Inferring spectroscopic features of galaxies from their broadband colors using Self Organizing Maps

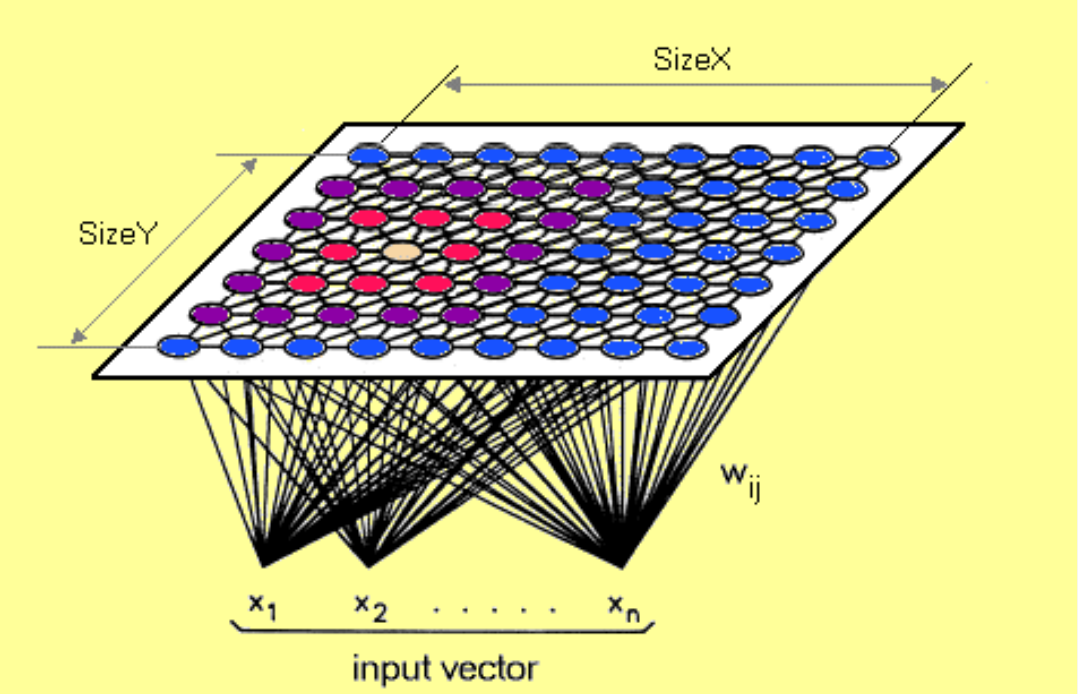
Marziye Jafariyazani

Postdoctoral Scholar Research Associate at IPAC/Caltech

Advisors: Dan Masters, Harry Teplitz and Andreas Faisst

Galaxy Evolution with the ESA Euclid Mission and ESO Telescopes
European Space Astronomy Centre
October 2022

Self-Organazing Maps (SOMs)



Reduce complex problems for easy interpretation

- Type of artificial neural network trained using unsupervised learning
- Used for dimensionality reduction, data visualization and clustering

What happens in SOM algorithm?

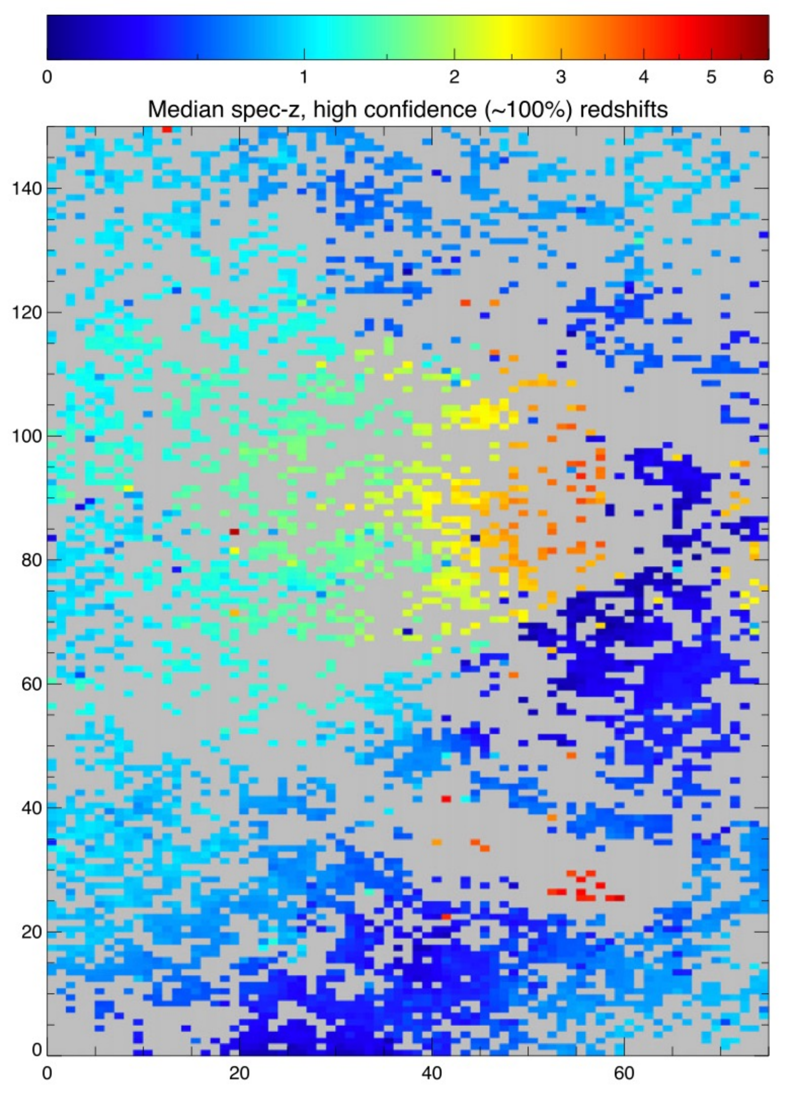
- Assume an input data with
  - n training examples, and
  - m features for each example
- Start with a 2D (X-D) map. Each piece on the map represents a **neuron.**
- Initialize these neurons randomly.
- Present each input data to the map and find the **best matching unit (BMU)**.
  - The **BMU** is the neuron that optimizes the value of a formula, usually the **Euclidean distance**, between the values of the example and the values held by the neuron.

