galsbi: A Python package for the GalSBI galaxy population model

Silvan Fischbacher[1], **Beatrice Moser**[1], **Tomasz Kacprzak**[1,2], **Joerg Herbel**[1], **Luca Tortorelli**[1,3], **Uwe Schmitt**[1], **Alexandre Refregier**[1], and **Adam Amara**[1,5]

1 ETH Zurich, Institute for Particle Physics and Astrophysics, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland **2** Swiss Data Science Center, Paul Scherrer Institute, Forschungsstrasse 111, 5232 Villigen, Switzerland **3** University Observatory, Faculty of Physics, Ludwig-Maximilian-Universität München, Scheinerstrasse 1, 81679 Munich, Germany **4** ETH Zurich, Scientific IT Services, Binzmühlestrasse 130, 8092 Zürich, Switzerland **5** School of Mathematics and Physics, University of Surrey, Guildford, Surrey, GU2 7XH, UK

DOI: 10.21105/joss.08766

Software

- Review
- Repository
- Archive

Editor: Josh Borrow

Reviewers:

- @aboucaud
- @jborrow

Submitted: 16 December 2024
Published: 25 October 2025

Summary

Large-scale structure surveys measure the shapes and positions of millions of galaxies in order to constrain the cosmological model with high precision. The resulting large data volume poses a challenge for the analysis of the data, from the estimation of photometric redshifts to the calibration of shape measurements. We present GalSBI, a model for the galaxy population, to address these challenges. This phenomenological model is constrained by observational data using simulation-based inference (SBI). The `galsbi` Python package provides an easy interface to generate catalogs of galaxies based on the GalSBI model, including their photometric properties, and to simulate realistic images of these galaxies using the `UFig` package.

Statement of need

The analysis of large-scale structure surveys can use realistic galaxy catalogs in various applications, such as the measurement of photometric redshifts, the calibration of shape measurements, also under the influence of source blending, or the modelling of complex selection functions. A promising approach to tackle these challenges is forward modeling of the galaxy population. Several such approaches have recently emerged in the literature, differing in their treatment of noise models, galaxy SED modeling, and population sampling strategies. Apart from our efforts with GalSBI, other approaches include SkyPy (Amara et al., 2021), PopSED (Li et al., 2024), and pop-cosmos (Alsing et al., 2024; Thorp et al., 2024).

GalSBI is a parametric galaxy population model constrained by data using SBI (see Cranmer et al., 2020 for a review). Based on one set of model parameters, a galaxy catalog is generated. This catalog is then rendered into a realistic astronomical image using the `UFig` package (Bergé et al., 2013; Fischbacher, Moser, et al., 2025). The realism of the image relies on two key components: accurate forward modelling of image systematics such as the point spread function (PSF) and the background noise, and a realistic galaxy catalog. For the former, we refer to (Bergé et al., 2013; Fischbacher, Moser, et al., 2025; Herbel et al., 2017), while the latter is provided by the GalSBI model.

To produce a realistic galaxy catalog, the galaxy population model must be constrained by data. The first version of the galaxy population model, described in Herbel et al. (2017), uses data from the Suprime-Cam instrument on the Subaru Telescope in the COSMOS field to constrain the model. This model was extended by Kacprzak et al. (2020) to measure cosmic shear with the Dark Energy Survey (for more details, see Bruderer et al., 2016, 2018; Chang et al., 2015). Tortorelli et al. (2020) uses the GalSBI framework to measure the B-band galaxy luminosity function using data from the Canada-France-Hawaii Telescope Legacy Survey. In Tortorelli et al. (2021), the model is applied to measure narrow-band galaxy properties of the PAU survey. Fagioli et al. (2018) and Fagioli et al. (2020) use the model to simulate galaxy spectra of the Sloan Digital Sky Survey CMASS sample. Berner et al. (2024) utilizes galaxies sampled from the GalSBI model to produce a realistic spatial distribution of galaxies using a subhalo-abundance matching approach. Further refinements to the model are described in Moser et al. (2024), where they use Hyper Suprime-Cam (HSC) deep fields to constrain the model to high redshift. The first public release of the phenomenological model, incorporating several model extensions is described in Fischbacher, Kacprzak, et al. (2025). Additionally, Tortorelli et al. (2025) presents a first version of the GalSBI model based on stellar population synthesis.

