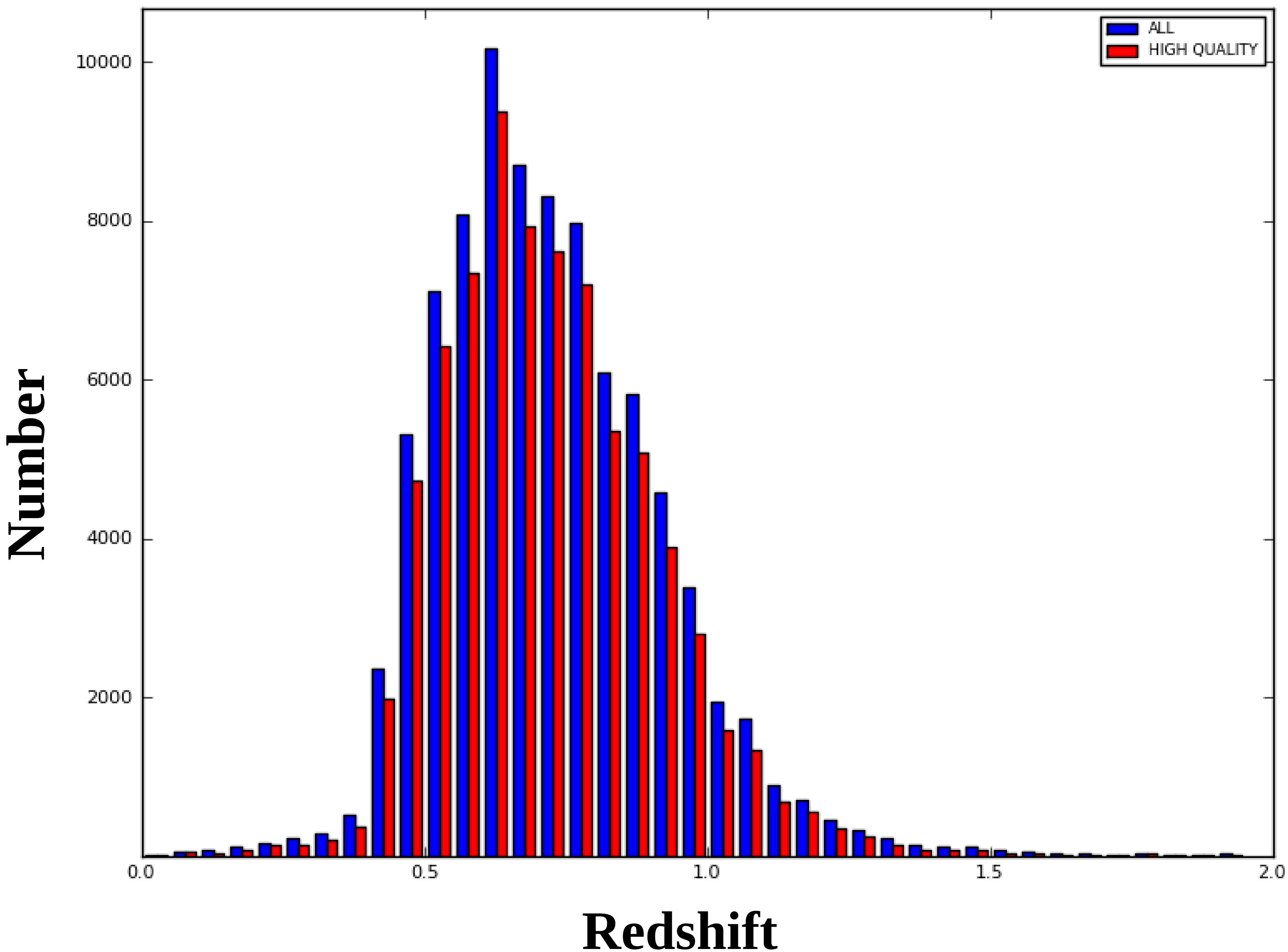
3D distribution of VIPERS galaxies

VIPERS spectra: examples



Guzzo et al. (2013); Garilli et al. (2014)

