The Journal of Open Source Software

SOAP: A Python Package for Calculating the Properties of Galaxies and Halos Formed in Cosmological Simulations

Robert McGibbon [1], **John C. Helly** [2], **Joop Schaye** [1], **Matthieu Schaller** [1,3], **and Bert Vandenbroucke** [1]

1 Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands **2** Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK **3** Lorentz Institute for Theoretical Physics, Leiden University, PO box 9506, 2300 RA Leiden, the Netherlands

DOI: 10.21105/joss.08252

Software

- Review
- Repository
- Archive

Editor: Ivelina Momcheva

Reviewers:

- @dnelson86
- @stuartmcalpine
- @nvanalfen

Submitted: 20 February 2025
Published: 23 July 2025

Summary

Cosmological simulations model the evolution of dark matter and baryons under the influence of gravitational and hydrodynamic forces. Beginning at high redshift, they capture the hierarchical formation of structures, where smaller structures form first and later merge into larger ones.These simulations incorporate hydrodynamics to evolve the gas and include a number of subgrid prescriptions for modelling important physical processes, such as star formation. Once the simulation has concluded, a halo-finding algorithm is used to identify bound structures (subhaloes) within the resulting particle distribution.

Here, we introduce SOAP (Spherical Overdensity & Aperture Processor), a Python package designed to compute halo and galaxy properties from simulations that have been post-processed with a halo finder. SOAP takes a subhalo catalogue as input and calculates a wide array of properties for each object. Its output is compatible with the swiftsimio package (Borrow & Borrisov, 2020), enabling seamless unit handling. SOAP has already been used to generate halo catalogues for the FLAMINGO simulation suite (Kugel et al., 2023; Schaye et al., 2023), which includes the largest cosmological hydrodynamic simulation to date. These catalogues have been used in more than 20 publications to date [1].

SOAP is hosted on GitHub. We strongly encourage contributions to SOAP, such as opening issues and submitting pull requests.

Statement of Need

Modern galaxy simulations are often analysed by a large number of researchers. However, due to the large volume of data generated, it is often impractical for individual users to compute the specific properties they require independently. SOAP addresses this challenge by producing comprehensive catalogues containing a wide range of properties that can be shared across the community.

Given the substantial volume of data, it is essential for the output to be processed in parallel. SOAP achieves this using the mpi4py library (Dalcín et al., 2005). This enables SOAP to scale efficiently across multiple compute nodes. SOAP is also designed to handle subvolumes of the simulation independently, allowing for large simulations to be processed sequentially if required. This approach reduces the need for high-memory computing nodes, making it possible to

[1] For a complete list, see https://flamingo.strw.leidenuniv.nl/papers.html

process simulation outputs without requiring a large number of high-memory resources. The ability to efficently process subhalos in parallel is a unique feature of SOAP when compared with other packages for computing galaxy properties (e.g. Dome, 2023; Narayanan et al., 2023; Pontzen & Tremmel, 2018).

A large number of halo finders are used by the community to identify bound structures within simulation outputs. These employ a variety of methods which can result in subhalo catalogues with significant differences (Forouhar Moreno et al., 2025). Therefore, it is important to be able to compare the results of various halo finders to help quantify the uncertainty associated with structure finding. However, along with the different structure identification methods, the halo finder codes often vary in their implementation of halo/galaxy property calculations and may even have different definitions (e.g. using inclusive/exclusive bound mass) for the same property. This can lead to further differences in the resulting catalogues, although in this case it is not due to the halo finding method itself. SOAP can take input from multiple halo finders and calculate properties consistently, thereby enabling actual differences between structure-finding algorithms to be identified. Currently SOAP supports HBT-HERONS (Forouhar Moreno et al., 2025; Han et al., 2018), SubFind (Springel et al., 2001), VELOCIraptor (Elahi et al., 2019), and ROCKSTAR(Behroozi et al., 2013). Adding a new halo finder requires a script to convert the subhalo catalogue into the standard format used by SOAP; no other code changes are necessary.