With the constrained model, we can generate realistic intrinsic galaxy catalogs for various applications. Rendering the catalogs into realistic astronomical images can help to calibrate the shape measurements of galaxies, also under the influence of source blending. Performing source extraction on the simulated images results in realistic measured galaxy catalogs including the redshift distribution. Furthermore, the impact of selection effects can be easily studied by applying the selection function to the catalogs and directly measuring the impact on the observables.

Features

The main `galsbi` layer allows the user to generate realistic galaxy catalogs based on published GalSBI models as described in Moser et al. (2024) or Fischbacher, Kacprzak, et al. (2025). With just a few lines of code, the user can generate an intrinsic catalog, simulate astronomical images or run one of the emulators described in Fischbacher, Kacprzak, et al. (2025) to obtain a measured catalog. We provide the configuration files of these prepared setups in the package to make it easy for the user to get started. However, starting with one of these setups, the user can easily modify the configuration files to adapt the model to their specific needs.

Furthermore, we provide the catalog generator `ucat` as a subpackage of `galsbi`. In `ucat`, the user can define their own galaxy population model using a variety of model choices with different parametrizations. `ucat` is used by the main `galsbi` layer to generate the catalogs and the user can use `ucat` directly to generate catalogs based on their own model. An overview of the different components described above is given in the Table below.

Component	Core functionality	Details	**`galsbi` connection**
`galsbi.GalSBI`	A convenience layer to load GalSBI models and create intrinsic and measured catalogs based from them.	Provides predefined configurations to run `ucat` and `UFig` plugins. Configurations can be easily customized.	The main interface for running and customizing workflows in the `galsbi` framework.
`galsbi.ucat`	A subpackage implementing the phenomenological galaxy population modeling.	Samples intrinsic galaxy properties like magnitudes, sizes, and ellipticities, and provides the `ucat` plugins	A subpackage in `galsbi` that is also called by the main interface.

The Journal of Open Source Software

Component	Core functionality	Details	`galsbi` connection
`UFig`	An external package (see Fischbacher, Moser, et al., 2025) to obtain a measured catalog based from an intrinsic catalog (e.g. generated by GalSBI)	Adds PSF and background to images, can render images and perform source extraction on them or emulate the transfer function from intrinsic to measured catalog.	`UFig` plugins are used in the predefined configuration files of the `galsbi` interface

Using the model from Fischbacher, Kacprzak, et al. (2025), sampling a catalog for an HSC deep field simulation in five bands takes about five seconds. This is faster than simulating a single band with `UFig`. However, the runtime depends on the simulation area, depth, and, to a lesser extent, the chosen galaxy population model.

The GalSBI model overview

In this section, we give a short overview of the GalSBI model. We focus on the constrained model as described in Fischbacher, Kacprzak, et al. (2025) but the package offers a variety of model choices and parametrizations that are described in the documentation. For interactive versions of the figures, please refer to the corresponding section in the documentation.

Luminosity functions

The initial galaxy catalog is sampled from two luminosity functions for the red and blue galaxy populations. The luminosity functions are described by a Schechter function with parameters ϕ^*, M^*, and α. The two parameters ϕ^* and M^* vary as a function of redshift and `galsbi` includes several parametrizations for these functions. Figure 1 shows the blue and red luminosity functions based on the model from Fischbacher, Kacprzak, et al. (2025) as a function of redshift as well as a simulated image for this specific choice of luminosity functions. The luminosity function determines the number of galaxies in a given area and the absolute luminosity and the redshift of each sampled galaxies.

Galaxy spectra

In order to obtain an apparent magnitude, each galaxy is assigned a spectrum using a linear combination of the kcorrect templates (Blanton & Roweis, 2007). The coefficients of the templates are drawn from a Dirichlet distribution such that they sum to one. The resulting total spectrum is then normalized to match the absolute magnitude of the galaxy. The apparent magnitude is calculated by applying reddening due to galactic extinction, redshifting the spectrum and integrating it over the filter band.