Applications of SOMs in Astronomy

Masters et al. 2015:

Optimal photometric redshift calibration strategies for cosmology surveys:

- Training SOMs with existing photometric data from the COSMOS survey selected to approximate the anticipated Euclid observations.
- Gray regions correspond to parts of galaxy color space for which no high-confidence spectroscopic redshifts currently exist.
- ***These regions will be of interest for spectroscopic observations for training and calibration.***

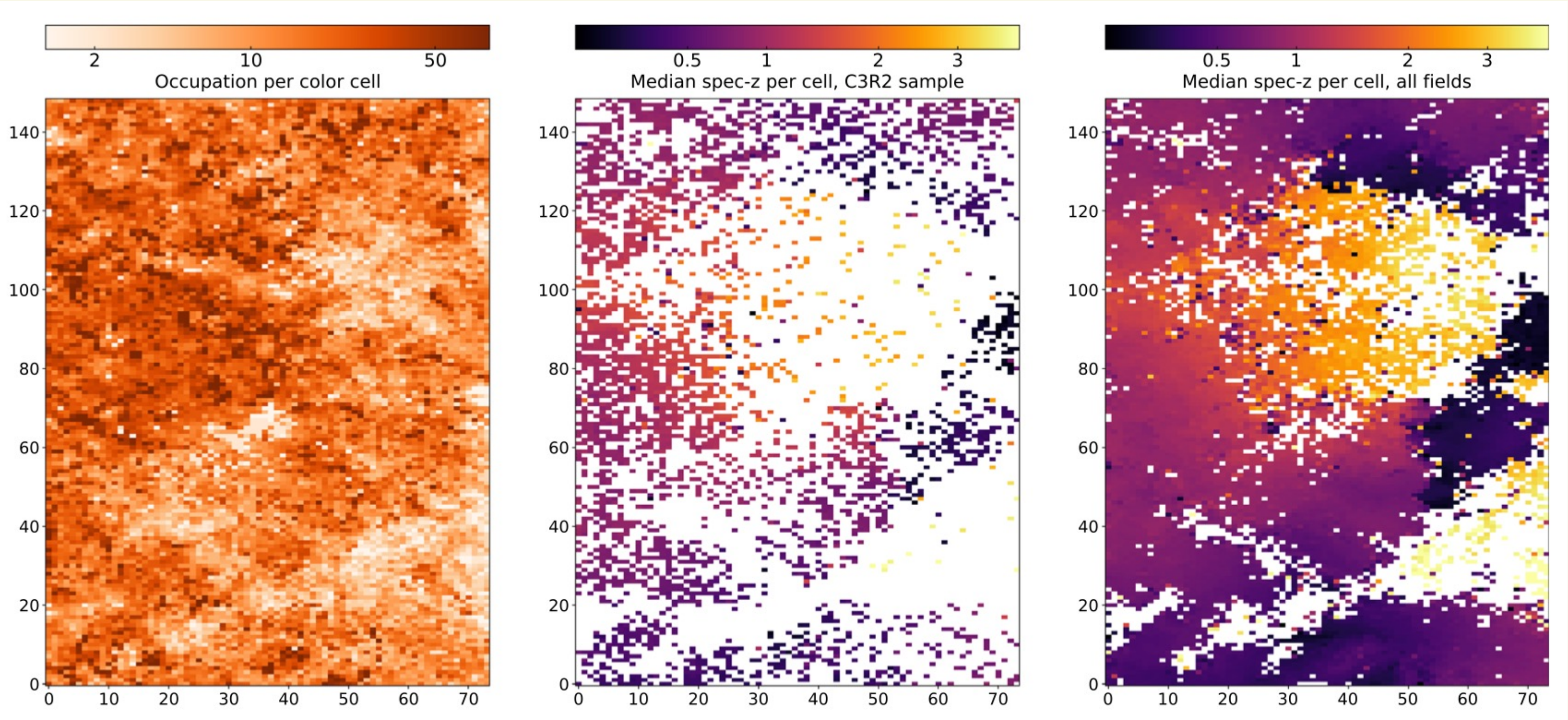


The median spectroscopic redshift of galaxies associating with each SOM cell, using very high confidence redshifts from the COSMOS master spectroscopic catalog (come from a variety of surveys that have targeted the COSMOS field).

Applications of SOMs in Astronomy

Masters et al. 2019:

The Complete Calibration of the Color-Redshift Relation (C3R2): aims to map the empirical relation of galaxy color to redshift to i ~ 24.5



Left: *Occupation density of galaxies on the SOM.*
Center: *C3R2 coverage of the color space. C3R2 has successfully observed >35% of the color cells.*
Right: *The current total coverage of the SOM, incorporating C3R2 with spectra from numerous other surveys. Roughly 76% of color cells are covered, while ~85% of galaxies live within calibrated cells.*

