The Journal of Open Source Software

PyAutoGalaxy: Open-Source Multiwavelength Galaxy Structure & Morphology

James. W. Nightingale[1], Aristeidis Amvrosiadis[1], Richard G. Hayes[1], Qiuhan He[1], Amy Etherington[1], XiaoYue Cao[2], Shaun Cole[1], Jonathan Frawley[3], Carlos S. Frenk[1], Sam Lange[1], Ran Li[2], Richard J. Massey[1], Mattia Negrello[4], and Andrew Robertson[1]

1 Institute for Computational Cosmology, Stockton Rd, Durham DH1 3LE **2** National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China **3** Advanced Research Computing, Durham University, Durham DH1 3LE **4** School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Nearly a century ago, Edwin Hubble famously classified galaxies into three distinct groups: ellipticals, spirals and irregulars (Hubble, 1926). Today, by analysing millions of galaxies with advanced image processing techniques Astronomers have expanded on this picture and revealed the rich diversity of galaxy morphology in both the nearby and distant Universe (Kormendy, 2015) (Vulcani et al., 2014) (Van Der Wel et al., 2012). `PyAutoGalaxy` is an open-source Python 3.8+ package for analysing the morphologies and structures of large multiwavelength galaxy samples, with core features including fully automated Bayesian model-fitting of galaxy two-dimensional surface brightness profiles, support for imaging and interferometer datasets and comprehensive tools for simulating galaxy images. The software places a focus on big data analysis, including support for hierarchical models that simultaneously fit thousands of galaxies, massively parallel model-fitting and an SQLite3 database that allows large suites of modeling results to be loaded, queried and analysed. Accompanying `PyAutoGalaxy` is the autogalaxy workspace, which includes example scripts, datasets and the `HowToGalaxy` lectures in Jupyter notebook format which introduce non-experts to studies of galaxy morphology using `PyAutoGalaxy`. Readers can try `PyAutoGalaxy` right now by going to the introduction Jupyter notebook on Binder or checkout the readthedocs for a complete overview of `PyAutoGalaxy`'s features.

Background

Galaxy morphology studies aim to understand the different luminous structures that galaxies are composed of (Lackner & Gunn, 2012) (Oh et al., 2017). Using large CCD imaging datasets of galaxies observed at ultraviolet, optical and near-infrared wavelengths from instruments like the Hubble Space Telescope (HST), Astronomers have uncovered the plentiful structures that make up a galaxy, such as bars, bulges, disks and rings (Graham, 2013) (Hodge et al., 2019) and revealed that evolving galaxies transition from disk-like structures to bulge-like elliptical galaxies (Coenda et al., 2018). At sub-mm and radio wavelengths interferometer datasets from instruments like the Atacama Large Millimeter Array (ALMA) have revealed the integral role that dust plays in forming galaxies in the distant Universe (Blain et al., 2002) (Casey et al., 2014), early in their lifetimes. Studies typically represent a galaxy's light using analytic functions such as the Sersic profile (Sersic, 1968), which quantify the global appearance of most galaxies into one of three groups: (i) bulge-like structures which follow a de Vaucouleurs profile (de Vaucouleurs, 1948); (ii) disk-like structures which follow an exponential profile or;

The Journal of Open Source Software

(iii) irregular morphologies which are difficult to quantify with symmetric and smooth analytic profiles. Galaxies are often composed of many sub-components which may be a combination of these different structures (Nightingale et al., 2019).

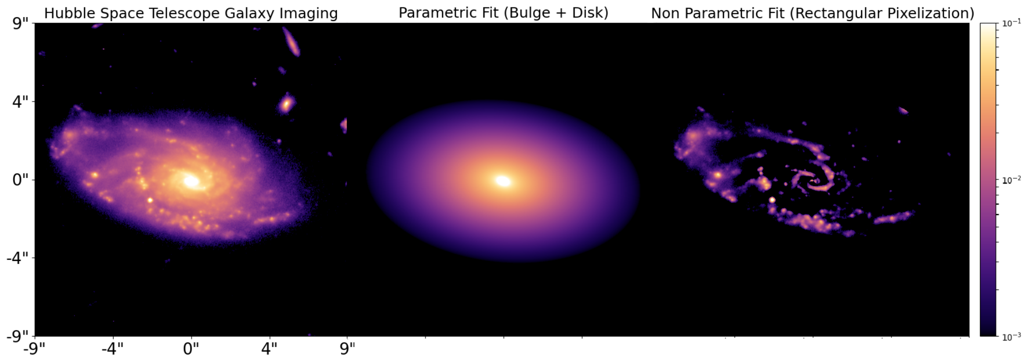

Figure 1: Hubble Space Telescope imaging of a spiral galaxy (left column), a parametric fit to its smooth bulge and disk components (middle column) and a non parametric fit to its asymmetric and irregular structures like its spiral arms (right column).