Galaxy morphology

The half-light radii of the galaxies are sampled from a log-normal distribution that depends on the absolute magnitude and redshift. Figure 2 shows the half-light radius as a function of redshift and absolute magnitude for the blue and red galaxy populations based on the model from Fischbacher, Kacprzak, et al. (2025).

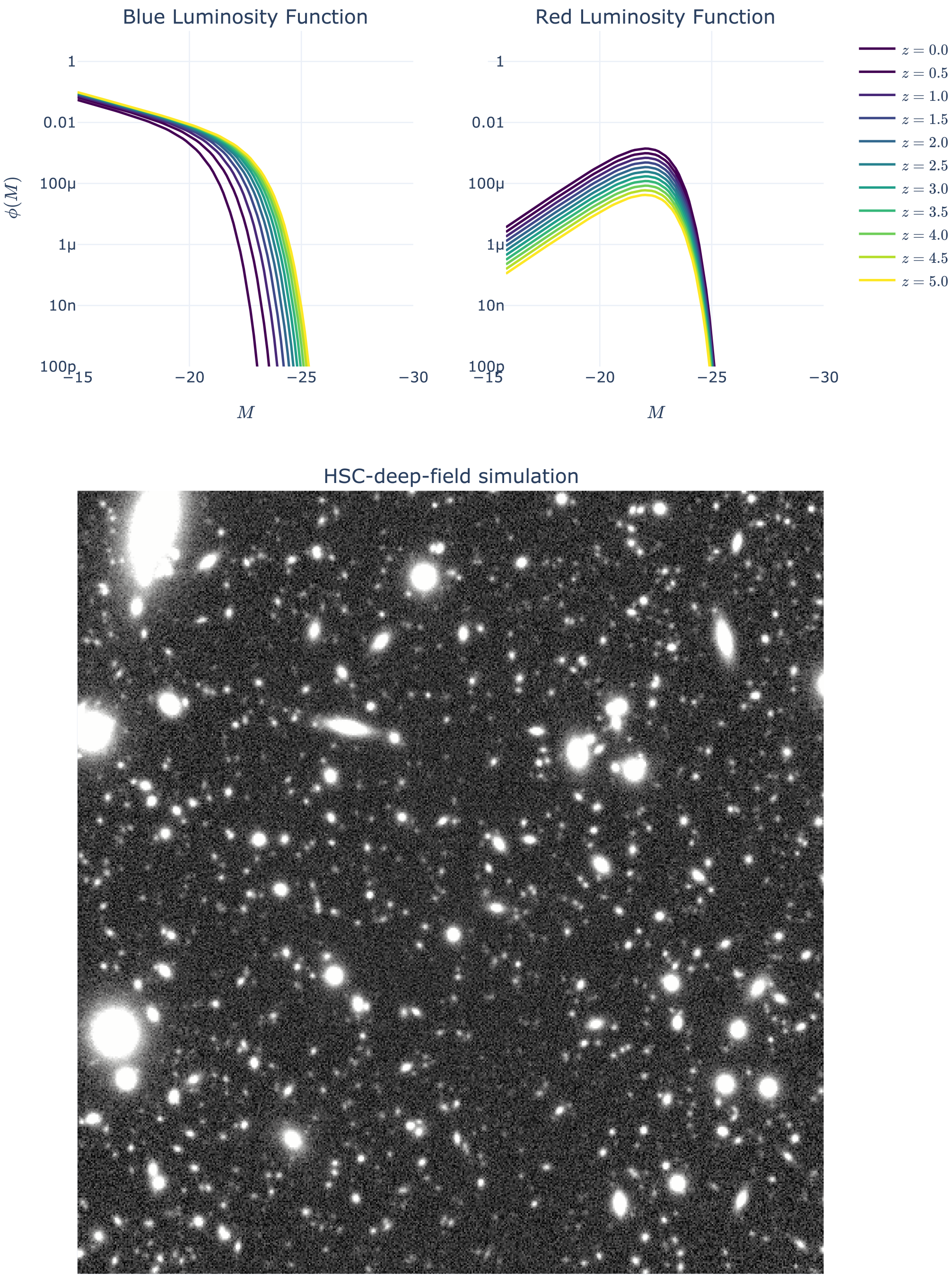

Figure 1: Luminosity functions of red and blue galaxies as a function of absolute magnitude M. The redshift evolution of the luminosity function is represented by the color gradient, transitioning from low redshift (blue) to high redshift (yellow). The lower panel displays an HSC deep field-like image generated using the above luminosity functions. An interactive version of this plot, including live updates to the image, is available in the documentation.