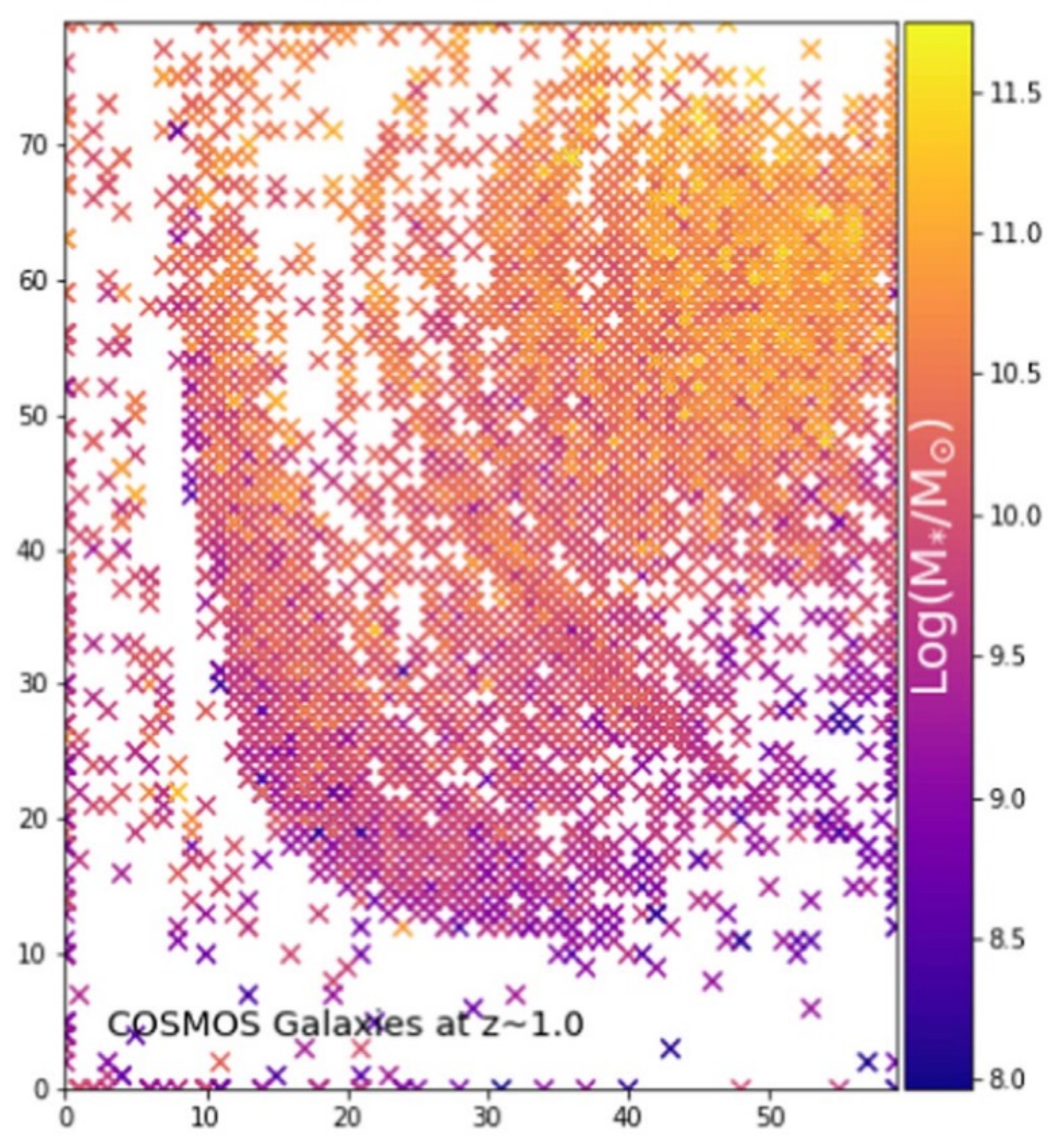
Applications of SOMs in Astronomy

Hemmati et al. 2019:

Tool for both visualizing and computationally accelerating the estimation of galaxy physical properties from photometric data:

- SOMs are used to estimate the physical parameters of 14,000 z~1 galaxies in the COSMOS field:

- **This method was able to estimate the full probability distribution functions for each galaxy up to $\sim 10^6$ times faster than direct model fitting.**



COSMOS galaxies at z~1 are mapped to the SOM trained with theoretical models. Empty regions of the SOM represent combinations of the theoretical parameter space that do not have observed counterparts

Train SOMs with broadband colors and inform them with spectroscopic observations to extract specific spectroscopic features.

COSMOS 2020

Multiband photometry (UV-optical-IR)

for 1.7 million sources

LEGA-C: The Large Early Galaxy Census

Spectroscopic Survey of ~3200 *K*-band selected galaxies

Conducted with VLT/VIMOS

Redshift: $0.6 < z < 1$

(K=21.08 at z=0.6 to K = 20.36 at z = 1.0)

zCOSMOS

Spectroscopic Survey of ~20000 *I*-band selected galaxies

$I_{AB} < 22.5$

Conducted with VLT/VIMOS

Redshifts $0.1 < z < 1.2$

Trained SOMs

SOMS trained with 10 broadband colors (u-band to IRAC/Ch2) for ~60000 galaxies that:

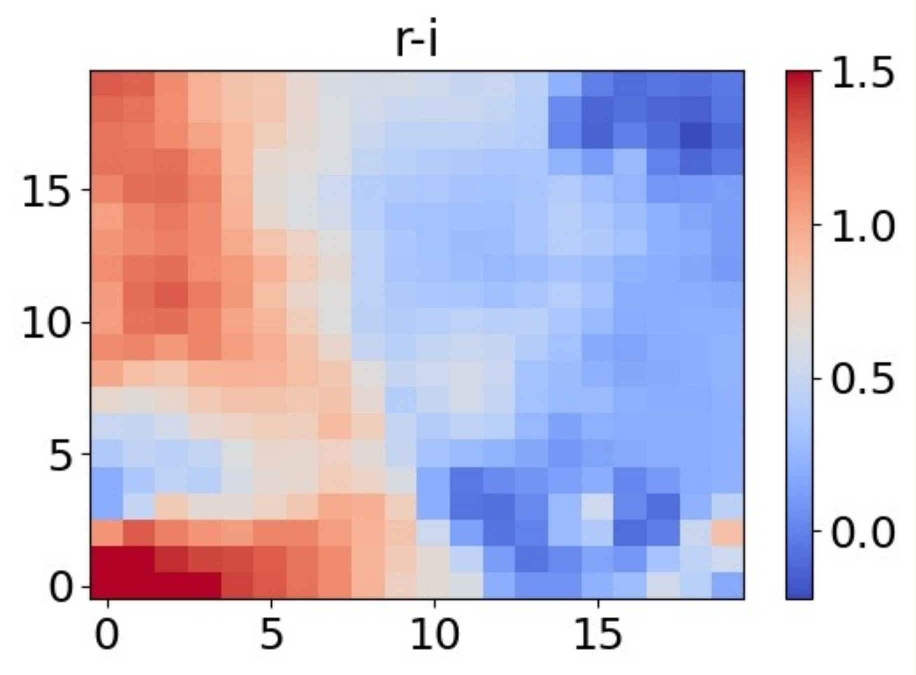
- Meet the criteria of the LEGA-C:
  K-band magnitude < 21.08

