




# CyMorph: fast non-parametric galaxy morphology package

Igor Kolesnikov<sup>1</sup>, Vitor Medeiros Sampaio<sup>2</sup>, and Reinaldo Ramos de Carvalho<sup>2</sup>

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

DOI: [00.00000/joss.00000](https://doi.org/00.00000/joss.00000)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: 

Submitted: 00 January 0000

Published: 00 January 0000

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

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

CyMorph is a fast and optimized package designed primarily to estimate five parameters characterizing the galaxy's morphology: 1) Concentration ( $C$ ); 2) Asymmetry ( $A$ ); 3) Smoothness ( $S$ ); 4) Entropy ( $H$ ); and the second moment of Gradient Pattern Analysis ( $G_2$ ). The input image is expected to contain a single galaxy at the center and cleaned from contaminants (stars, other galaxies, etc). CyMorph has been used also to quantify the morphology of the temperature maps of the Intra-cluster medium (ICM) of clusters of galaxies. Finally, by design, metrics are source agnostic and will evaluate any 2d array supplied to the program.

CyMorph requires only an input image and, if needed, tuning of the parameters involved in the metrics definition. Preprocessing, such as cleaning and creation of the segmented image is required, but not included in CyMorph code.

## Statement of need

### Morphometry and physical proprieties of galaxies

When CyMorph is used to quantify galactic morphology, each metric is supposed to represent how a given physical mechanism alters the shape of the system and we know that this is true in several cases Conselice (2003). CyMorph estimates the following parameters:

- Concentration ( $C$ ) - measures how tight is the flux distribution of a given galaxy;
- Asymmetry ( $A$ ) - as name says, it quantifies how symmetric the galaxy is;
- Smoothness ( $S$ ) (also called Clumpiness) determines how clumpy the distribution is, being sensitive to the the presence of star-forming regions in a late-type galaxy;
- Entropy ( $H$ ) - in this case we measure the Shannon entropy which reflects the average amount of information in the image, and
- $G_2$  - is the second moment of the image gradient pattern which gives the integral of all the assymetric vectors.

One of the most traditional systems designed 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 is critical to distinguish between early- and late-type systems Conselice (2003). Also, as shown by

Graham et al. (2001),  $C$  well correlates with central 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. Conselice (2003) studies how the  $S$  parameter may be associated to the star formation history. Early-type galaxies (ETG) exhibit little signs of star formation and should result in low values of smoothness, while late-type galaxies (LTG) should have more pronounced star-forming regions, and by consequence, higher smoothness values, and in fact that is what is observed. The CAS system, together with  $H$  and  $G_2$ , is very effective in distinguishing between early- and late-type galaxies and a combination of them may provide a solid intuition to find other classes still unknown. More important than having indicators that distinguish early- and late-type galaxies 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 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) conduct a thorough study of ML using the CASHG2 metrics extracted with previous version of CyMorph as the primary input information to provide the classification between 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%.

## Intra-cluster medium

CyMorph has been used to study the temperature maps of the Intra-cluster medium (ICM) of rich clusters of galaxies. This use has proved valuable in understanding how cool-core and non cool-core clusters relate to the overall evolution of the system.

## Acknowledgements

We thank CAPES and FAPESP for the financial support throughout this research.

## 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>.