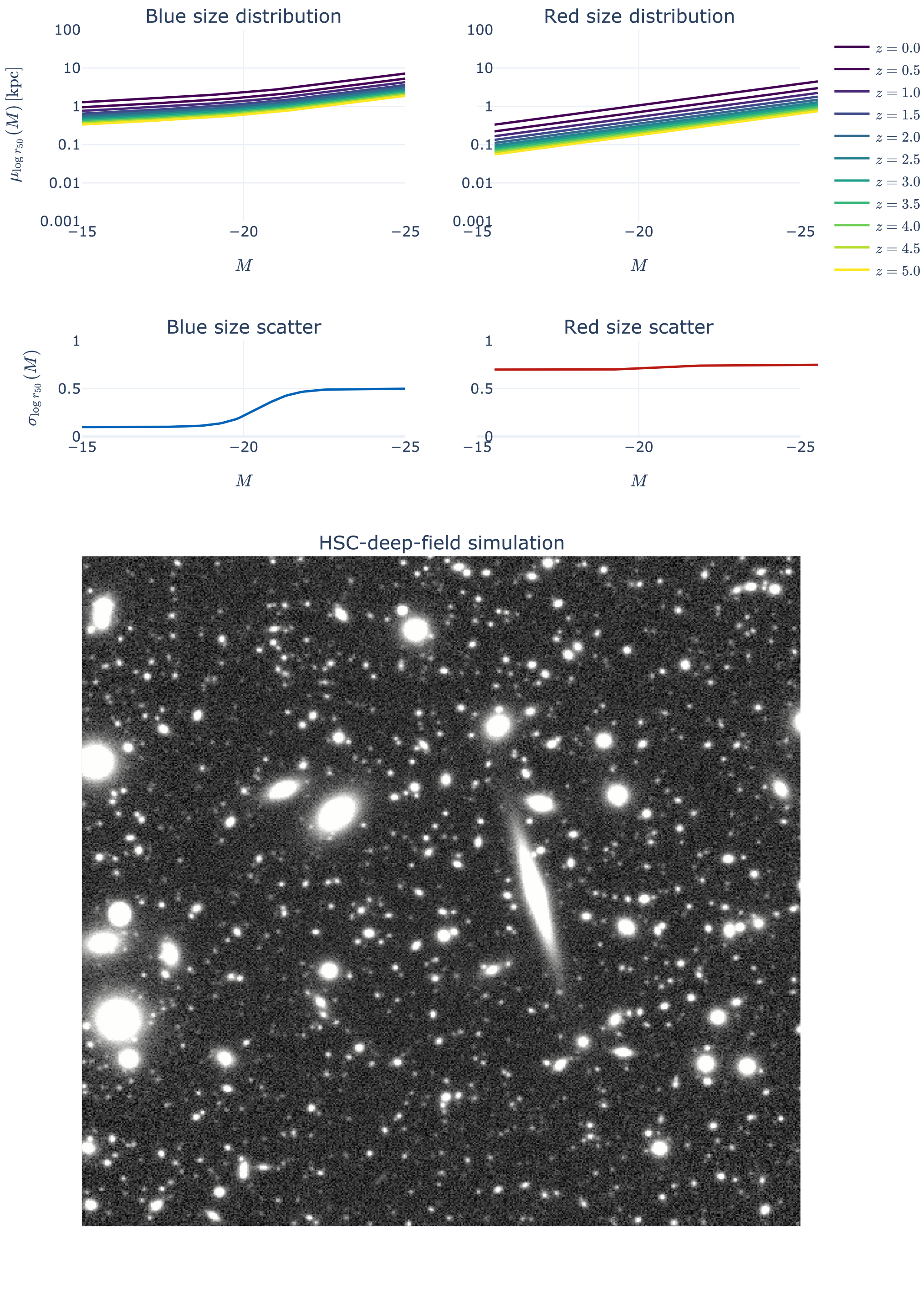

Figure 2: Mean and standard deviation of the log-normal size distribution as a function of absolute magnitude M for red and blue galaxies. The redshift evolution of the mean is represented by the color gradient, transitioning from low redshift (blue) to high redshift (yellow). The lower panel displays an HSC deep field-like image generated using the above size model. An interactive version of this plot, including live updates to the image, is available in the documentation.

The ellipticity of the galaxies is defined as a complex number $e = e_1 + ie_2$. UFig requires the two components e_1 and e_2 to render an image. Depending on the sampling method, the two components are either sampled from Gaussian distributions or the absolute ellipticity $|e| = \sqrt{e^1 + e^2}$ is sampled using different prescriptions and the phase is sampled uniformly.

Finally, each galaxy is assigned a light profile characterized by its Sersic index. In Fischbacher, Kacprzak, et al. (2025), the Sersic index is sampled from a beta prime distribution.

For more details on the available model choices and parametrizations, please refer to the documentation. A more comprehensive description of the physical motivation of the model can be found in Fischbacher, Kacprzak, et al. (2025).

Acknowledgements

This project was supported in part by grant 200021_143906, 200021_169130 and 200021_192243 from the Swiss National Science Foundation.

