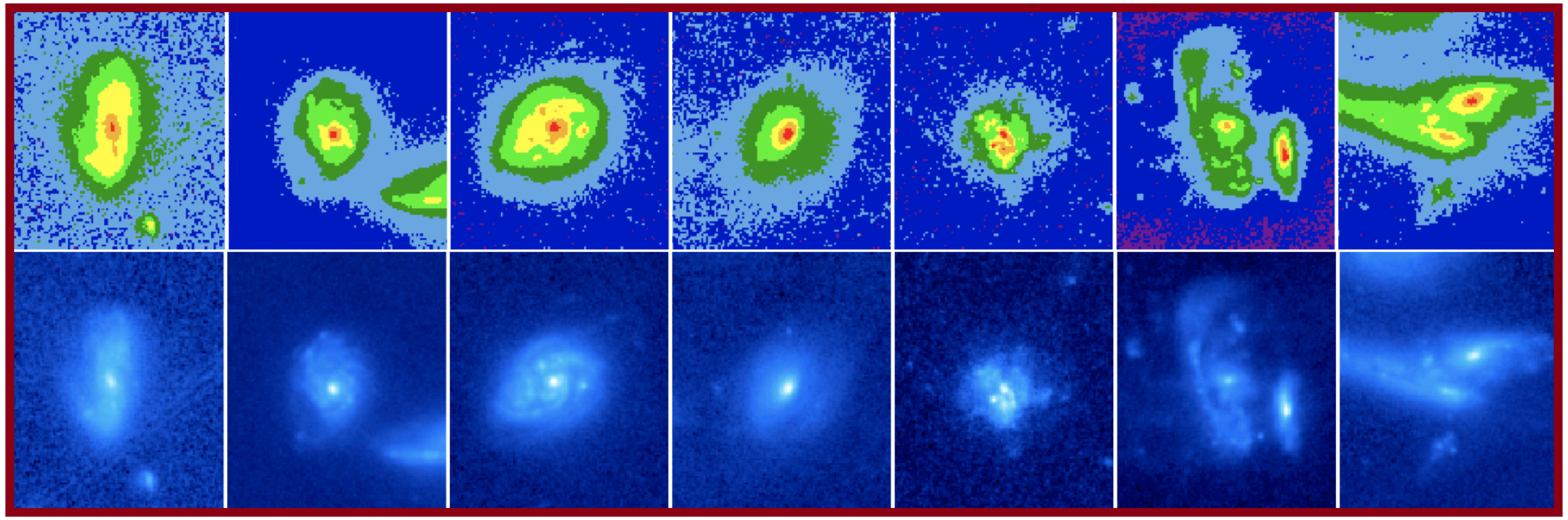
Non-parametric morphological analysis with STATMORPH



Bonaventura et al.

P. Rinaldi, EURECA — Feb 14th, 2025

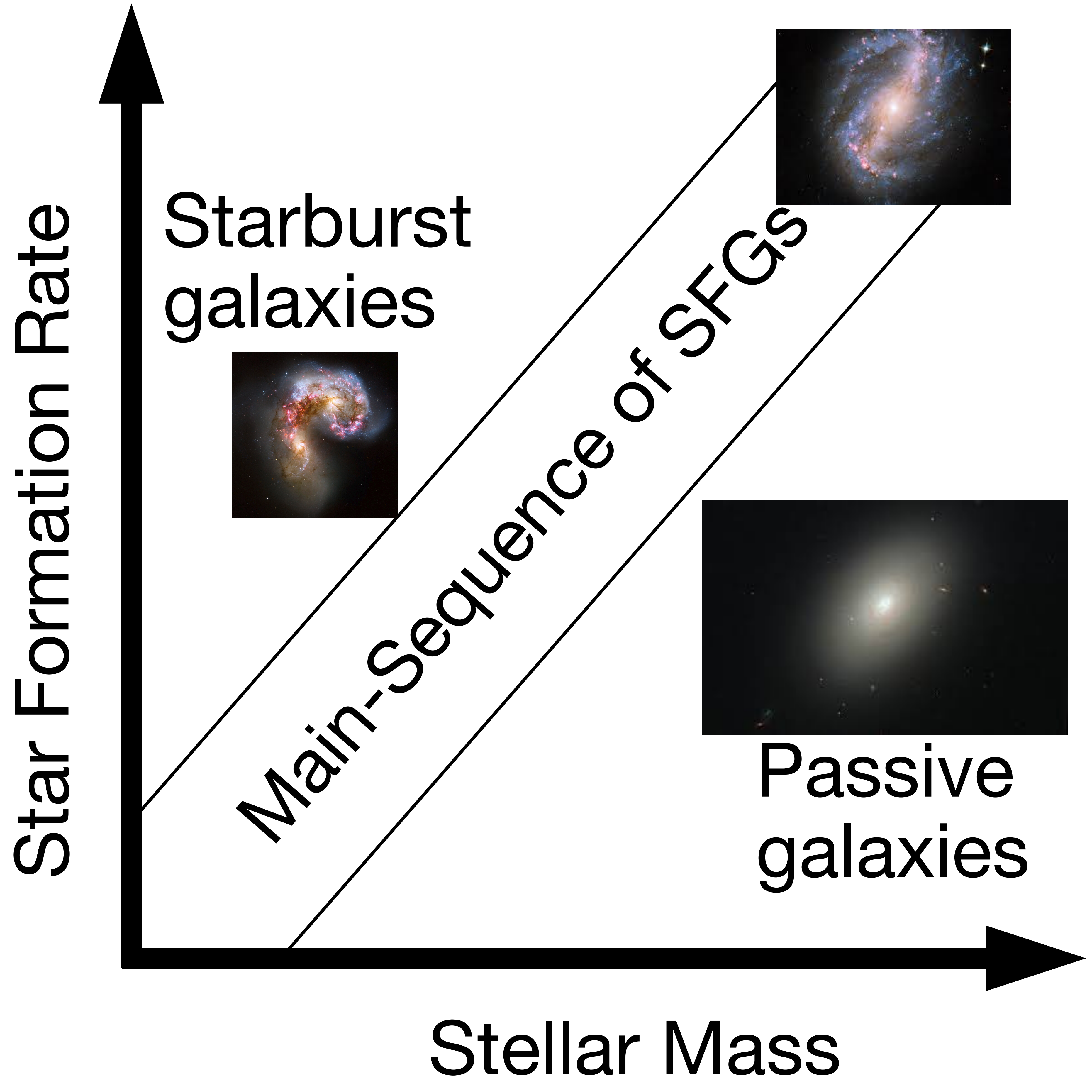
Galaxy Morphology: why is it so important?

- Morphology is the foundation of standard classification schemes. It is the basis of Hubble’s tuning fork and modern classification systems.
- It reveals how galaxies formed and evolved over cosmic time, thus it is a powerful tracer of galaxy transformation as a function of redshift.
- It provides a visual fingerprint of a galaxy’s structure and dynamics.
- It is also strongly connected to key physical properties (e.g.):
  - Colors
  - Stellar mass ($M_\star$)
  - Star Formation History (SFH)
  - Local environment (e.g., you can think about galaxies within galaxy clusters)
- It is fundametnal for studying **secular evolution, mergers, and interactions** and thus for improving our **theories of hierarchical structure formation**.

Galaxy M... important?