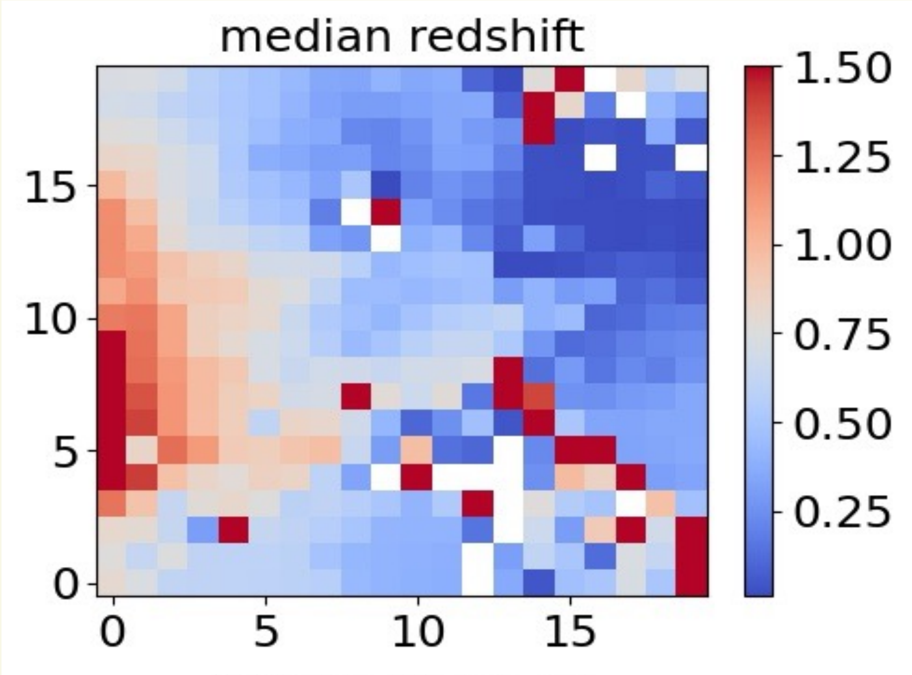
OR

- Meet the criteria of the zCOSMOS:
  I-band magnitude < 22.5

AND

- Have at least 4 measured colors in COSMOS 2020

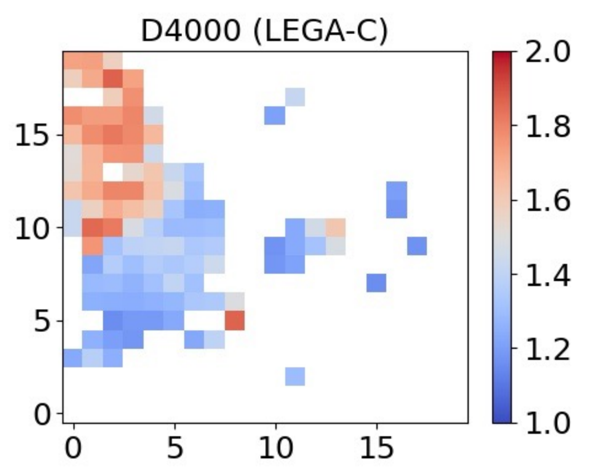


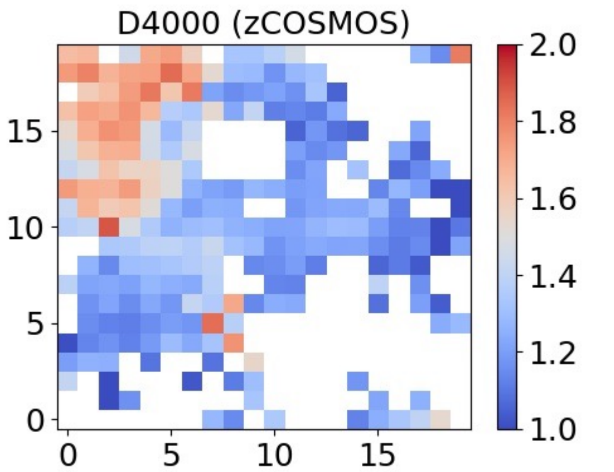


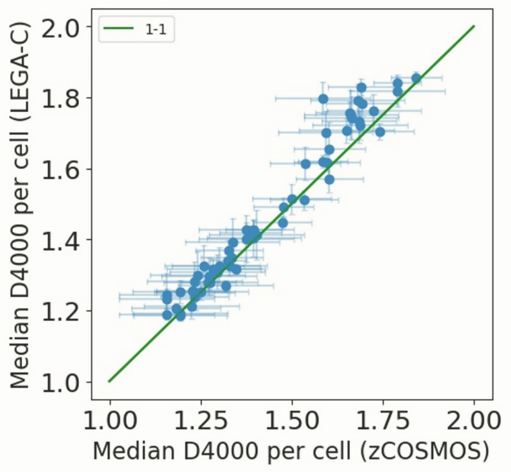
Case Study: strength of the 4000 Å break (D4000)

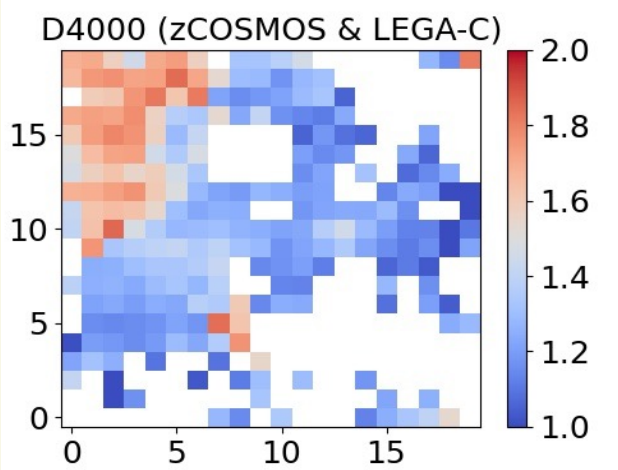
$$D4000_{narrow} = \frac{F\,(4000\text{–}4100\ A)}{F\,(3850\text{–}3950\ A)} \quad \text{(Balogh et al. 1999)}$$

- Strong D4000 is indicative of lack of young stars and abundance of metal absorbents.
- LEGA-C Survey has measured D4000 for 2832 galaxies with mean error of 0.054.
- zCOSMOS has measured D4000 for 11342 galaxies with mean error of 0.078.