We acknowledge the use of the following software packages: Astropy (Astropy Collaboration et al., 2013), *healpy* (Zonca et al., 2019), NumPy (Van Der Walt et al., 2011), PyCosmo (Moser et al., 2022; Refregier et al., 2018; Tarsitano et al., 2021), SciPy (Virtanen et al., 2020), and `ufig` (Bergé et al., 2013; Fischbacher, Moser, et al., 2025). For the plots in this paper and the documentation, we used Matplotlib (Hunter, 2007), Plotly (Plotly Technologies Inc., 2015) and `trianglechain` (Fischbacher et al., 2023; Kacprzak & Fluri, 2022).

References

Alsing, J., Thorp, S., Deger, S., Peiris, H. V., Leistedt, B., Mortlock, D., & Leja, J. (2024). Pop-cosmos: A comprehensive picture of the galaxy population from COSMOS data. *The Astrophysical Journal Supplement Series*, *274*(1), 12. https://doi.org/10.3847/1538-4365/ad5c69

Amara, A., Bella, L. F. de la, Birrer, S., Bridle, S., Cordero, J. P., Favole, G., Harrison, I., Harry, I. W., Hartley, W. G., Krawczyk, C., Lundgren, A., Nord, B., Nuttall, L. K., Rollins, R. P., Sudek, P., Tam, S.-I., Tessore, N., Tolley, A. E., Umetsu, K., … Wolz, L. (2021). SkyPy: A package for modelling the Universe. *Journal of Open Source Software*, *6*(65), 3056. https://doi.org/10.21105/joss.03056

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. *Astronomy and Astrophysics*, *558*, A33. https://doi.org/10.1051/0004-6361/201322068

Bergé, J., Gamper, L., Réfrégier, A., & Amara, A. (2013). An ultra fast image generator (UFig) for wide-field astronomy. *Astronomy and Computing*, *1*, 23–32. https://doi.org/10.1016/j.ascom.2013.01.001

Berner, P., Refregier, A., Moser, B., Tortorelli, L., Machado Poletti Valle, L. F., & Kacprzak, T. (2024). Fast forward modelling of galaxy spatial and statistical distributions. *Journal of Cosmology and Astroparticle Physics*, *2024*(04), 023. https://doi.org/10.1088/1475-7516/2024/04/023