- Morpholo... hemes. It is the basis of H... stems.
- It reveals ... ne, thus it is a powerful ... dshift.
- It provide... ynamics.
- It is also s... .):
  - Colors
  - Stellar m...
  - Star For...
  - Local en... in galaxy clusters)
- It is funda... **and interactions** and thus ... **ure formation.**

An easy way to think about it:
How do galaxies look like when looking at them on a SFR-M* plane?



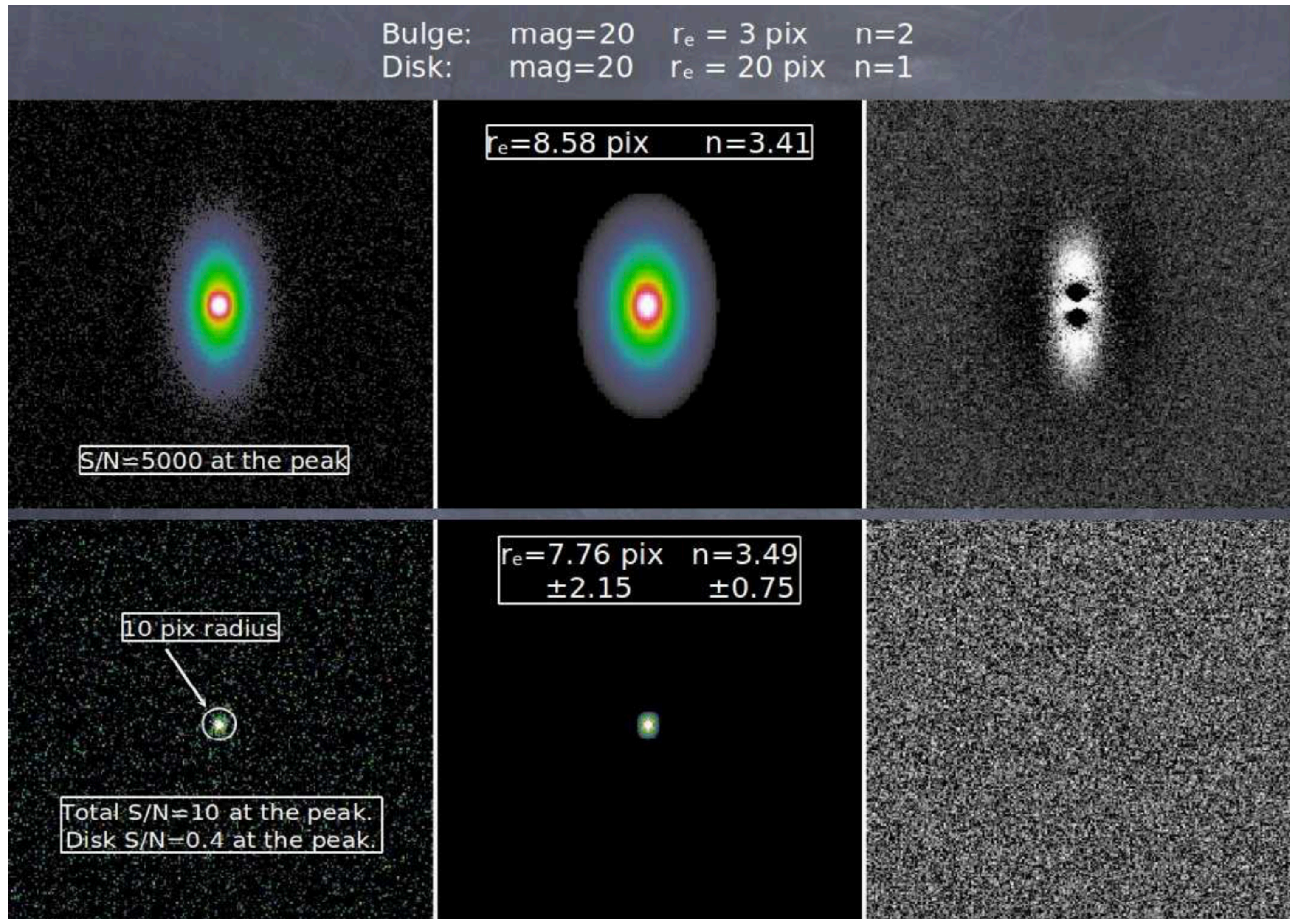
Beyond Parametric Modeling: A Non-Parametric Perspective

Quantifying galaxy morphology is typically done through morphological indicators derived from analytical models fitted to the surface brightness distribution. Common approaches include:

- Single-component fits: Effective radii and Sérsic parameters.
- Multi-component decompositions: Bulge-to-disk ratios.

The common tool used to perform this kind of analysis is **GALFIT**.

CAVEAT: it requires a lot of time because you need to fit each individual object — imagine big surveys like Euclid…



GALFIT example output

Although this “classic” approach has led to numerous studies and results over time, it relies on fitting galaxy images with **predefined models**, such as Sérsic profiles, de Vaucouleurs profiles, or multi-component models. In contrast, non-parametric methods provide a model-independent way to quantify galaxy structure, making them particularly useful for analyzing irregular or disturbed morphologies.

The most popular sets of non-parametric indicators

- Concentration–Asymmetry–Smoothness (CAS)
- Gini–M20 indices
- MID statistics

Why the non-parametric apporach?

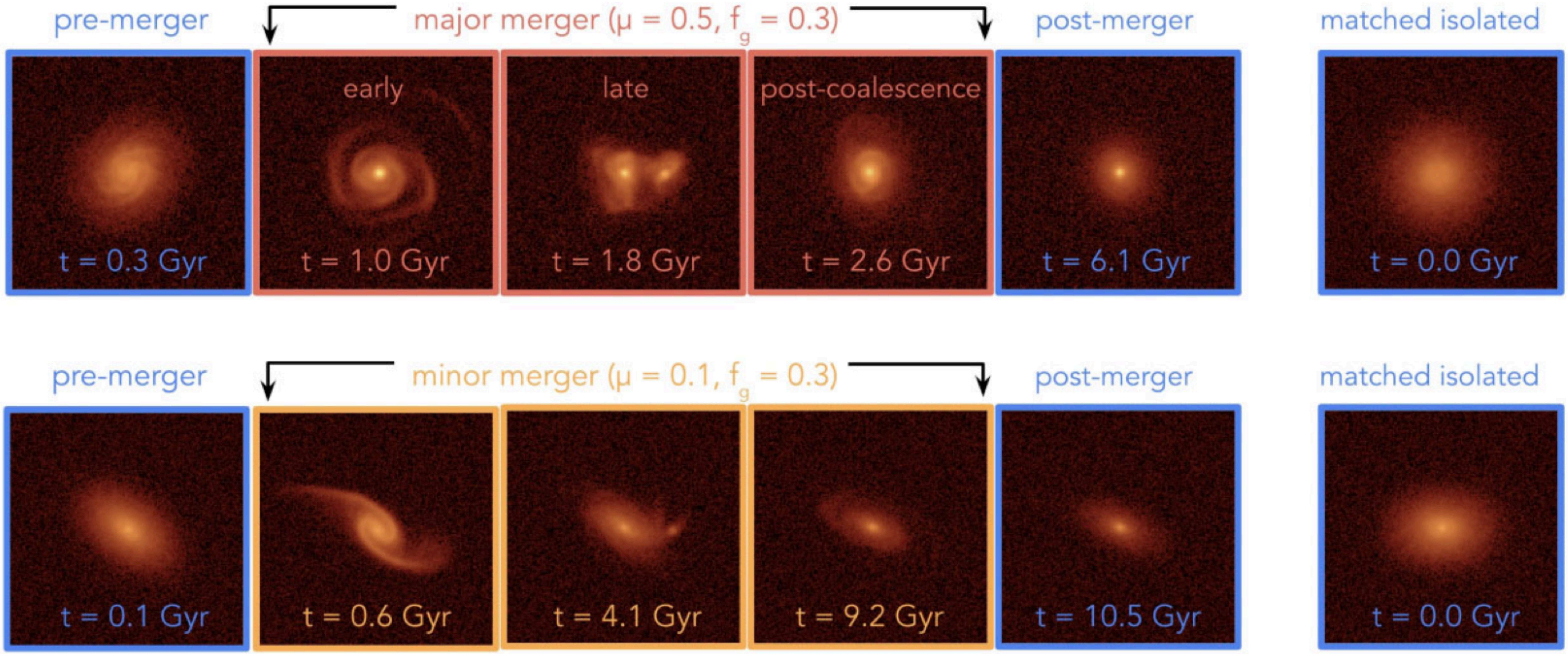
- **Broad Applicability:** We can adopt it to any image, including those not dominated by stellar emission, enabling objective and quantitative comparisons of different galaxy components.

