# CyMorph: fast non-parametric galaxy morphology package

Igor Kolesnikov[1], Vitor Medeiros Sampaio[2], and Reinaldo Ramos de Carvalho[2]

1 Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil **2** NAT-Universidade Cruzeiro do Sul / Universidade Cidade de São Paulo, Rua Galvão Bueno, 868, 01506-000, São Paulo, SP, Brazil

Summary

We present the Cython Morphology (CyMorph), a package that extracts quantitative metrics from the supplied 2d array. This revised and updated version is based on code used in Sautter (2018) and Barchi et al. (2020). Metric extractions is done in a non-parametric way, such that there are no underlying assumptions about the 2d array content. The package relies heavily on the Cython implementation, which makes it much faster than the plain Python version, making it suitable for large data analysis.

The primarily objective is to address galaxy morphology, CyMorph is a fast and optimized package that provides five morphometrics already known to the scientific community: Concentration (C), Asymmetry (A), Smoothness (S), Entropy (H), and second moment of Gradient Pattern Analysis (G2). The input image is expected to contain a single galaxy in the center, free of contaminants and secondary sources. Each extracted metric from such image will quantify a given physical propriety of the galaxy. Additionally to galaxy morphology, CyMorph can also be used with gas clusters and short γ-ray bursts. Finally, by design, metrics are source agnostic and will evaluate any 2d array supplied to the program.

To use CyMorph, it is necessary to supply an image and, if needed, tune the parameters of the metrics. The image preprocessing, such as cleaning or segmentation, could be required, but it is not included in CyMorph code and is the end user's responsibility.

Statement of need

Morphometry and physical proprieties of the galaxies

When CyMorph is used to quantify galactic morphology, each metric measures patterns and forms in the image:

- Concentration measures how tightly pixels are distributed in the galaxy structure.
- Asymmetry measures the irregularity of the shape of the galaxy's disk and bulge.
- Smoothness (also called as Clumpiness) evaluates the presence of the small structures in the galaxy's disk (star-forming regions).
- Entropy is the measure of the heterogeneity of the pixel distribution.
- G2 analyses the vector field of the flux distribution and counts the variation of flux intensity.

One of the most traditional methods to systematize the observed varieties of morphologies is based on Concentration, Asymmetry, and Smoothness e.g., Conselice (2003). These parameters have been used to study the physical properties of galaxies. Concentration can be an important discriminant between spirals and ellipticals Conselice (2003). Besides, Graham et al. (2001) has shown that C correlate with velocity dispersion, galaxy size, luminosity, and black hole

mass. (Conselice, Bershady, and Jangren 2000; Conselice 2003) indicate that A is a good morphological indicator of galaxy interactions and mergers. These processes vary with redshift, so A may be valuable in extending the morphological analysis to the high redshift range. Also, Conselice (2003) studied how the S indicator may reveal how clumpy galaxies are and how this reflects their star formation history. Ellipticals exhibit little signs of star formation and should result in low values of smmothness, while spirals with active star formation should have a more pronounced star-forming regions, and by consequence, higher smoothness values.

The effectiveness in distinguishing Early-type galaxies (ETGs; elliptical) from Late-type galaxies (LTGs; spiral) is explained by the fact when the metrics applied to an image with a clearly elliptical galaxy, usually, C will have high values, and A, S, H, and G2 will have low values. While in the case of a clearly spiral galaxy, it would be expected the opposite. A combination of these parameters can provide a solid intuition of the class of unknown galaxies. More important than having indicators that separate ellipticals and spirals is that they reflect physical properties that can be examined in different environments and at different redshifts, probing the way galaxies evolve.

Morphometry and Machine learning

Besides using the morphometric parameters to study the actual physical processes molding galaxies directly, these parameters can also be used in Machine Learning (ML) methods to discriminate ETGs from LTGs. Barchi et al. (2020) conducts a thorough study of ML using the CASHG2 metrics extracted with previous version of CyMorph as the primary input information to provide the classification netween multiple combinations of morphological classes. Several ML algorithms were tested: Decision Tree (DT), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). DT had a slightly better performance than the other two, with an overall accuracy (OA) of 98.5%, when dealing only with bright galaxies. When applying Deep Learning (DL) technique, they find only a tiny increase in OA, 99.5%. However, when classifying fainter galaxies, DL outperforms ML with 98.7% against 94.8%. The critical point here is that while losing by a thin margin, ML methods trained on the CASHG2 metrics preserved meaningful features, contrary to DL, which was more of a black-box scenario.

Intracluster medium

CyMorph can also be applied to X-Ray map tracing gas from the Intracluster medium (ICM). By combining the results of the galaxy's morphology and ICM hot gas X-Ray morphology properties, it is possible to investigate how the cluster's ICM morphological properties (high asymmetry, for example) affect member galaxies' morphology.

Short γ-ray burst(GRB)

CyMorph was also used by Santana-Silva et al. (in prep.) to trace galaxy properties that are typical for γ-ray burst(GRB) hosts in order to prioritize targets with similar characteristics during gravitational wave optical follow-ups.

Acknowledgements

References

Barchi, PH, RR de Carvalho, RR Rosa, RA Sautter, M Soares-Santos, BAD Marques, E Clua, TS Gonçalves, C de Sá-Freitas, and TC Moura. 2020. "Machine and Deep Learning Applied to Galaxy Morphology-a Comparative Study." *Astronomy and Computing* 30: 100334.

Conselice, Christopher J. 2003. "The Relationship Between Stellar Light Distributions of Galaxies and Their Formation Histories." *The Astrophysical Journal Supplement Series* 147 (1): 1.

Conselice, Christopher J, Matthew A Bershady, and Anna Jangren. 2000. "The Asymmetry of Galaxies: Physical Morphology for Nearby and High-Redshift Galaxies." *The Astrophysical Journal* 529 (2): 886.

Graham, Alister W, Peter Erwin, Nicola Caon, and Ignacio Trujillo. 2001. "A Correlation Between Galaxy Light Concentration and Supermassive Black Hole Mass." *The Astrophysical Journal Letters* 563 (1): L11.

Sautter, Rubens Andreas. 2018. "Gradient Pattern Analysis: New Methodological and Computational Features with Applications." Master's thesis, São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE). http://urlib.net/rep/8JMKD3MGP3W34R/3R3SGF2.