Blanton, M. R., & Roweis, S. (2007). K-corrections and filter transformations in the ultraviolet, optical, and near infrared. *The Astronomical Journal*, *133*(2), 734–754. https://doi.org/10.1086/510127

Bruderer, C., Chang, C., Refregier, A., Amara, A., Berge, J., & Gamper, L. (2016). Calibrated ultra fast image simulations for the Dark Energy Survey. *The Astrophysical Journal*, *817*(1),

25. https://doi.org/10.3847/0004-637X/817/1/25

Bruderer, C., Nicola, A., Amara, A., Refregier, A., Herbel, J., & Kacprzak, T. (2018). Cosmic shear calibration with forward modeling. *Journal of Cosmology and Astroparticle Physics*, *2018*(08), 007. https://doi.org/10.1088/1475-7516/2018/08/007

Chang, C., Busha, M. T., Wechsler, R. H., Refregier, A., Amara, A., Rykoff, E., Becker, M. R., Bruderer, C., Gamper, L., Leistedt, B., Peiris, H., Abbott, T., Abdalla, F. B., Balbinot, E., Banerji, M., Bernstein, R. A., Bertin, E., Brooks, D., Carnero, A., … Zuntz, J. (2015). Modeling the transfer function for the Dark Energy Survey. *The Astrophysical Journal*, *801*, 73. https://doi.org/10.1088/0004-637X/801/2/73

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062. https://doi.org/10.1073/pnas.1912789117

Fagioli, M., Riebartsch, J., Nicola, A., Herbel, J., Amara, A., Refregier, A., Chang, C., Gamper, L., & Tortorelli, L. (2018). Forward modeling of spectroscopic galaxy surveys: Application to SDSS. *Journal of Cosmology and Astroparticle Physics*, *2018*(11), 015. https://doi.org/10.1088/1475-7516/2018/11/015

Fagioli, M., Tortorelli, L., Herbel, J., Zürcher, D., Refregier, A., & Amara, A. (2020). Spectro-imaging forward model of red and blue galaxies. *Journal of Cosmology and Astroparticle Physics*, *2020*(06), 050. https://doi.org/10.1088/1475-7516/2020/06/050

Fischbacher, S., Kacprzak, T., Blazek, J., & Refregier, A. (2023). Redshift requirements for cosmic shear with intrinsic alignment. *Journal of Cosmology and Astroparticle Physics*, *2023*(01), 033. https://doi.org/10.1088/1475-7516/2023/01/033

Fischbacher, S., Kacprzak, T., Tortorelli, L., Moser, B., Refregier, A., Gebhardt, P., & Gruen, D. (2025). GalSBI: Phenomenological galaxy population model for cosmology using simulation-based inference. *Journal of Cosmology and Astroparticle Physics*, *2025*(06), 007. https://doi.org/10.1088/1475-7516/2025/06/007

Fischbacher, S., Moser, B., Kacprzak, T., Tortorelli, L., Herbel, J., Bruderer, C., Schmitt, U., Refregier, A., Berge, J., Gamper, L., & Amara, A. (2025). UFig v1: The ultra-fast image generator. *Journal of Open Source Software*, *10*(113), 8697. https://doi.org/10.21105/joss.08697

Herbel, J., Kacprzak, T., Amara, A., Refregier, A., Bruderer, C., & Nicola, A. (2017). The redshift distribution of cosmological samples: A forward modeling approach. *Journal of Cosmology and Astroparticle Physics*, *2017*(08), 035–035. https://doi.org/10.1088/1475-7516/2017/08/035

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

Kacprzak, T., & Fluri, J. (2022). DeepLSS: Breaking parameter degeneracies in large-scale structure with deep-learning analysis of combined probes. *Physical Review X*, *12*, 031029. https://doi.org/10.1103/PhysRevX.12.031029

Kacprzak, T., Herbel, J., Nicola, A., Sgier, R., Tarsitano, F., Bruderer, C., Amara, A., Refregier, A., Bridle, S. L., Drlica-Wagner, A., Gruen, D., Hartley, W. G., Hoyle, B., Secco, L. F., Zuntz, J., Annis, J., Avila, S., Bertin, E., Brooks, D., … Weller, J. (2020). Monte Carlo control loops for cosmic shear cosmology with DES Year 1. *Physical Review D*, *101*(8), 082003. https://doi.org/10.1103/PhysRevD.101.082003