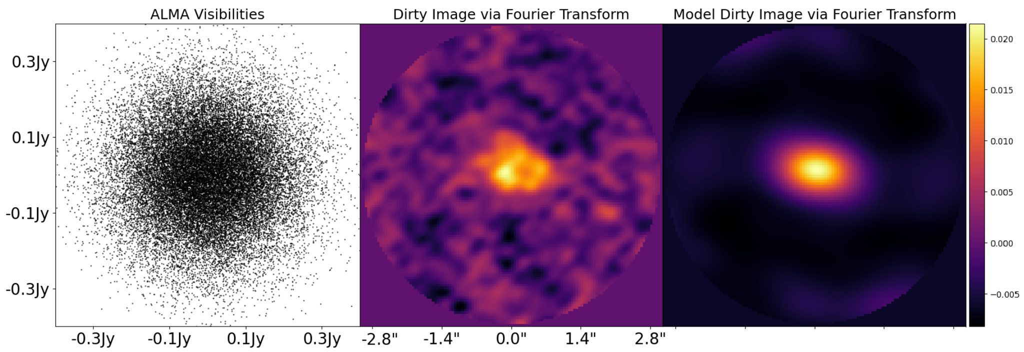

Figure 2: Atacama Large Millimeter Array interferometer visibilities data of a high redshift galaxy (left column), its dirty image created in real space via a Fourier transform (middle column) and a parametric fit to its smooth bulge and disk components which is performed directly on the visibility data (right column).

Figure 1 shows example `PyAutoGalaxy` models of two galaxies taken with two different datasets. The top row shows a structural decomposition of HST imaging of a galaxy, where `PyAutoGalaxy` has decomposed the galaxy into two distinct parametric components, a bulge and disk, whilst simultaneously using a non parametric model to represent the galaxy's irregular and asymmetric spiral arms. Instrumental effects like diffraction from the telescope optics are fully accounted for. Figure 2 shows a `PyAutoGalaxy` fit to ALMA interferometry, where the model galaxy's light is fitted directly in the complex uv-plane and Fourier transformed to real-space for visualization.

Statement of Need

In the next decade, wide field surveys such as Euclid, the Vera Rubin Observatory and Square Kilometer Array are poised to observe images of *billions* of galaxies. Analysing these extremely large galaxy datasets demands advanced Bayesian model-fitting techniques which can scale-up in a fully automated manner. Equally, the James Webb Space Telescope, thirty-meter class ground telescopes and Square Kilometer Array radio interferometer will observe galaxies at an unprecedented resolution and level of detail. This demands more flexible modeling techniques that can accurately represent the complex irregular structures such high resolution observations reveal. PyAutoGalaxy aims to meet both these needs, by interfacing galaxy model-fitting with the probabilistic programming language PyAutoFit to provide Bayesian fitting tools suited to big data analysis alongside image processing tools that represent irregular galaxy structures using non parametric models.

Software API and Features

At the heart of the PyAutoGalaxy API is the Galaxy object, which groups together one or more LightProfile objects at an input redshift. Passing these objects a Grid2D returns an image of the galaxy(s), which can subsequently be passed through Operator objects to apply a 2D convolution or Fast Fourier Transform and thereby compare the Galaxy's image to an imaging or interferometer dataset. The inversion package contains non parametric models which fit a galaxy's light using Bayesian linear matrix algebra. These were originally developed to reconstruct the source galaxies of strong gravitational lenses in PyAutoGalaxy's child project PyAutoLens (Nightingale & Dye, 2015) (Nightingale, Hayes, Kelly, et al., 2021). PyAutoGalaxy includes a comprehensive visualization library for the analysis of both direct imaging and interferometer datasets and tools for preprocessing data to formats suitable to galaxy model-fitting. The astropy cosmology module handles unit conversions and calculations are optimized using the packages NumPy (van der Walt et al., 2011), numba (Lam et al., 2015), PyNUFFT (Lin, 2018) and PyLops (Ravasi & Vasconcelos, 2019).

To perform model-fitting, PyAutoGalaxy adopts the probabilistic programming language PyAutoFit (https://github.com/rhayes777/PyAutoFit). PyAutoFit allows users to compose a model from LightProfile and Galaxy objects, customize the model parameterization and fit it to data via a non-linear search, for example dynesty (Speagle, 2020), emcee (Foreman-Mackey et al., 2013) or PySwarms (Miranda, 2018). By composing a model with Pixelization objects, the galaxy's light is reconstructed using a non parametric rectangular grid that accounts for irregular galaxy morphologies. Multiple images of the same galaxy can be fitted simultaneously, using models whose parameters vary across wavelength.