- **Ideal for Large Galaxy Surveys:** Efficient and scalable, making them suitable for analyzing large datasets (e.g., Euclid, Roman, etc.). Unlike parametric models, non-parametric metrics (e.g., Gini, M20, concentration, asymmetry) can be computed quickly and fully automated. They are also more robust for irregular and merging galaxies that do not conform to simple models.

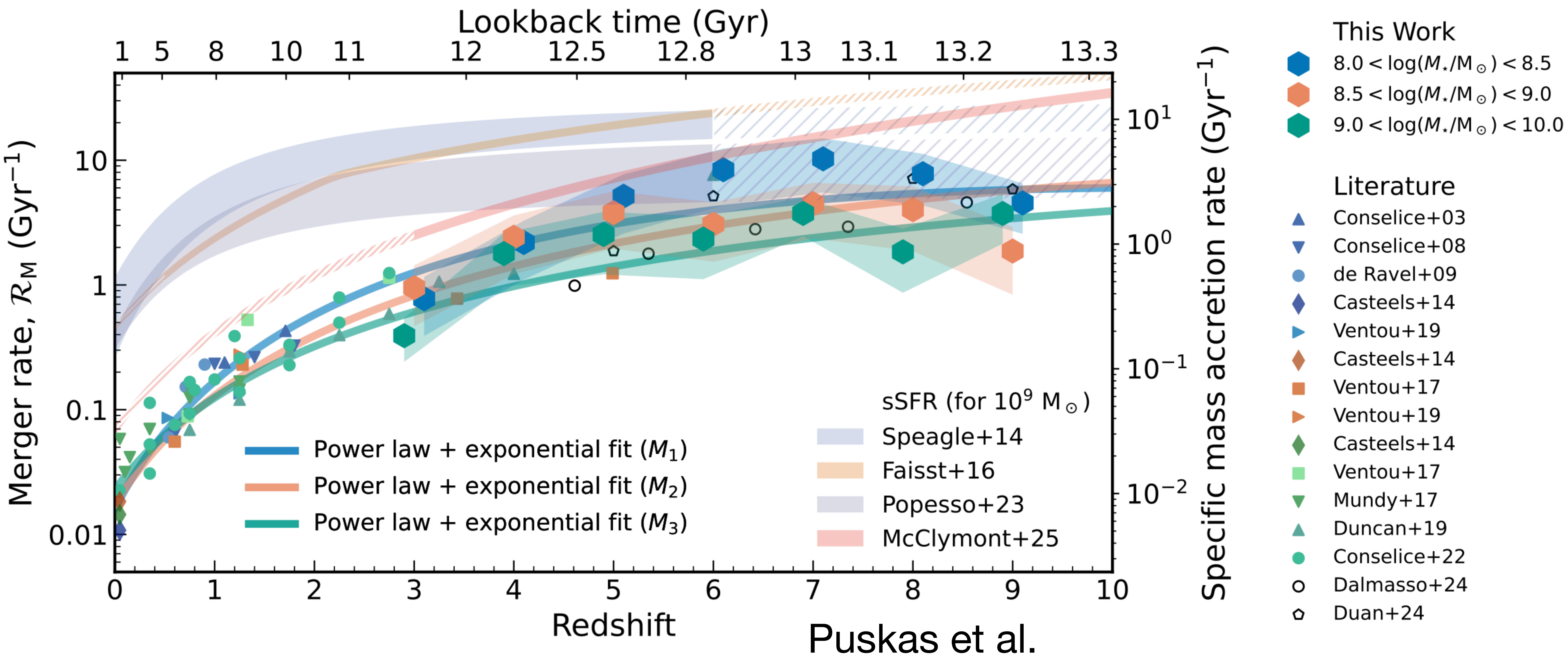
- **Balancing Non-Parametric and Parametric Approaches:** While parametric models remain useful for detailed structural analysis (e.g., bulge-to-disk ratios), a hybrid approach is often used—applying parametric fits to bright, well-resolved galaxies while using non-parametric methods for the broader population.

Why the non-parametric apporach?

It is the most suitable apporach to identify galaxy mergers



If we study mergers, we can trace the stellar mass assembly…

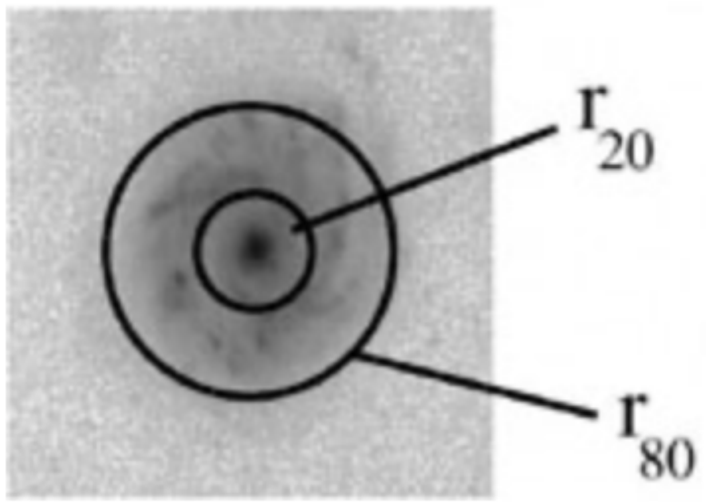


Puskas et al.