Li, J., Melchior, P., Hahn, C., & Huang, S. (2024). PopSED: Population-level inference for galaxy properties from broadband photometry with neural density estimation. *The Astronomical Journal*, *167*(1), 16. https://doi.org/10.3847/1538-3881/ad0be4

Moser, B., Kacprzak, T., Fischbacher, S., Refregier, A., Grimm, D., & Tortorelli, L. (2024).

Simulation-based inference of deep fields: Galaxy population model and redshift distributions. *Journal of Cosmology and Astroparticle Physics*, *2024*(05), 049. https://doi.org/10.1088/1475-7516/2024/05/049

Moser, B., Lorenz, C. S., Schmitt, U., Réfrégier, A., Fluri, J., Sgier, R., Tarsitano, F., & Heisenberg, L. (2022). Symbolic implementation of extensions of the PyCosmo Boltzmann solver. *Astronomy and Computing*, *40*, 100603. https://doi.org/10.1016/j.ascom.2022.100603

Plotly Technologies Inc. (2015). *Collaborative data science*. Plotly Technologies Inc. https://plot.ly

Refregier, A., Gamper, L., Amara, A., & Heisenberg, L. (2018). PyCosmo: An integrated cosmological Boltzmann solver. *Astronomy and Computing*, *25*, 38–43. https://doi.org/10.1016/j.ascom.2018.08.001

Tarsitano, F., Schmitt, U., Refregier, A., Fluri, J., Sgier, R., Nicola, A., Herbel, J., Amara, A., Kacprzak, T., & Heisenberg, L. (2021). Predicting cosmological observables with PyCosmo. *Astronomy and Computing*, *36*, 100484. https://doi.org/10.1016/j.ascom.2021.100484

Thorp, S., Alsing, J., Peiris, H. V., Deger, S., Mortlock, D. J., Leistedt, B., Leja, J., & Loureiro, A. (2024). Pop-cosmos: Scaleable inference of galaxy properties and redshifts with a data-driven population model. *The Astrophysical Journal*, *975*(1), 145. https://doi.org/10.3847/1538-4357/ad7736

Tortorelli, L., Fagioli, M., Herbel, J., Amara, A., Kacprzak, T., & Refregier, A. (2020). Measurement of the b-band galaxy luminosity function with approximate Bayesian computation. *Journal of Cosmology and Astroparticle Physics*, *2020*(09), 048. https://doi.org/10.1088/1475-7516/2020/09/048

Tortorelli, L., Fischbacher, S., Refregier, A., Bellstedt, S., Robotham, A. S. G., & Kacprzak, T. (2025). *GalSBI-SPS: A stellar population synthesis-based galaxy population model for cosmology and galaxy evolution applications*. *arXiv:2505.21610*. https://doi.org/10.48550/arXiv.2505.21610

Tortorelli, L., Siudek, M., Moser, B., Kacprzak, T., Berner, P., Refregier, A., Amara, A., García-Bellido, J., Cabayol, L., Carretero, J., Castander, F. J., Vicente, J. D., Eriksen, M., Fernandez, E., Gaztanaga, E., Hildebrandt, H., Joachimi, B., Miquel, R., Sevilla-Noarbe, I., … Wright, A. H. (2021). The PAU survey: Measurement of narrow-band galaxy properties with approximate bayesian computation. *Journal of Cosmology and Astroparticle Physics*, *2021*(12), 013. https://doi.org/10.1088/1475-7516/2021/12/013

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

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J. van der, Brett, M., Wilson, J., Millman, K. J., 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*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Zonca, A., Singer, L. P., Lenz, D., Reinecke, M., Rosset, C., Hivon, E., & Gorski, K. M. (2019). Healpy: Equal area pixelization and spherical harmonics transforms for data on the sphere in Python. *Journal of Open Source Software*, *4*(35), 1298. https://doi.org/10.21105/joss.01298