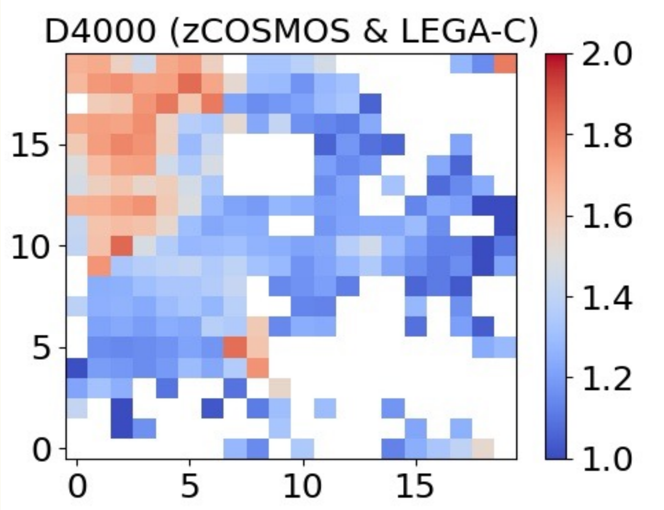


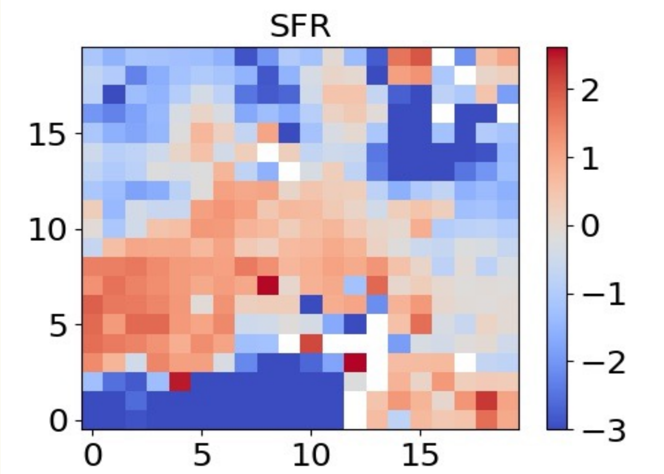


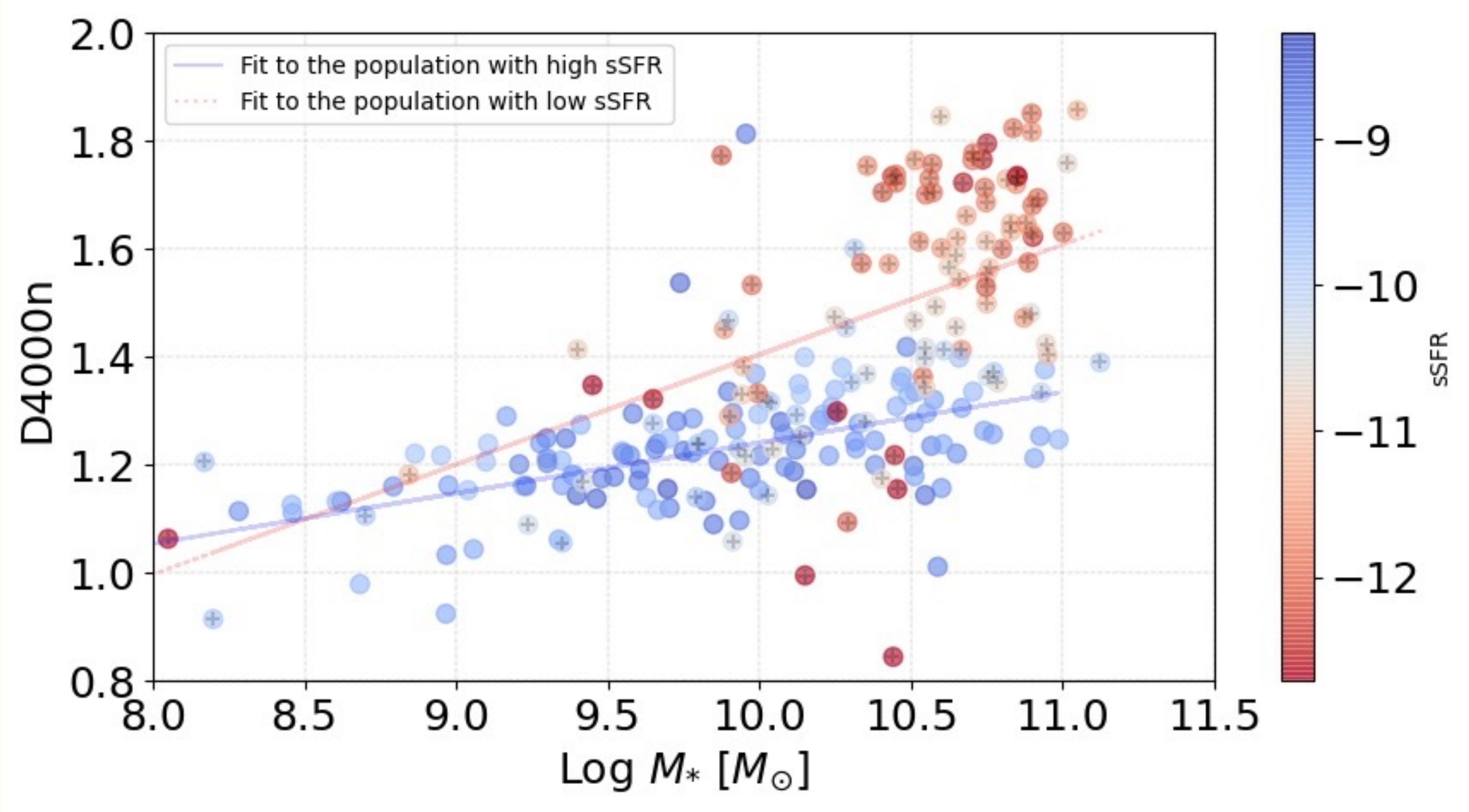


Case Study: strength of the 4000 Å break (D4000)