The most common definition of a halo is based on spherical overdensities (SO), regions of the universe which have a much larger density than the average. The overdensity of a region is based on all the particles within it, whether bound or unbound, and is therefore not always output by halo finders. SOAP determines spherical overdensity radii by constructing expanding spheres until the target density limit is reached. It then calculates the properties of each halo using all the particles within its SO radius. SOAP also calculates properties for several other definitions of a halo: subhalo properties (using all particles bound to a subhalo), fixed physical projected apertures (using all bound particles within a projected radius); and two types of fixed physical apertures (using all/bound particles within a sphere of the same radius for all objects). These various types give users the freedom to select the most appropriate definition for their scientific use case e.g. the type of observational data they are comparing with.

Overview of Features

- SOAP can currently calculate over 250 halo and galaxy properties. Users can easily add new properties to tailor the tool to their specific scientific needs. When combined with the four different halo definitions, this makes SOAP exceptionally versatile.
- SOAP is compatible with both dark matter-only (DMO) and full hydrodynamic simulations. For DMO runs, any properties which are irrelevant (e.g. gas mass) are automatically excluded, requiring no changes to the parameter file.
- SOAP makes it easy to enable or disable specific halo definitions and properties using the SOAP parameter file. This is possible because all properties are lazily defined within the code and are only computed if required. Additionally, if certain objects require further analysis, SOAP can be run on a subset of subhalos.
- Properties can be assigned filters so that they are only calculated for objects that meet certain criteria (e.g. only calculate the halo concentration if a subhalo has a minimum number of bound particles of a particular type). This improves the runtime of SOAP and also reduces the data volume of the final output catalogues.
- SOAP was originally written to run on Swift simulation snapshots (Schaller et al., 2024), utilising their metadata for unit handling and spatial sorting to enable efficient loading of the data. However, it has also been used to create halo catalogues from the EAGLE simulation (Schaye et al., 2015) snapshots (which use a modified GADGET format (Springel, 2005)). Supporting additional snapshot formats requires a conversion script to be written.

- The output is saved as an HDF5 file which is spatially sorted, enabling quick loading of simulation subvolumes for analysis without requiring the entire dataset.
- The catalogues can be read with the swiftsimio package (Borrow & Borrisov, 2020), which provides unit conversion (including handling comoving vs. physical coordinates) and a number of visualisation tools. All datasets are output in units that are *h*-free.
- When provided with a parameter file, `SOAP` can automatically generate a corresponding PDF document with a detailed description of all the output properties. This ensures that the documentation of the generated catalogues (e.g., the property names, units, compression level, etc.) always reflects the specific setup of the current `SOAP` run.

Acknowledgements

We gratefully acknowledge contributions to the code from Joey Braspenning, Jeger Broxterman, Evgenii Chaikin, Camila Correa, Victor Forouhar Moreno, and Roi Kugel. `SOAP` relies heavily on the following packages: mpi4py (Dalcín et al., 2005), NumPy (Harris et al., 2020), H5py (Collette, 2013), and unyt (Goldbaum et al., 2018).

References

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. (2013). The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophysical Journal*, *762*(2), 109. https://doi.org/10.1088/0004-637X/762/2/109

Borrow, J., & Borrisov, A. (2020). Swiftsimio: A python library for reading SWIFT data. *Journal of Open Source Software*, *5*(52), 2430. https://doi.org/10.21105/joss.02430

Collette, A. (2013). *Python and HDF5*. O'Reilly.

Dalcín, L., Paz, R., & Storti, M. (2005). MPI for python. *Journal of Parallel and Distributed Computing*, *65*(9), 1108–1115. https://doi.org/10.1016/j.jpdc.2005.03.010

Dome, T. (2023). CosmicProfiles: A Python package for radial profiling of finitely sampled dark matter halos and galaxies. *Journal of Open Source Software*, *8*(85), 5008. https://doi.org/10.21105/joss.05008

Elahi, P. J., Cañas, R., Poulton, R. J. J., Tobar, R. J., Willis, J. S., Lagos, C. del P., Power, C., & Robotham, A. S. G. (2019). Hunting for galaxies and halos in simulations with VELOCIraptor. *Publications of the Astronomical Society of Australia*, *36*, e021. https://doi.org/10.1017/pasa.2019.12