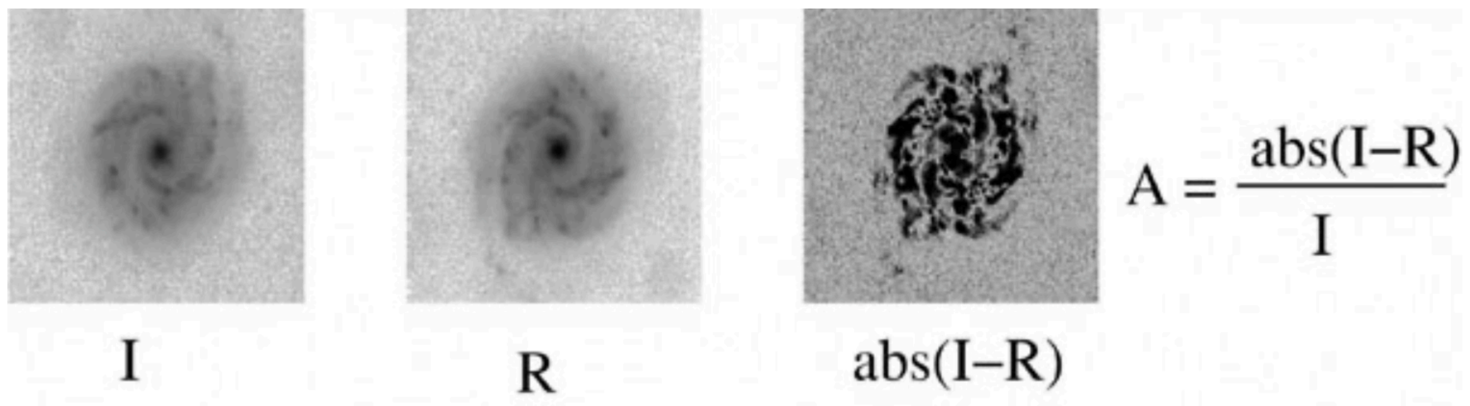
I know, it’s boring…, but useful

Concentration–Asymmetry–Smoothness (CAS)

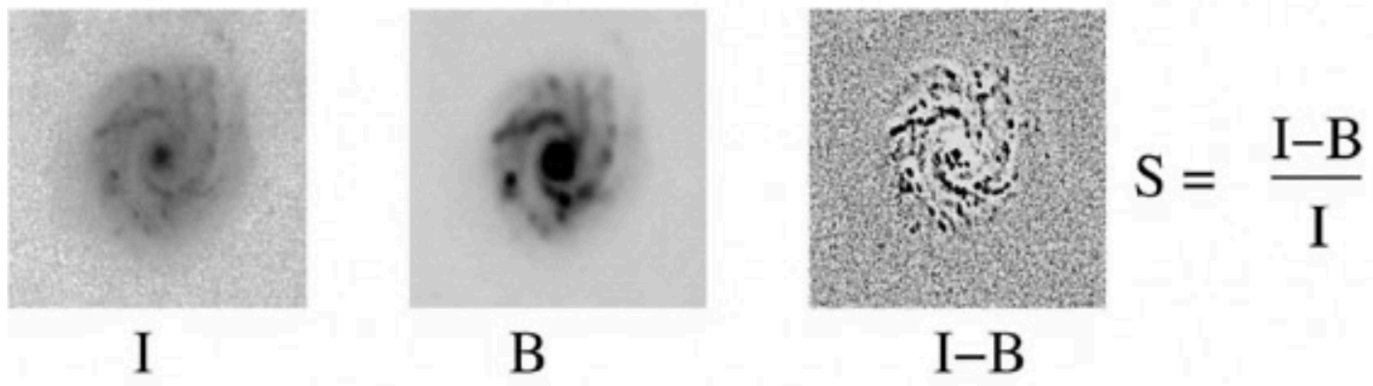
- **Concentration:** This parameter is a measure of how concentrated (or diffuse) the central bulge component is with respect to the total flux of the galaxy. The concentration of a galaxy is determined by locating the radii of the circular apertures that contain 20% ($r_{0.2}$) and 80% ($r_{0.8}$) of the total flux of the galaxy, and then considering $C = 5\log(r_{0.8}/r_{0.2})$. [**it correlates strongly with Sérsic n values**]



- **Asymmetry:** It quantifies how symmetric or asymmetric a galaxy is relative to its total flux. Galactic asymmetries primarily arise from star formation, particularly in structural features such as spiral arms, and serve as strong indicators of merger history. The asymmetry itself is calculated by rotating the image 180° around the defined center and subtracting the rotated image from the original. The absolute residuals are then normalized by the total flux of the galaxy. To account for background noise, the same method is applied to a background-only region of the image, with the resulting value scaled to match the galaxy’s size. This background asymmetry is then subtracted from the measured galactic asymmetry to obtain a final, corrected value. Then $A = (\sum I - I_{180}/\sum I) - (\sum B - B_{180}/\sum I)$, where I is the Image flux and B is the background flux.



- **Clumpiness Parameter or Smoothness (S):** It measures the high-frequency light distribution in a galaxy, directly linked to star formation. It is computed by subtracting a smoothed, degraded-resolution version of the image from the original. Background clumpiness is subtracted for correction. To avoid bias from central bulges, the flux contribution from this region is excluded based on the galaxy’s light profile. Then $S = 10 \times (\sum I - I_S/\sum I) - (\sum B - B_S/\sum I)$, I is the image flux, Is is the smoothed image flux, B is the background, and Bs is the smoothed background.



A real application for the CAS statistics

Tohill et al.

GINI-M20 indices

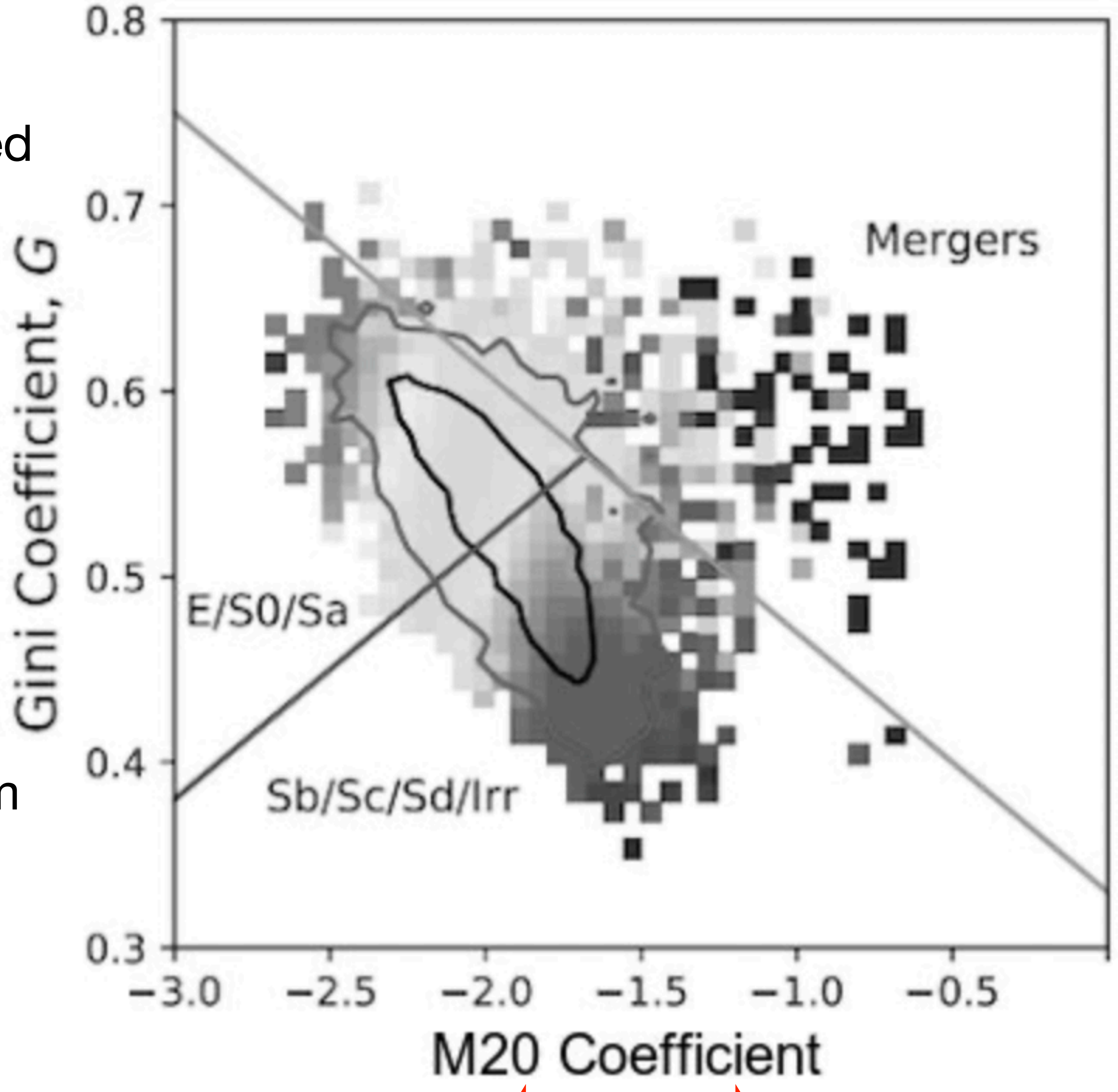
- **Gini parameter:** As applied to galaxy morphologies, it is a measurement of the inequality of the distribution of light in a galaxy. Conceptually, this measurement is made by ordering the pixels that make up the image of a galaxy in ascending order by flux and then comparing the resulting cumulative distribution function to what would be expected from a perfectly even flux distribution. A low Gini coefficient (close to zero) means that the distribution of light is fairly uniform, while a high Gini coefficient (close to one) means that most of the galaxy's light is contained in only a small fraction of the pixels. Then, $G = \frac{1}{|\bar{X}|n(n-1)} \sum_{i=1}^{n} (2i - n - 1)|X_i|$, where X_i is the value of the ith pixel when ordered by flux, $\bar{X}$ is the mean of the pixel values, and n is the total number of pixels.