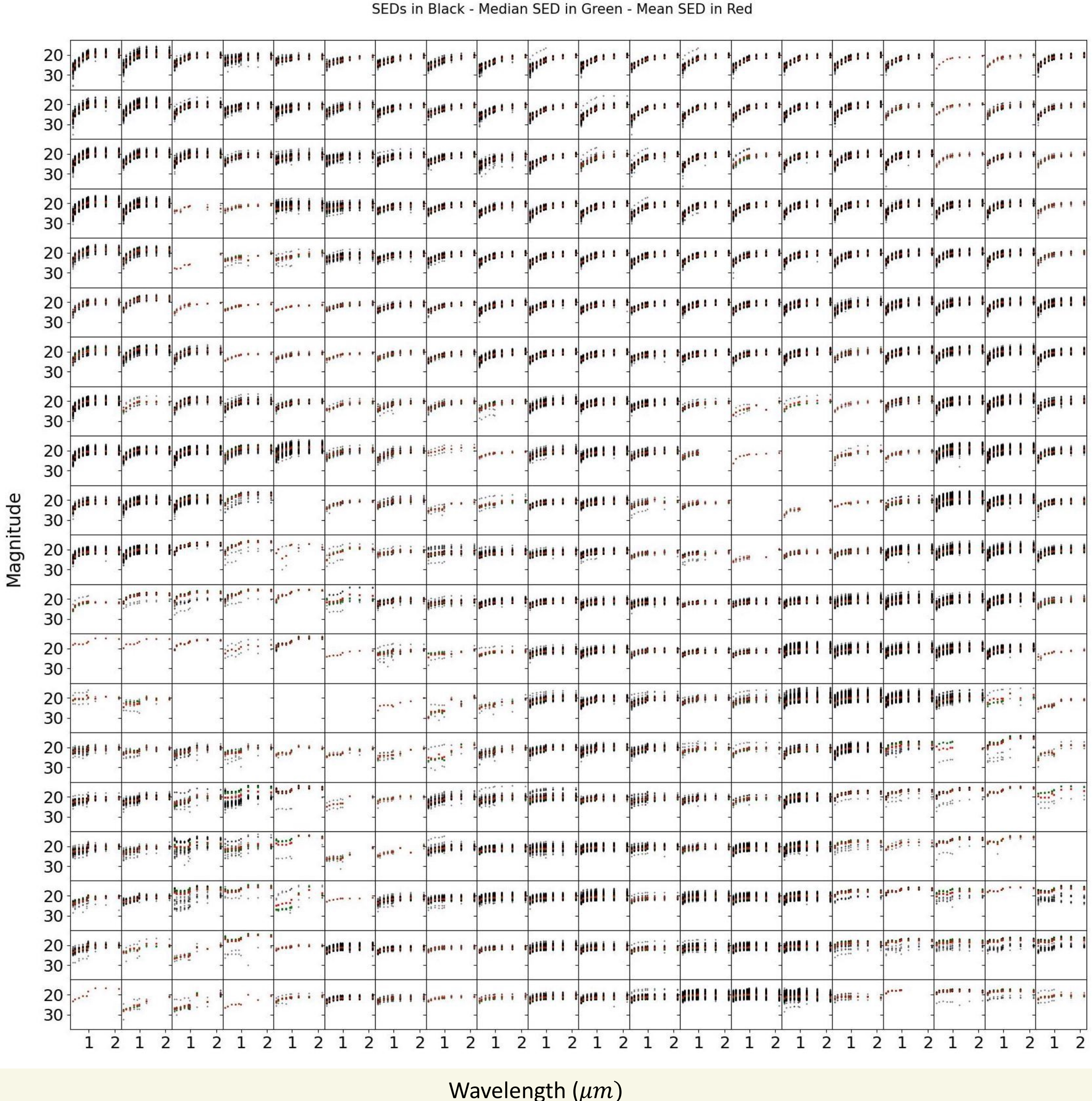
$$D4000_{narrow} = \frac{F\,(4000\text{–}4100\,A)}{F\,(3850\text{–}3950\,A)}$$