VIPERS redshift distribution



k-means++ algorithm

1. **Choose one center uniformly at random from sample data points.**

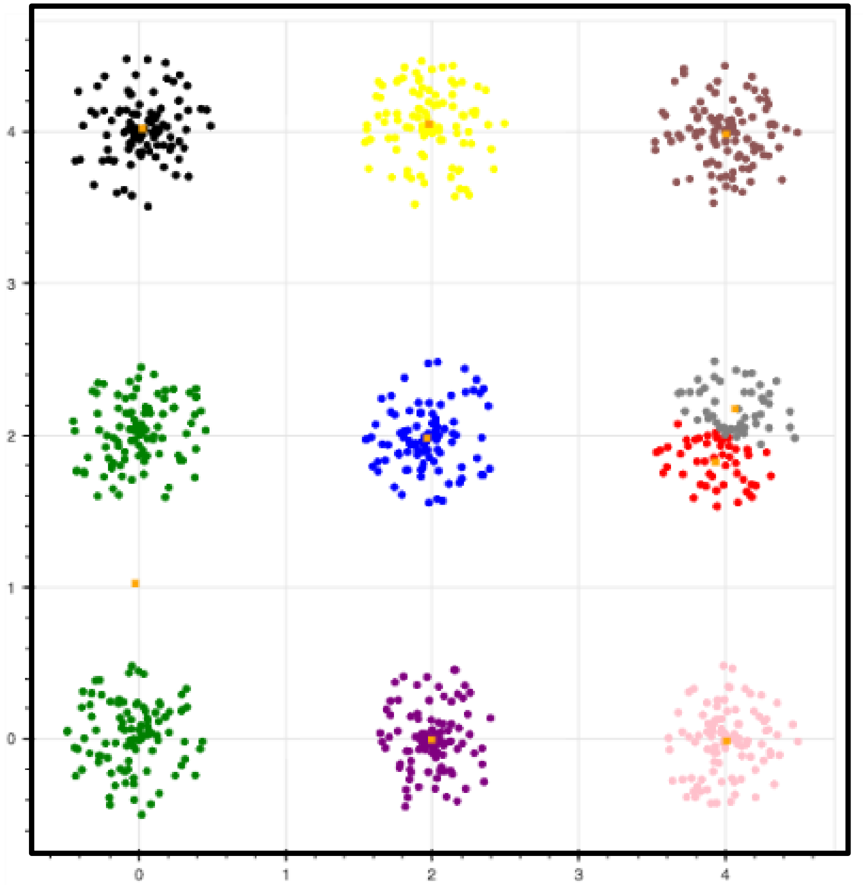
2. **For each data point x, compute $D(x)$, the distance between x and the nearest center that has already been chosen.**

3. **Choose one new data point at random as a new center, using a weighted probability distribution where a point x is chosen with probability proportional to $D(x)^2$.**

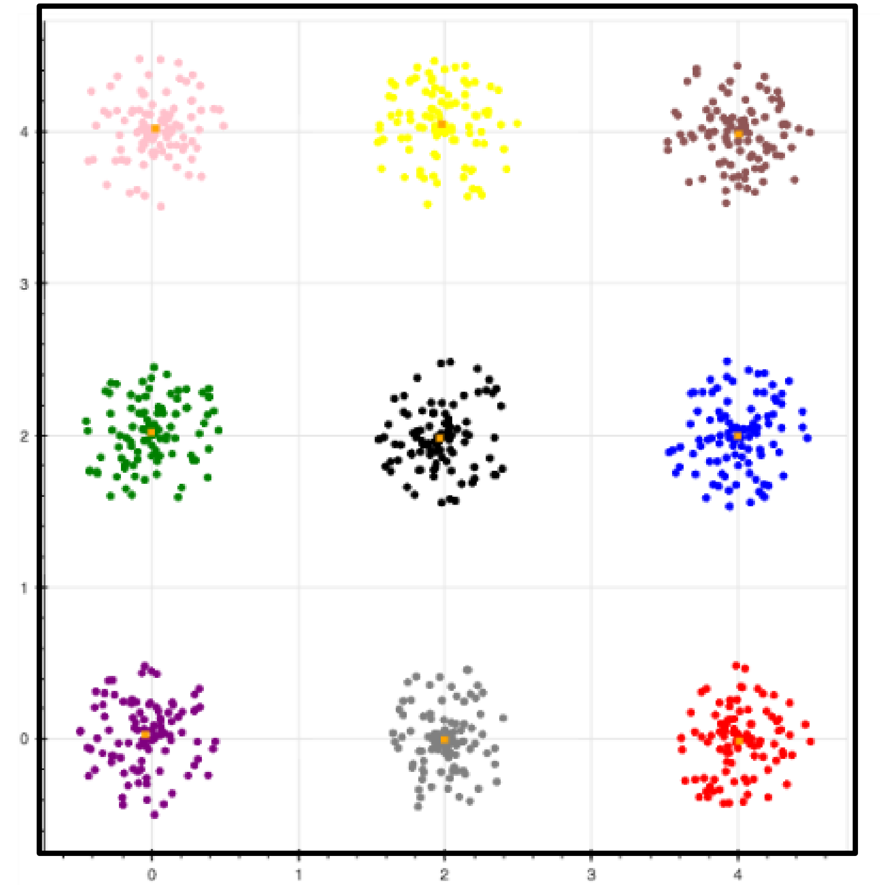
4. **Repeat Steps 2 and 3 until k centers have been chosen.**