- M_{20}: It measures the spatial concentration of a galaxy's brightest regions by computing the second-order moment of the flux within its brightest 20%. It traces structures such as nuclei, bars, spiral arms, and off-center star clusters. A lower M20 value indicates a centrally concentrated brightness distribution, while higher values correspond to more extended or irregular light distributions.Then,

$$M_{20} \equiv \log_{10}\left(\frac{\sum_i M_i}{M_{tot}}\right), \quad \text{while} \sum_i f_i < 0.2 f_{tot}.$$

More concentrated light distribution

More uniform light distribution



More concentrated, e.g. elliptical

Multiple bright regions, e.g. Irregular

Other (alternative) non-parametric indicators for identify mergers

MID and A_s

- **Multimode (M) Statistic:** The M statistic quantifies how a galaxy's light is distributed into distinct regions. It identifies separate pixel groups above a given intensity threshold (q), ranks them by size, and computes the ratio of the second-largest group's area ($A_{q,2}$) to the first ($A_{q,1}$). A high M value (≈ 1) suggests a double-nucleus system, while a low value (≈ 0) indicates a single, centrally concentrated source. This makes M particularly useful for identifying mergers and galaxies with multiple bright components. Then, $M = \max_q(A_{q,2}/A_{q,1})$.

- **Intensity (I) Statistic:** The I statistic measures the brightness contrast between the two most luminous regions in a galaxy. This provides a complementary measure to M, as I identifies compact bright regions that may be too small to significantly affect the M statistic. It is particularly useful for detecting dual AGN or bright star-forming regions.

- **Deviation (D) Statistic:** The D statistic measures how offset the brightest region of a galaxy is from its overall centroid. It is calculated as the normalized distance between the galaxy's flux-weighted center (as determined from the segmentation map) and the brightest intensity peak found via the I statistic. High D values indicate morphological disturbances, such as late-stage mergers or galaxies undergoing strong interactions, while low D values suggest a symmetric, ordered structure like a disk or elliptical galaxy.

- **Shape Asymmetry (A_S) Statistic:** This is similar to the classic A but is computed on a binary detection mask instead of the flux image. This means it measures the spatial outline of a galaxy rather than variations in pixel brightness. Because it assigns equal weight to all detected structures, A_S is particularly sensitive to low-surface-brightness features along the galaxy's edges. It is computed by rotating the galaxy's mask 180° and subtracting it from the original, with background asymmetry subtracted for correction. This makes A_S a useful tool for detecting faint tidal features or extended low-brightness structures that may not significantly affect the traditional asymmetry parameter. Usually, we classify all objects with $A_S > 0.2$ as strongly spatially disturbed.

What is the difference among all of them?

- **CAS:** Quantifies morphology based on light concentration, asymmetry, and small-scale structure. It is intuitive—does the galaxy look compact, asymmetric, or clumpy? Best for distinguishing bulge-dominated vs. irregular galaxies.

- **Gini-M20:** Measures how flux is distributed, distinguishing bulge-dominated, disk, and merging systems. It is statistical—how is brightness distributed? Best for separating mergers from disks and bulge-dominated galaxies.

- **MID:** Detects mergers and irregular galaxies by tracking multiple bright regions, intensity peaks, and deviations from symmetry. It is merger-sensitive—are there multiple bright zones and irregularities? Best for detecting ongoing mergers and disturbed morphologies.

Does it exist a tool that computes all these parameters for us?

Yes, it's STATMORPH

Statmorph is written in Python and is straightforward to use. A detailed description of its input parameters and measurements can be found in Rodriguez-Gomez et al. (2019).

The code requires three primary input files:

- Science image.
- Segmentation map.
- Weight map (representing the standard deviation, i.e. σ, of each pixel value).

CAVEAT: The code does not provide error estimates for its outputs. A common workaround is to generate mock images by perturbing each pixel based on its associated σ. By repeating this process multiple times (e.g., 1000), the 16th and 84th percentiles of the resulting measurement distributions can be used as error estimates. Basically, a Monte Carlo approach.

Does it exist a tool that computes all these parameters for us?

Yes, it’s STATMORPH

Optionally, you can also provide:

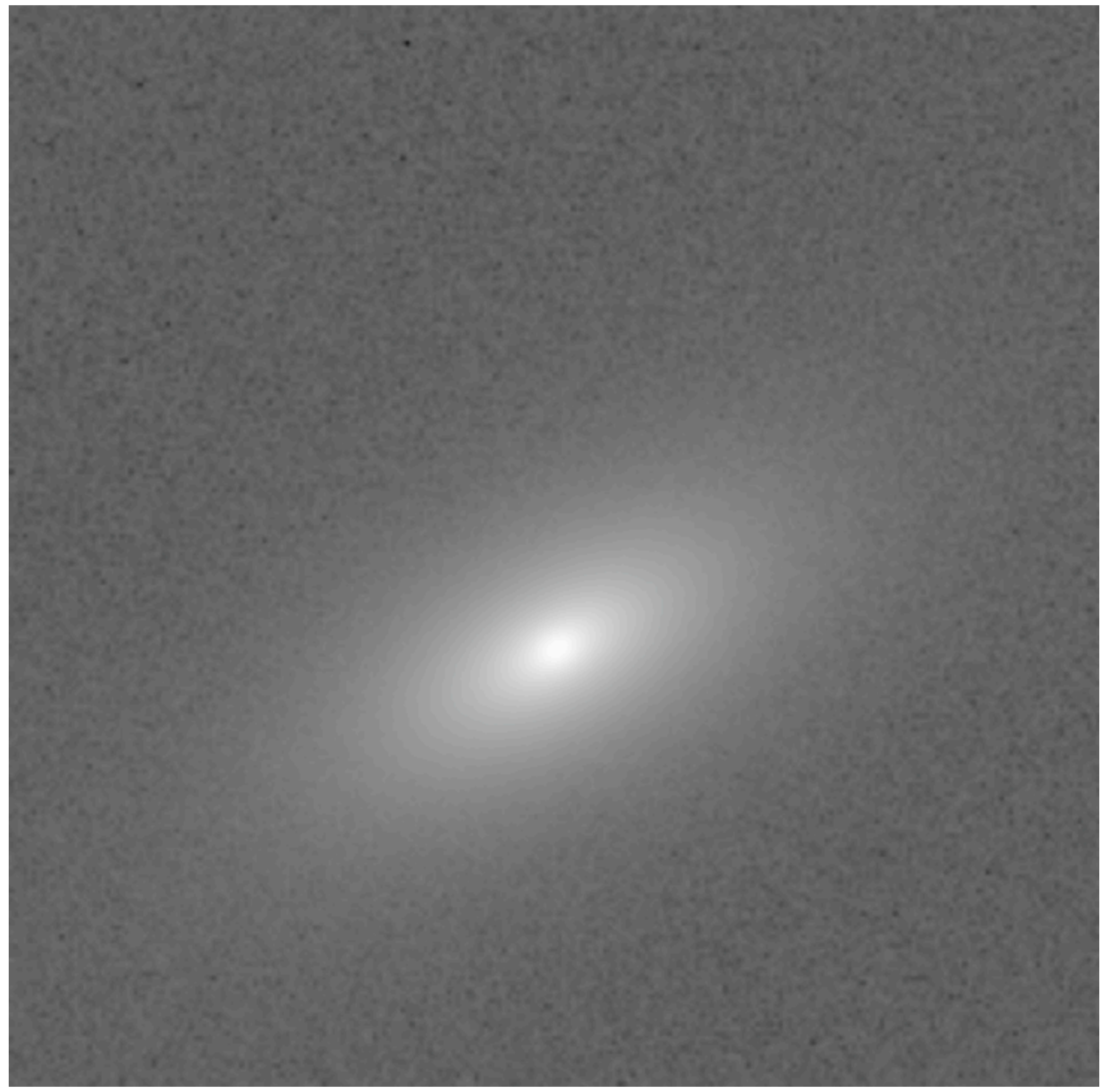
- Mask: this allows you to mask out sources for which you do not want to make any estimate.
- PSF: this is used when fitting 2D Sersic profiles. (this means that statmorph behaves like GALFIT somehow; so it is versatile in this sense)

What do we get as an output from STATMORPH?

- Morphological parameters (no errors!)
- FLAG: this can 0, 1, or 2. 0 indicates that there were no problem with your measurements. 1 means that you have a GINI segmap which is discontinuous (e.g., due to a secondary source that was not properly labeled/masked). 2 indicates that there were problems with the measurements.
- FLAG_SERSIC: as before, but it refers to your Sersic profile fit.

Basic example with STATMORPH

```
import numpy as np
import matplotlib.pyplot as plt
from astropy.visualization import simple_norm
from astropy.modeling.models import Sersic2D
from astropy.convolution import convolve, Gaussian2DKernel
from photutils.segmentation import detect_threshold, detect_sources
import time
import statmorph
%matplotlib inline
```