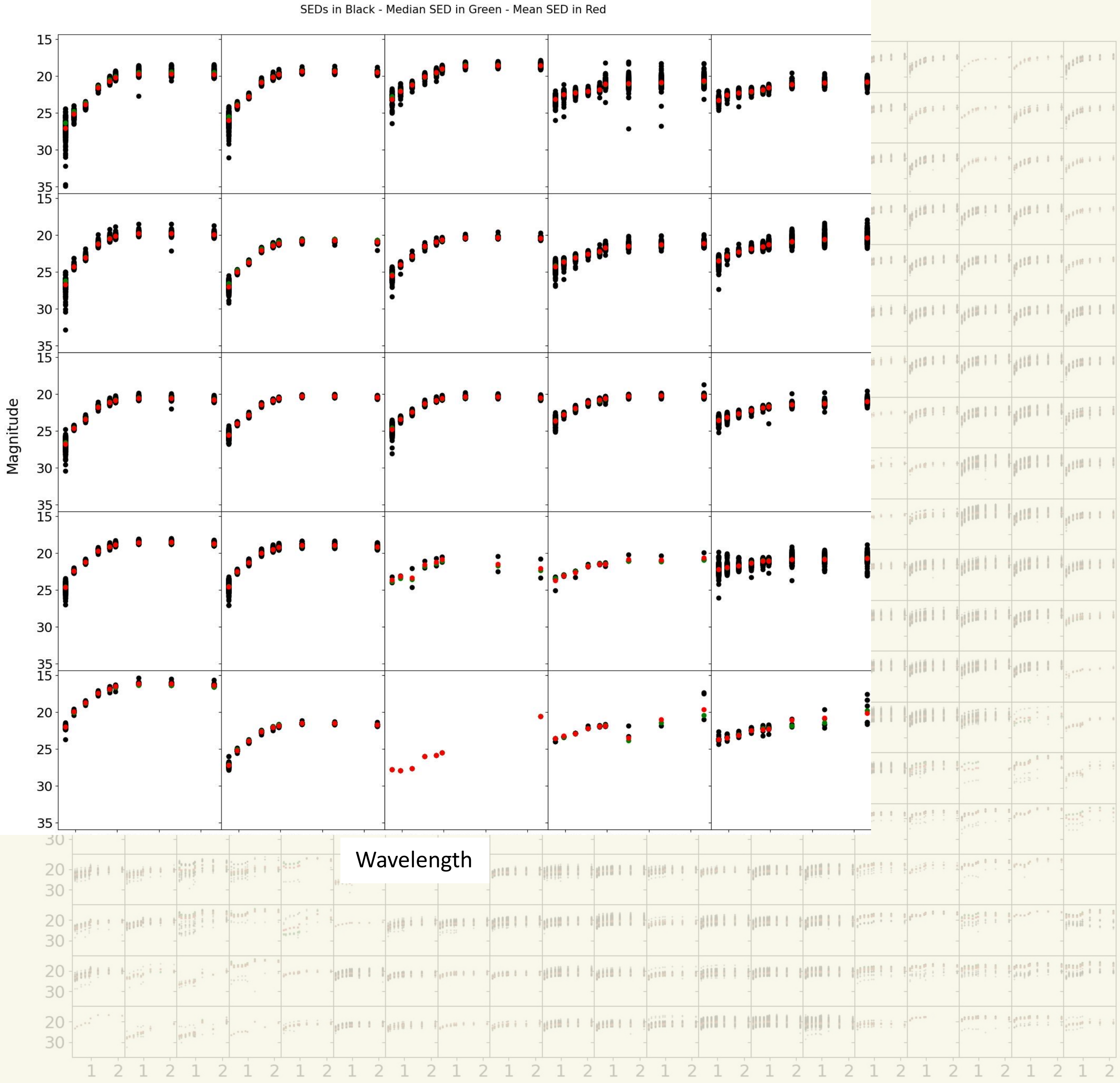
Galaxy
SEDs
In each
SOM
Cell.



Galaxy SEDs

In each

SOM Cell.