5. **Now that the initial centers have been chosen, proceed using standard k-means clustering.**

k-means++ algorithm



***k*-means**



***k*-means++**

https://qiita.com/NoriakiOshita/items/cbd46d907d196efe64a3

Gaussian mixture model

For *m* variables and *n* Gaussians,

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)$$

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k) = \frac{1}{(2\pi)^{\frac{m}{2}} (\det \boldsymbol{\sigma}_k)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^{\mathrm{T}} \boldsymbol{\sigma}_k (\boldsymbol{x} - \boldsymbol{\mu}_k)\right]$$

x : [x_1, x_2, x_3, … x_m]

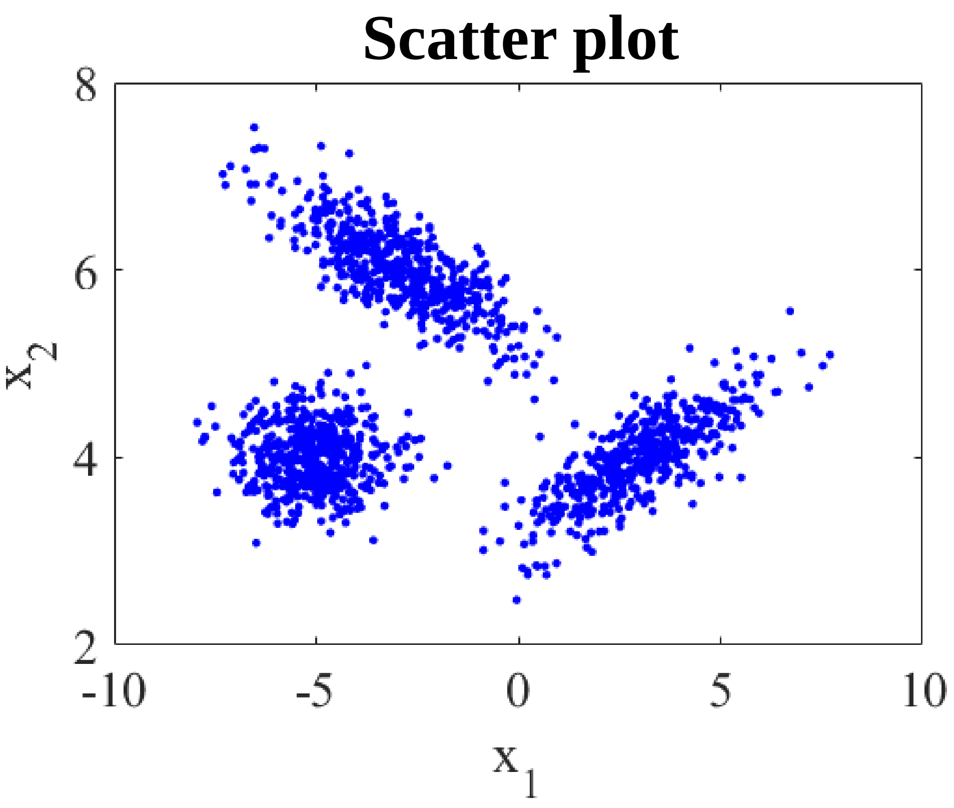
μ_k : mean vector (1 $\times$ *m*) of *k*-th Gaussian