PyAutoFit's graphical modeling framework allows one to fit a hierarchical model to images of thousands of galaxies simultaneously. Using a technique called expectation propagation (Vehtari et al., 2020), this fits each galaxy dataset one-by-one and combines the results of every fit into a global model using a self-consistent Bayesian framework. Automated fitting of complex galaxy models is possible using PyAutoFit's search chaining, which breaks the fitting of a galaxy into a a chained sequence of non-linear searches. These fits pass information gained about simpler models fitted by earlier searches to subsequent searches, which fit progressively more complex models. By granularizing the model-fitting procedure, automated pipelines that fit complex galaxy models without human intervention can be carefully crafted, with example pipelines found on the autogalaxy workspace. To ensure the analysis and interpretation of fits to large galaxy datasets is feasible, PyAutoFit's database tools write modeling results to a relational database which can be queried from a storage drive to a Python script or Jupyter notebook. This uses memory-light Python generators, ensuring it is practical for results containing hundreds of thousands of galaxies.

Workspace and HowToGalaxy Tutorials

PyAutoGalaxy is distributed with the autogalaxy workspace, which contains example scripts for modeling and simulating galaxies and tutorials on how to preprocess imaging and interferometer datasets before a PyAutoGalaxy analysis. Also included are the HowToGalaxy tutorials, a four chapter lecture series composed of over 20 Jupyter notebooks aimed at non-experts,

The Journal of Open Source Software

introducing them to galaxy morphology analysis, Bayesian inference and teaching them how to use `PyAutoGalaxy` for scientific study. The lectures are available on Binder and may therefore be taken without a local `PyAutoGalaxy` installation.

Software Citations

`PyAutoGalaxy` is written in Python 3.8+ (Van Rossum & Drake, 2009) and uses the following software packages:

- `Astropy` (Astropy Collaboration et al., 2013) (Price-Whelan et al., 2018)
- `COLOSSUS` (Diemer, 2018)
- `corner.py` (Foreman-Mackey, 2016)
- `dynesty` (Speagle, 2020)
- `emcee` (Foreman-Mackey et al., 2013)
- `Matplotlib` (Hunter, 2007)
- `numba` (Lam et al., 2015)
- `NumPy` (van der Walt et al., 2011)
- `PyAutoFit` (Nightingale, Hayes, & Griffiths, 2021)
- `PyLops` (Ravasi & Vasconcelos, 2019)
- `PyNUFFT` (Lin, 2018)
- `pyprojroot` (https://github.com/chendaniely/pyprojroot)
- `PySwarms` (Miranda, 2018)
- `scikit-image` (Van der Walt et al., 2014)
- `scikit-learn` (Pedregosa et al., 2011)
- `Scipy` (Virtanen et al., 2020)

Related Software

- `PyAutoLens` (Nightingale et al., 2018) (Nightingale, Hayes, Kelly, et al., 2021)
- `galfit` https://users.obs.carnegiescience.edu/peng/work/galfit/galfit.html (Peng et al., 2002)
- `GaLight` https://github.com/sibirrer/lenstronomy (Ding et al., 2021)
- `GIM2D` http://www.astro.uvic.ca/~simard/GIM2D/
- `imfit` https://github.com/perwin/imfit
- `megamorph` https://www.nottingham.ac.uk/astronomy/megamorph/ (Häußler et al., 2013)
- `ProFit` https://github.com/ICRAR/ProFit (Robotham et al., 2017)
- `SourceXtractor++` https://github.com/astrorama/SourceXtractorPlusPlus

Acknowledgements

JWN and RJM are supported by the UK Space Agency, through grant ST/V001582/1, and by InnovateUK through grant TS/V002856/1. RGH is supported by STFC Opportunities grant ST/T002565/1. AA, QH, CSF and SMC are supported by ERC Advanced In-vestigator grant, DMIDAS [GA 786910] and also by the STFCConsolidated Grant for Astronomy at Durham [grant numbersST/F001166/1, ST/I00162X/1,ST/P000541/1]. AE and SL aresupported by STFC via grants ST/R504725/1 and ST/T506047/1. RJM is supported by a Royal Society University Research Fellowship. AR is supported by the ERC Horizon2020 project 'EWC' (award AMD-776247-6). MN has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 707601. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., … Streicher, O. (2013). Astropy: A community Python package for astronomy. *558*, A33. https://doi.org/10.1051/0004-6361/201322068