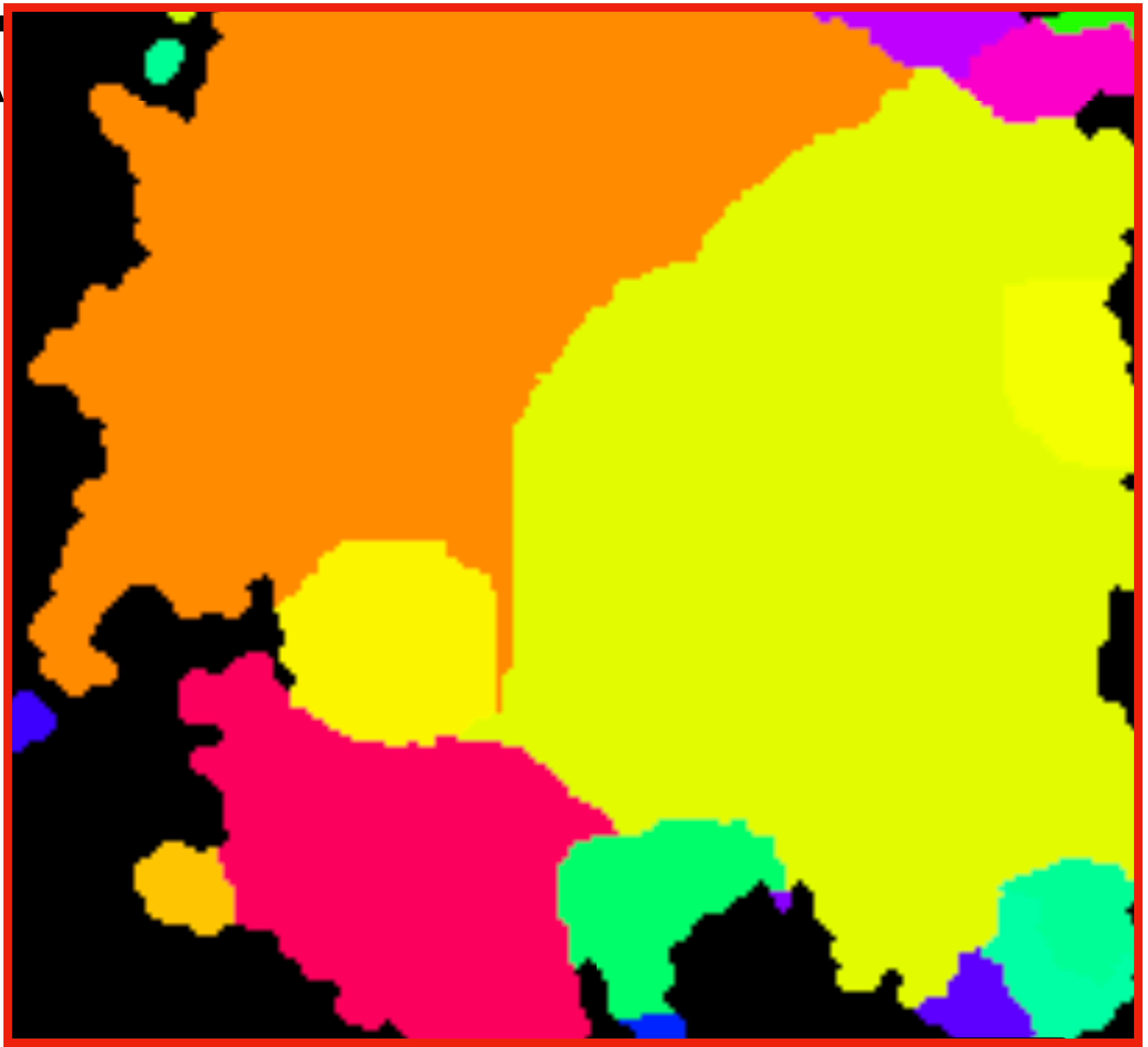


A mock galaxy created with astropy

Its segmentation map

STA

t_sources



REALITY IS NOT THAT EASY AND THINGS ARE MUCH MORE COMPLICATED

Basic example with STATMORPH

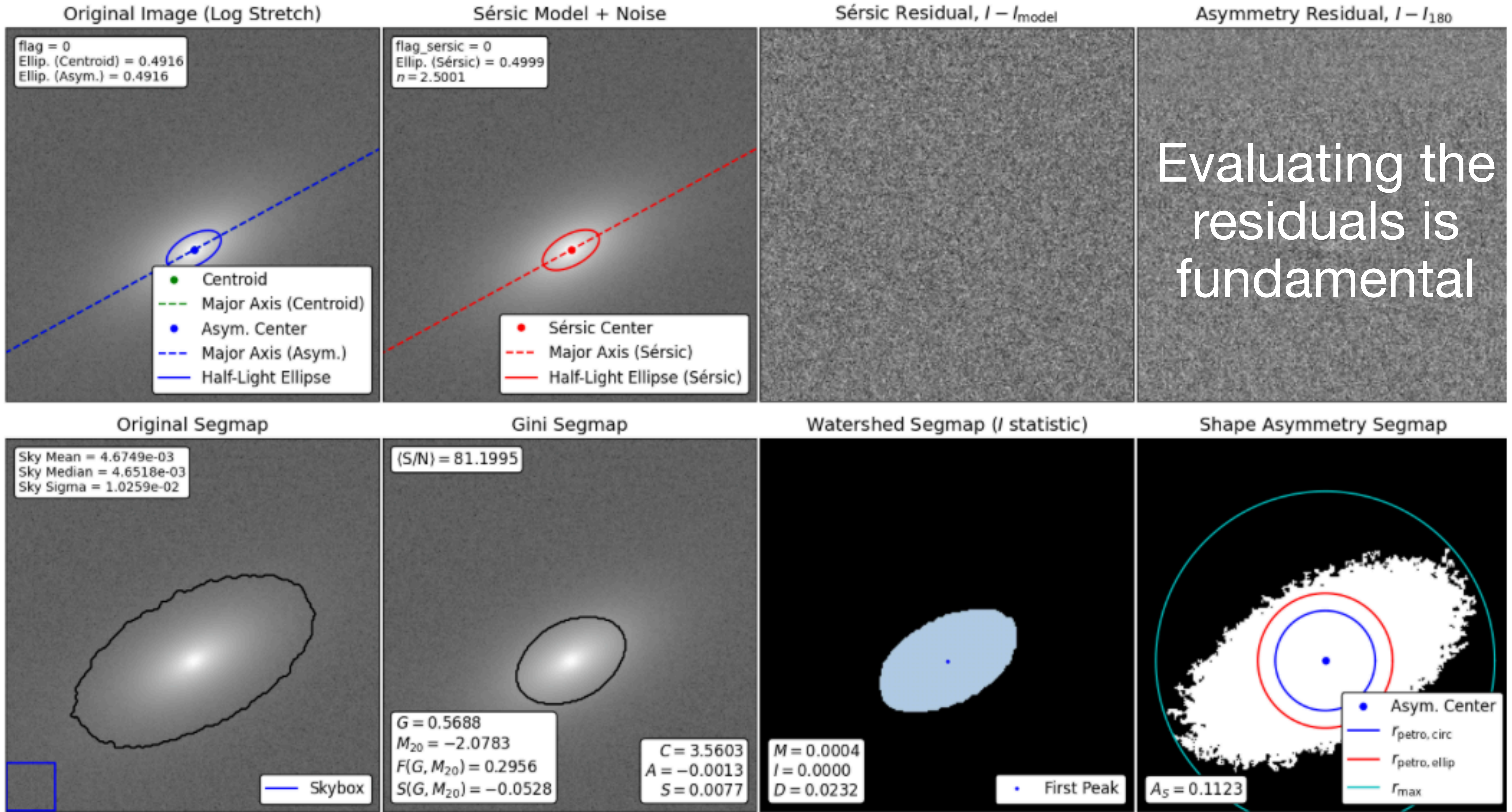
```
start = time.time()
source_morphs = statmorph.source_morphology(
    image, segmap, gain=gain, psf=psf)
print('Time: %g s.' % (time.time() - start))
```

GAIN is your weightmap

PSF is not necessary, but useful

…and that's it!

```
print('BASIC MEASUREMENTS (NON-PARAMETRIC)')
print('xc_centroid =', morph.xc_centroid)
print('yc_centroid =', morph.yc_centroid)
print('ellipticity_centroid =', morph.ellipticity_centroid)
print('elongation_centroid =', morph.elongation_centroid)
print('orientation_centroid =', morph.orientation_centroid)
print('xc_asymmetry =', morph.xc_asymmetry)
print('yc_asymmetry =', morph.yc_asymmetry)
print('ellipticity_asymmetry =', morph.ellipticity_asymmetry)
print('elongation_asymmetry =', morph.elongation_asymmetry)
print('orientation_asymmetry =', morph.orientation_asymmetry)
print('rpetro_circ =', morph.rpetro_circ)
print('rpetro_ellip =', morph.rpetro_ellip)
print('rhalf_circ =', morph.rhalf_circ)
print('rhalf_ellip =', morph.rhalf_ellip)
print('r20 =', morph.r20)
print('r80 =', morph.r80)
print('Gini =', morph.gini)
print('M20 =', morph.m20)
print('F(G, M20) =', morph.gini_m20_bulge)
print('S(G, M20) =', morph.gini_m20_merger)
print('sn_per_pixel =', morph.sn_per_pixel)
print('C =', morph.concentration)
print('A =', morph.asymmetry)
print('S =', morph.smoothness)
print()
print('SERSIC MODEL')
print('sersic_amplitude =', morph.sersic_amplitude)
print('sersic_rhalf =', morph.sersic_rhalf)
print('sersic_n =', morph.sersic_n)
print('sersic_xc =', morph.sersic_xc)
print('sersic_yc =', morph.sersic_yc)
print('sersic_ellip =', morph.sersic_ellip)
print('sersic_theta =', morph.sersic_theta)
print('sersic_chi2_dof =', morph.sersic_chi2_dof)
print()
print('OTHER')
print('sky_mean =', morph.sky_mean)
print('sky_median =', morph.sky_median)
print('sky_sigma =', morph.sky_sigma)
print('flag =', morph.flag)
print('flag_sersic =', morph.flag_sersic)
```



Basic example with STATMORPH

```
start = time.time()
source_morphs = statmorph.source_morphology(
    image, segmap, gain=gain, psf=psf)
print('Time: %g s.' % (time.time() - start))
```