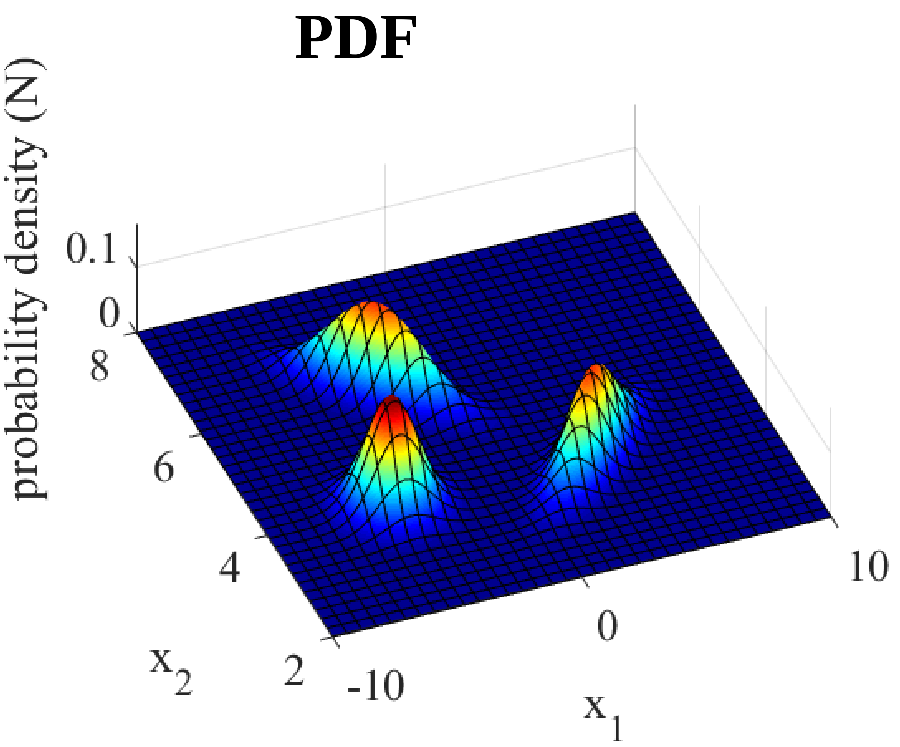
Σ_k : covariance matrix (*m* $\times$ *m*) of *k*-th Gaussian