Blain, A. W., Smail, I., Ivison, R. J., Kneib, J. P., & Frayer, D. T. (2002). Submillimeter galaxies. *Physics Report*, *369*(2), 111–176. https://doi.org/10.1016/S0370-1573(02)00134-5

Casey, C. M., Narayanan, D., & Cooray, A. (2014). Dusty star-forming galaxies at high redshift. *Physics Reports*, *541*(2), 45–161. https://doi.org/10.1016/j.physrep.2014.02.009

Coenda, V., Martínez, H. J., & Muriel, H. (2018). Green valley galaxies as a transition population in different environments. *Monthly Notices of the Royal Astronomical Society*, *473*(4), 5617–5629. https://doi.org/10.1093/mnras/stx2707

de Vaucouleurs, G. (1948). Recherches sur les Nebuleuses Extragalactiques. *Annales d'Astrophysique*, *11*, 247.

Diemer, B. (2018). COLOSSUS: A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos. *The Astrophysical Journal Supplement Series*, *239*(2), 35. https://doi.org/10.3847/1538-4365/aaee8c

Ding, X., Birrer, S., Treu, T., & Silverman, J. D. (2021). *Galaxy shapes of Light (GaLight): a 2D modeling of galaxy images*. *8583*. http://arxiv.org/abs/2111.08721

Foreman-Mackey, D. (2016). Corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, *1*(2), 24. https://doi.org/10.21105/joss.00024

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee : The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, *125*(925), 306–312. https://doi.org/10.1086/670067

Graham, A. W. (2013). *Elliptical and disk galaxy structure and modern scaling laws* (Vol. 6, pp. 91–139). https://doi.org/10.1007/978-94-007-5609-0_2

Häußler, B., Bamford, S. P., Vika, M., Rojas, A. L., Barden, M., Kelvin, L. S., Alpaslan, M., Robotham, A. S. G., Driver, S. P., Baldry, I. K., Brough, S., Hopkins, A. M., Liske, J., Nichol, R. C., Popescu, C. C., & Tuffs, R. J. (2013). Megamorph - multiwavelength measurement of galaxy structure: Complete Sérsic profile information from modern surveys. *Monthly Notices of the Royal Astronomical Society*, *430*(1), 330–369. https://doi.org/10.1093/mnras/sts633

Hodge, J. A., Smail, I., Walter, F., Cunha, E. da, Swinbank, A. M., Rybak, M., Venemans, B., Brandt, W. N., Rivera, G. C., Chapman, S. C., Chen, C.-C., Cox, P., Dannerbauer, H., Decarli, R., Greve, T. R., Knudsen, K. K., Menten, K. M., Schinnerer, E., Simpson, J. M., … Weiss, A. (2019). ALMA Reveals Potential Evidence for Spiral Arms, Bars, and Rings in High-redshift Submillimeter Galaxies. *The Astrophysical Journal*, *876*(2), 130. https://doi.org/10.3847/1538-4357/ab1846

Hubble, E. P. (1926). Extragalactic nebulae. *The Astrophysical Journal*, *64*, 321. https://doi.org/10.1086/143018

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Kormendy, J. (2015). Elliptical galaxies and bulges of disc galaxies: Summary of progress and outstanding issues. *Galactic Bulges*, *418*, 431–477. https://doi.org/10.1007/978-3-319-19378-6_16

Lackner, C. N., & Gunn, J. E. (2012). Astrophysically motivated bulge-disc decompositions of Sloan Digital Sky Survey galaxies. *Monthly Notices of the Royal Astronomical Society*, *421*(3), 2277–2302. https://doi.org/10.1111/j.1365-2966.2012.20450.x

Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: a LLVM-based Python JIT compiler. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC - LLVM '15*, 1–6. https://doi.org/10.1145/2833157.2833162

Lin, J. M. (2018). Python non-uniform fast fourier transform (PyNUFFT): An accelerated non-cartesian MRI package on a heterogeneous platform (CPU/GPU). *Journal of Imaging*, *4*(3), 1–22. https://doi.org/10.3390/jimaging4030051

Miranda, L. J. V. (2018). PySwarms, a research-toolkit for Particle Swarm Optimization in Python. *Journal of Open Source Software*, *3*. https://doi.org/10.21105/joss.00433

Nightingale, J. W., & Dye, S. (2015). Adaptive semi-linear inversion of strong gravitational lens imaging. *MNRAS*, *452*(3), 2940–2959. https://doi.org/10.1093/mnras/stv1455