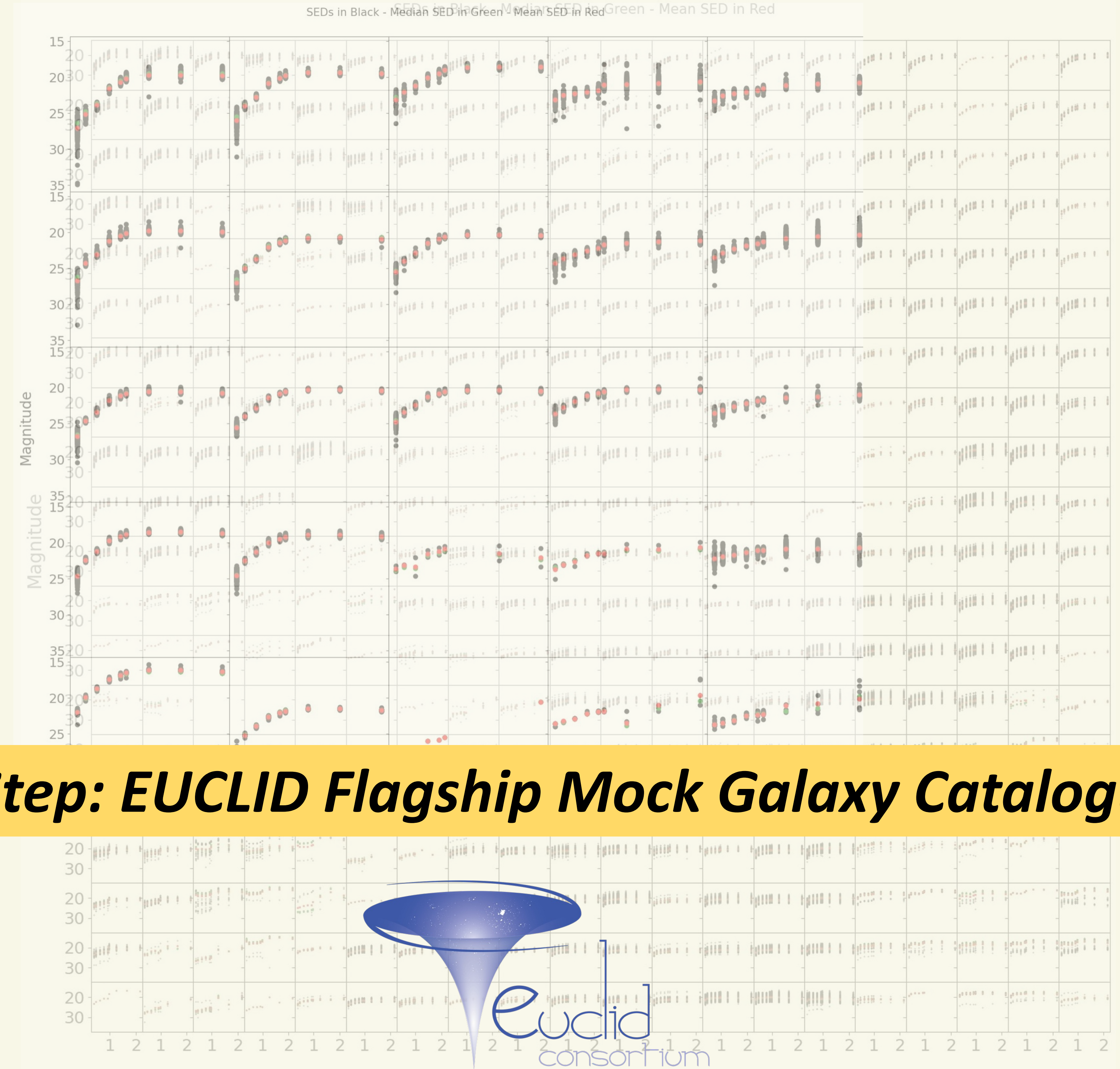
Galaxy

SEDs

In each

SOM

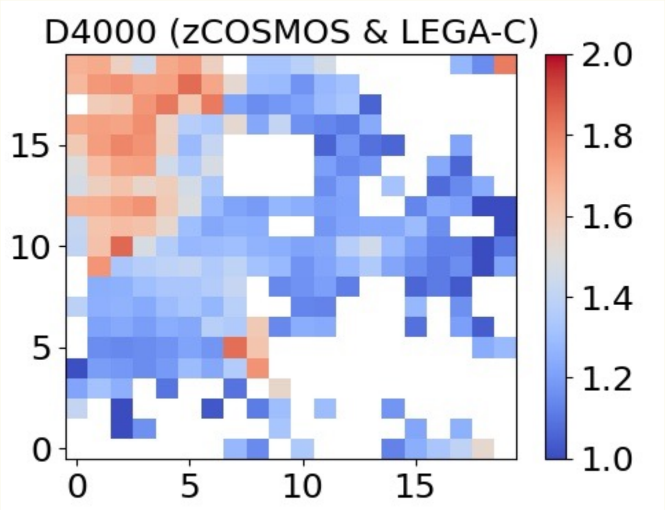
Cell.



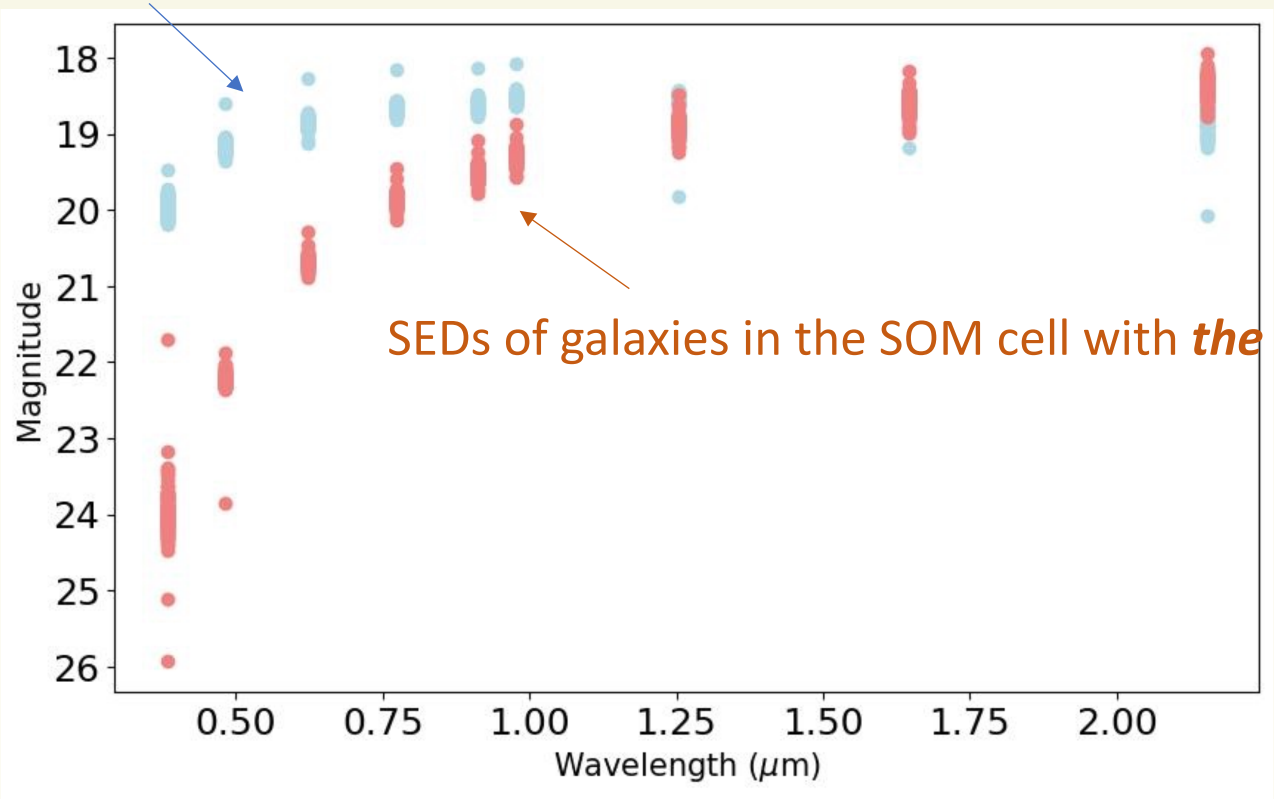
Next Step: EUCLID Flagship Mock Galaxy Catalogue

Case Study: strength of the 4000 Å break (D4000)

$$D4000_{narrow} = \frac{F\,(4000\text{–}4100\ A)}{F\,(3850\text{–}3950\ A)}$$



SEDs of galaxies in the SOM cell with ***the lowest*** D4000



SEDs of galaxies in the SOM cell with ***the highest*** D4000

Case Study: strength of the 4000 Å break (D4000)

Takeaway Point

***S**elf-**O**rganizing **M**aps are powerful tools to predict spectroscopic features of statistically large sample of galaxies for which we have broadband colors, and spectroscopy has been done on a fraction of the sample.*