π_k : mixing coefficient

$$\sum_{k=1}^{n} \pi_k = 1$$

Gaussian mixture model: example



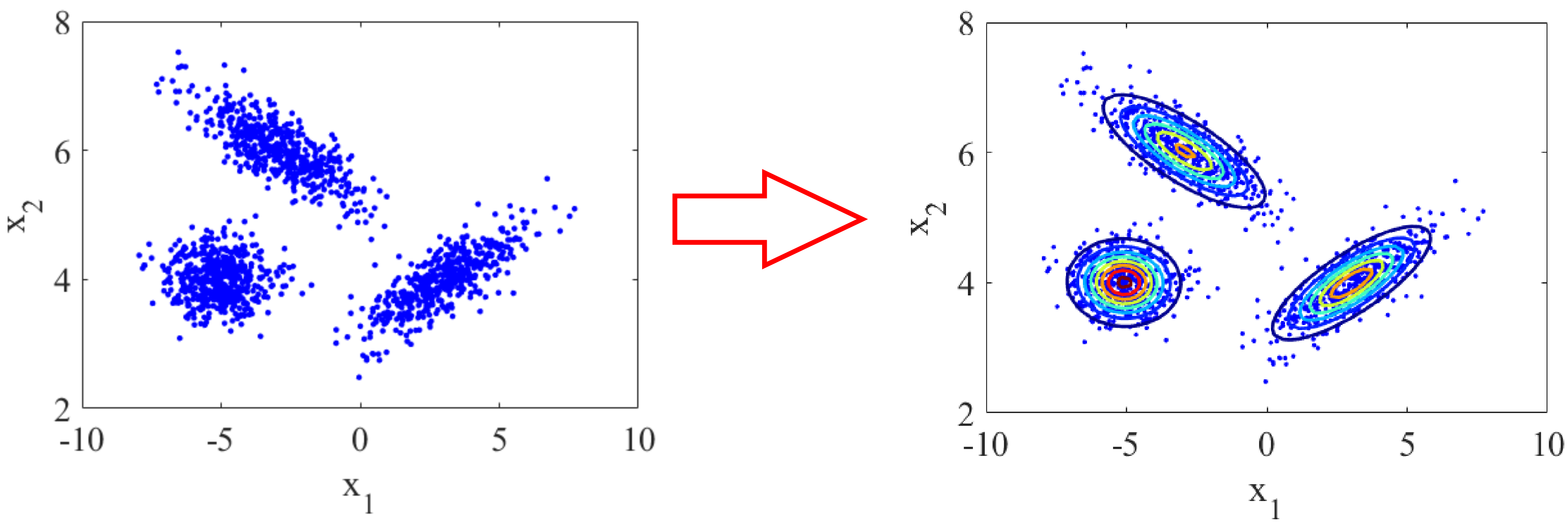


Estimation of the parameters for Gaussian mixture model

Given a dataset x : [x_1, x_2, x_3, … x_m], we estimate

μ_k : mean vector,
σ_k : covariance matrix, and
π_k : mixing coefficient

via maximum likelihood method.



Estimation of the parameters for Gaussian mixture model

When N data are divided into K clusters, a log-likelihood function is

$$\ln p(\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\sigma}) = \ln \prod_{n=1}^{N} \left[\sum_{k=1}^{K} \boldsymbol{\pi}_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\sigma}_k) \right]$$

$$= \sum_{n=1}^{N} \ln \left[\sum_{k=1}^{K} \boldsymbol{\pi}_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\sigma}_k) \right]$$

We calculate a contribution function $\boldsymbol{\gamma}(\boldsymbol{z_{nj}})$

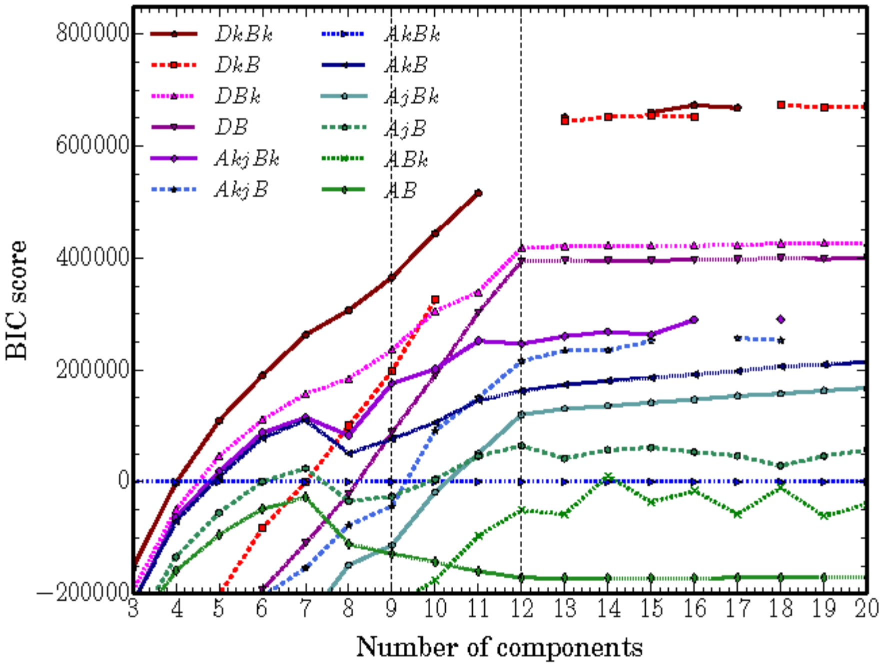
$$\gamma\left(z_{nj}\right) = \frac{\boldsymbol{\pi}_j \mathcal{N}\left(\boldsymbol{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\sigma}_j\right)}{\sum_{k=1}^{K} \boldsymbol{\pi}_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\sigma}_k)}$$

(z : latent variable, $0 \le z \le 1$)

Selection of best model and optimal number of classes

The number of classes is knot known a priori: a major difficulty in unsupervised algorithm.