Nightingale, J. W., Dye, S., & Massey, R. J. (2018). AutoLens: Automated modeling of a strong lens's light, mass, and source. *MNRAS*, *478*(4), 4738–4784. https://doi.org/10.1093/mnras/sty1264

Nightingale, J. W., Hayes, R. G., & Griffiths, M. (2021). 'PyAutoFit': A classy probabilistic programming language for model composition and fitting. *Journal of Open Source Software*, *6*(58), 2550. https://doi.org/10.21105/joss.02550

Nightingale, J. W., Hayes, R. G., Kelly, A., Amvrosiadis, A., Etherington, A., He, Q., Li, N., Cao, X., Frawley, J., Cole, S., Enia, A., Frenk, C. S., Harvey, D. R., Li, R., Massey, R. J., Negrello, M., & Robertson, A. (2021). 'PyAutoLens': Open-source strong gravitational lensing. *Journal of Open Source Software*, *6*(58), 2825. https://doi.org/10.21105/joss.02825

Nightingale, J. W., Massey, R. J., Harvey, D. R., Cooper, A. P., Etherington, A., Tam, S. I., & Hayes, R. G. (2019). Galaxy structure with strong gravitational lensing: Decomposing the internal mass distribution of massive elliptical galaxies. *MNRAS*, *489*(2), 2049–2068. https://doi.org/10.1093/mnras/stz2220

Oh, S., Greene, J. E., & Lackner, C. N. (2017). Testing the Presence of Multiple Photometric Components in Nearby Early-type Galaxies Using SDSS. *ApJ*, *836*(1), 115. https://doi.org/10.3847/1538-4357/836/1/115

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. (2002). Detailed Structural Decomposition of Galaxy Images. *The Astronomical Journal*, *124*(1), 266–293. https://doi.org/10.1086/340952

Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Paper Contributors, (Primary., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., … Contributors, (Astropy. (2018). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *156*, 123. https://doi.org/10.3847/1538-3881/aabc4f

Ravasi, M., & Vasconcelos, I. (2019). *PyLops – A Linear-Operator Python Library for large scale optimization*. http://arxiv.org/abs/1907.12349

Robotham, A. S. G., Taranu, D. S., Tobar, R., Moffett, A., & Driver, S. P. (2017). Profit: Bayesian profile fitting of galaxy images. *Monthly Notices of the Royal Astronomical Society*, *466*(2), 1513–1541. https://doi.org/10.1093/mnras/stw3039

Sersic, J. L. (1968). *Atlas de galaxias australes*. http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1968adga.book.....S&link_type=CITATIONS%5Cnpapers://dcc533b5-8613-47b7-b88c-2b0c0d39c33f/Paper/p10728

Speagle, J. S. (2020). dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *MNRAS*, *493*(3), 3132–3158. https://doi.org/10.1093/mnras/staa278

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array2D: A structure for efficient numerical computation. *Computing in Science Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in python. *PeerJ*, *2*, e453.

Van Der Wel, A., Bell, E. F., Häussler, B., McGrath, E. J., Chang, Y. Y., Guo, Y., McIntosh, D. H., Rix, H. W., Barden, M., Cheung, E., Faber, S. M., Ferguson, H. C., Galametz, A., Grogin, N. A., Hartley, W., Kartaltepe, J. S., Kocevski, D. D., Koekemoer, A. M., Lotz, J., … Peng, C. Y. (2012). Structural parameters of galaxies in candels. *Astrophysical Journal, Supplement Series*, *203*(2). https://doi.org/10.1088/0067-0049/203/2/24

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace. ISBN: 1441412697

Vehtari, A., Gelman, A., Sivula, T., Jylänki, P., Tran, D., Sahai, S., Blomstedt, P., Cunningham, J. P., Schiminovich, D., & Robert, C. P. (2020). Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data. *Journal of Machine Learning Research*, *21*, 1–53. https://arxiv.org/abs/1412.4869

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Contributors, S. 1. 0. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vulcani, B., Bamford, S. P., Häußler, B., Vika, M., Rojas, A., Agius, N. K., Baldry, I., Bauer, A. E., Brown, M. J. I., Driver, S., Graham, A. W., Kelvin, L. S., Liske, J., Loveday, J., Popescu, C. C., Robotham, A. S. G., & Tuffs, R. J. (2014). Galaxy and mass assembly (GAMA): The wavelength-dependent sizes and profiles of galaxies revealed by megamorph. *Monthly Notices of the Royal Astronomical Society*, *441*(2), 1340–1362. https://doi.org/10.1093/mnras/stu632