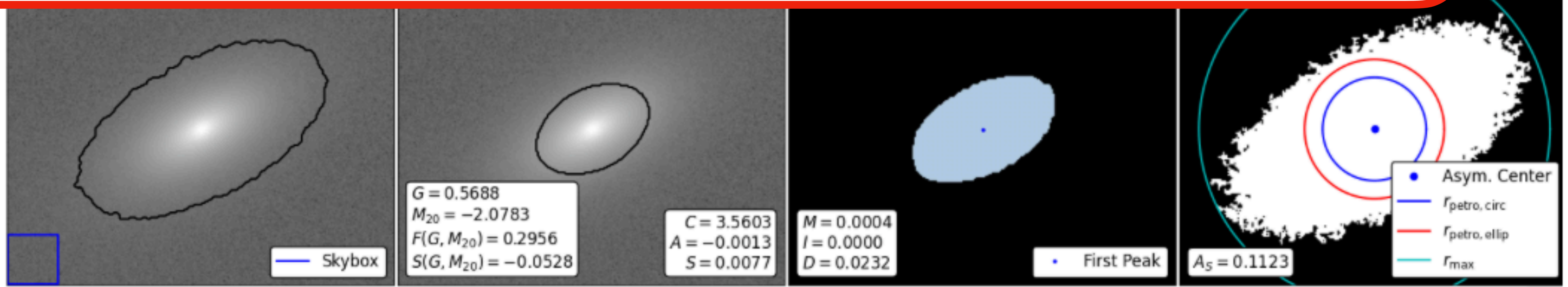
GAIN is your weightmap

PSF is not necessary, but useful

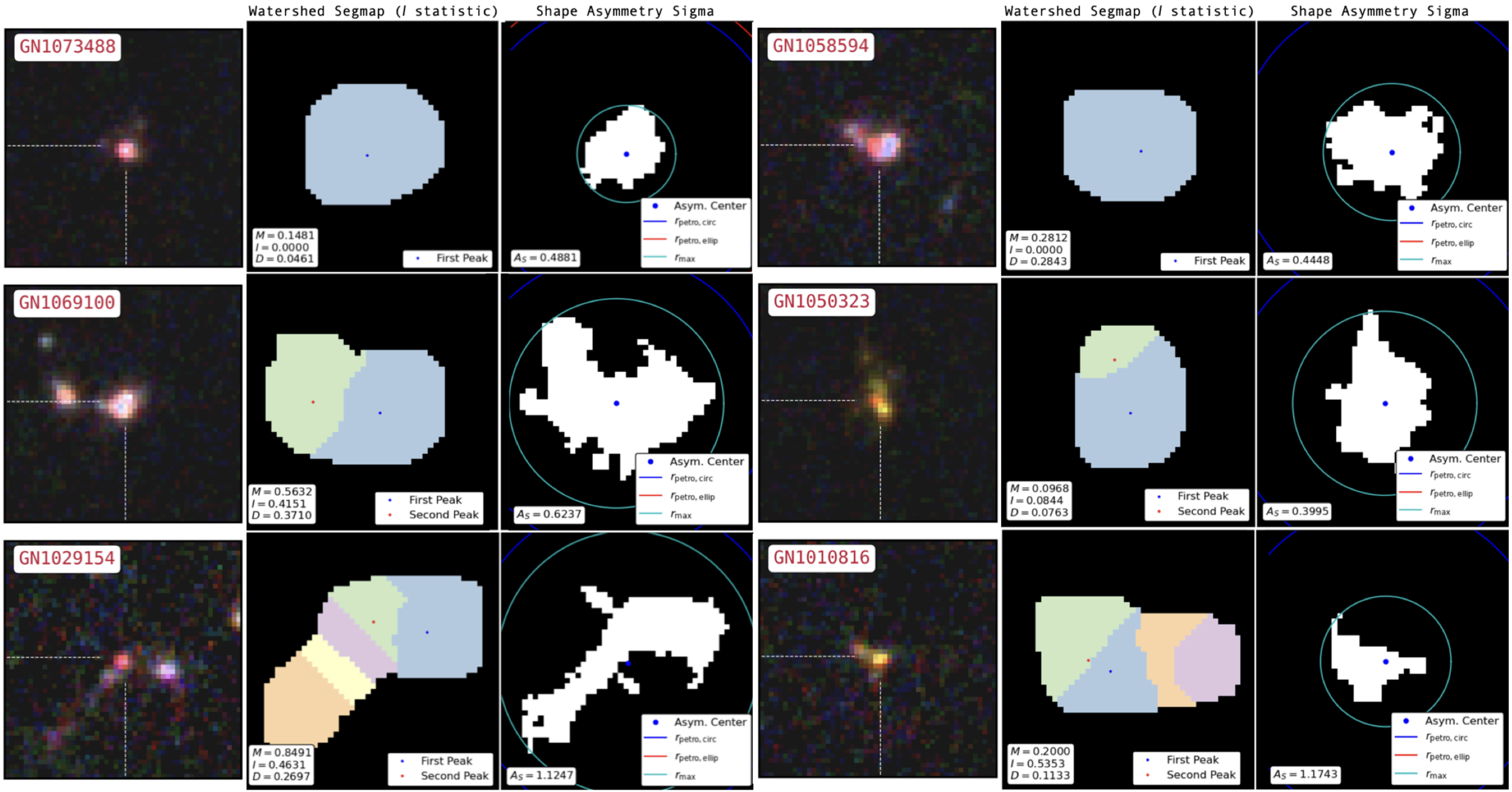
…and that's it!

```
print('BASIC M
print('xc_cent
print('yc_cent
print('ellipti
print('elongat
print('orienta
print('xc_asym
print('yc_asym
print('ellipti
print('elongat
print('orienta
print('rpetro_
print('rpetro_
print('rhalf_c
print('rhalf_e
print('r20 =',
print('r80 =',
print('Gini ='
print('M20 =',
print('F(G, M2
print('S(G, M2
print('sn_per_
print('C =', m
print('A =', m
print('S =', m
print()
print('SERSIC
print('sersic_
print('sersic_
print('sersic_n =
print('sersic_xc =', morph.sersic_xc)
print('sersic_yc =', morph.sersic_yc)
print('sersic_ellip =', morph.sersic_ellip)
print('sersic_theta =', morph.sersic_theta)
print('sersic_chi2_dof =', morph.sersic_chi2_dof)
print()
print('OTHER')
print('sky_mean =', morph.sky_mean)
print('sky_median =', morph.sky_median)
print('sky_sigma =', morph.sky_sigma)
print('flag =', morph.flag)
print('flag_sersic =', morph.flag_sersic)
```

When galaxies are isolated, estimating morphological parameters is relatively easy. However, when things are complicated (e.g., mergers, very crowed fields, etc.), the most difficult thing is to create a realistic segmentaiton map that actually captures and deblend carefully each soruce



A real application (from Rinaldi, Bonaventura et al.)



Useful references

- Abraham et al. 2003 (for the GINI/M20 indices)
- Conselice 2003 (for the CAS statistics)
- Lotz et al. 2004 (for the GINI/M20 indices)
- Freeman et al. 2013 (for the MID statistics)
- Nevin et al. 2019 (useful for connecting galaxy morphology to galaxy formation and evolution)
- Rodriguez-Gomez et al. 2019 (for STATMORPH)
- Nevin et al. 2023 (useful for connecting galaxy morphology to galaxy formation and evolution)

Some applications (very recent literature with JWST data):

- Bonaventura et al. 2024 (looking at the merger fraction in AGNs)
- Rinaldi et al. 2024 (looking at the UV morphology of Little Red Dots)
- Puskas et al. 2025 (looking at the merger history at $z \approx 3 - 9$)