In this work, the best model and the range of possible classes are chosen based on three statistical model-selection criteria: AIC (Akaike 1974), BIC (Schwarz 1978), and ICL (Baudry 2012).

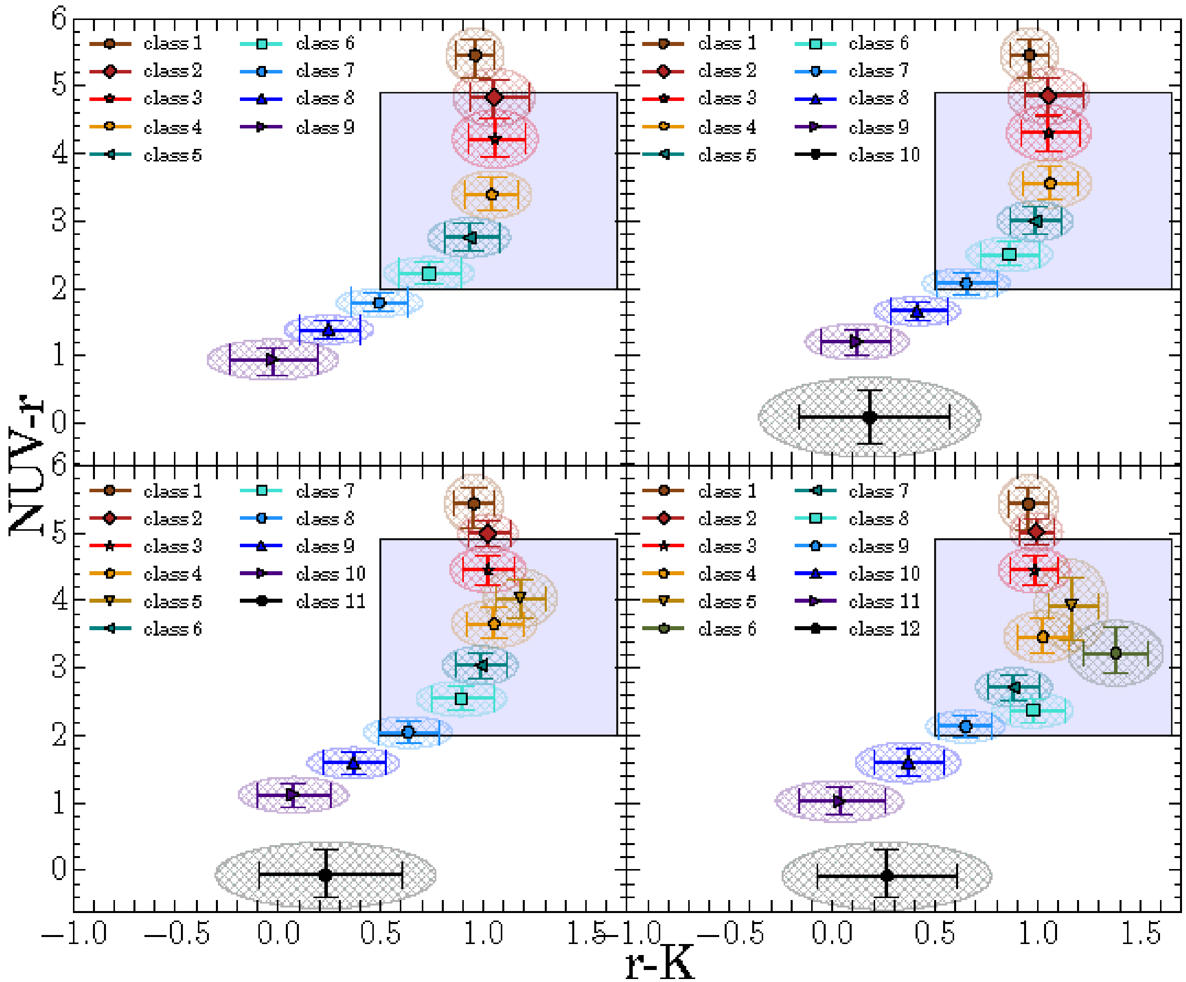


Siudek et al. (2018)

Selection of best model and optimal number of classes

The result can be intuitively verified on a NUV*rK* color-color diagram.