Forouhar Moreno, V. J., Helly, J., McGibbon, R., Schaye, J., Schaller, M., Han, J., & Kugel, R. (2025). Assessing subhalo finders in cosmological hydrodynamical simulations. *arXiv e-Prints*, arXiv:2502.06932. https://arxiv.org/abs/2502.06932

Goldbaum, N. J., ZuHone, J. A., Turk, M. J., Kowalik, K., & Rosen, A. L. (2018). Unyt: Handle, manipulate, and convert data with units in python. *Journal of Open Source Software*, *3*(28), 809. https://doi.org/10.21105/joss.00809

Han, J., Cole, S., Frenk, C. S., Benitez-Llambay, A., & Helly, J. (2018). HBT+: an improved code for finding subhaloes and building merger trees in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, *474*(1), 604–617. https://doi.org/10.1093/mnras/stx2792

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Kugel, R., Schaye, J., Schaller, M., Helly, J. C., Braspenning, J., Elbers, W., Frenk, C. S., McCarthy, I. G., Kwan, J., Salcido, J., van Daalen, M. P., Vandenbroucke, B., Bahé, Y. M., Borrow, J., Chaikin, E., Huško, F., Jenkins, A., Lacey, C. G., Nobels, F. S. J., & Vernon, I. (2023). FLAMINGO: calibrating large cosmological hydrodynamical simulations with machine learning. *Monthly Notices of the Royal Astronomical Society*, *526*(4), 6103–6127. https://doi.org/10.1093/mnras/stad2540

Narayanan et al. (2023). CAESAR. In *GitHub repository*. https://github.com/dnarayanan/caesar; GitHub.

Pontzen, A., & Tremmel, M. (2018). TANGOS: The Agile Numerical Galaxy Organization System. *The Astrophysical Journal Supplement Series*, *237*(2), 23. https://doi.org/10.3847/1538-4365/aac832

Schaller, M., Borrow, J., Draper, P. W., Ivkovic, M., McAlpine, S., Vandenbroucke, B., Bahé, Y., Chaikin, E., Chalk, A. B. G., Chan, T. K., Correa, C., van Daalen, M., Elbers, W., Gonnet, P., Hausammann, L., Helly, J., Huško, F., Kegerreis, J. A., Nobels, F. S. J., … Xiang, Z. (2024). SWIFT: A modern highly-parallel gravity and smoothed particle hydrodynamics solver for astrophysical and cosmological applications. *Monthly Notices of the Royal Astronomical Society*, *530*(2), 2378–2419. https://doi.org/10.1093/mnras/stae922

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., Baes, M., Booth, C. M., Camps, P., Navarro, J. F., Qu, Y., Rahmati, A., … Trayford, J. (2015). The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Monthly Notices of the Royal Astronomical Society*, *446*(1), 521–554. https://doi.org/10.1093/mnras/stu2058

Schaye, J., Kugel, R., Schaller, M., Helly, J. C., Braspenning, J., Elbers, W., McCarthy, I. G., van Daalen, M. P., Vandenbroucke, B., Frenk, C. S., Kwan, J., Salcido, J., Bahé, Y. M., Borrow, J., Chaikin, E., Hahn, O., Huško, F., Jenkins, A., Lacey, C. G., & Nobels, F. S. J. (2023). The FLAMINGO project: cosmological hydrodynamical simulations for large-scale structure and galaxy cluster surveys. *Monthly Notices of the Royal Astronomical Society*, *526*(4), 4978–5020. https://doi.org/10.1093/mnras/stad2419

Springel, V. (2005). The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society*, *364*(4), 1105–1134. https://doi.org/10.1111/j.1365-2966.2005.09655.x

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. (2001). Populating a cluster of galaxies - I. Results at z=0. *Monthly Notices of the Royal Astronomical Society*, *328*(3), 726–750. https://doi.org/10.1046/j.1365-8711.2001.04912.x