9 classes **10 classes** **11 classes** **12 classes**



Siudek et al. (2018)

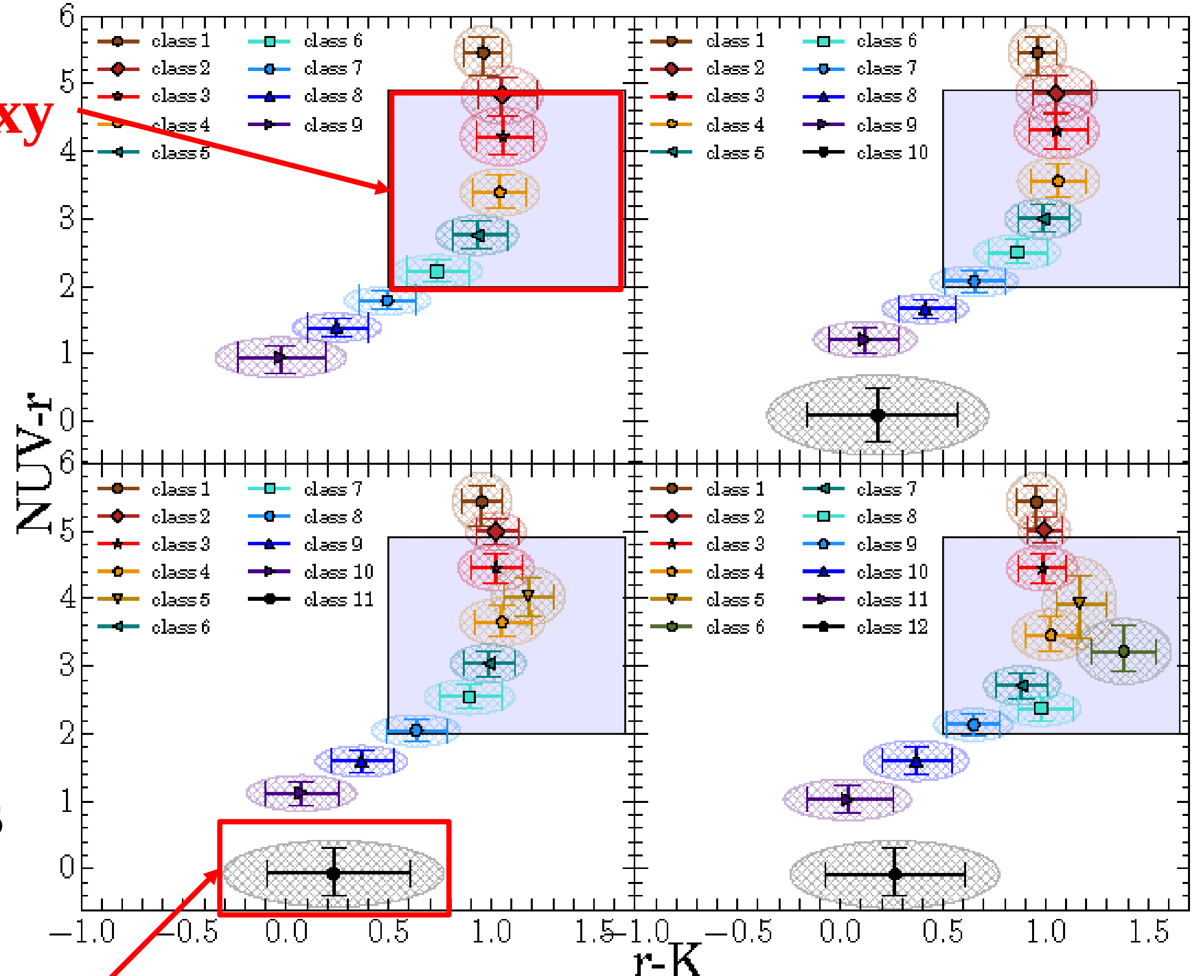
Selection of best model and optimal number of classes

The result can be intuitively verified on a NUV*rK* color-color diagram.

Dusty galaxy

9 classes

10 classes



11 classes

12 classes

High-*z* AGNs

Siudek et al. (2018)

The NUV*rK* diagram

The NUV*rK* diagram (Arnouts et al. 2013) is similar to the *UVJ* plane but allows for a better separation between passive and star-forming galaxies.