The Journal of Open Source Software

CoReCon: an open, community-powered collection of Reionization constraints

Enrico Garaldi[1]

1 Post-doctoral Fellow, Max-Planck-Institüt für Astrophysik

DOI: 10.21105/joss.02495

Software

- Review
- Repository
- Archive

Editor: Eloisa Bentivegna

Reviewers:

- @ngnedin
- @afialkov
- @Stefan-Heimersheim

Submitted: 07 July 2020
Published: 20 July 2020

Summary

The Epoch of Reionization (EoR) is the last global phase transition in the history of the Universe, and it represent the current fronteer in the study of galaxy formation, as it radically altered the environment in which cosmic structures formed and evolved thereafter. It transformed the intergalactic medium (IGM) that fills the space between galaxies from a cold and neutral gas to a hot and ionized one. Despite its importance, little is known about this period of time, mostly as a consequence of the intrinsic difficulties in observing such a distant epoch. The latter is the main reason why a plethora of different methods have been devised to extract information from the limited observations available. However, these data are typically scattered in many different publications, using inhomogeneous unit systems, and sampling strategies (e.g. volume- or mass-averaged quantities in the intergalactic medium). Hence, employing these data in a scientifically-sound way often requires (i) retrieving the aforementioned information from different publications, and (ii) homogenise them. Moreover, it is often challenging to grasp a complete picture of available constraints for a given physical quantity. The situation is made worse by the lack of any systematic collection of data. Additionally, constraints on physical quantities often need to be retrieved from published *plots*, as they are not explicitly reported in the publication itself, a tedious and error-prone practice. While in principle the latter limitation can be overcome by directly contacting the authors of the publication, the extreme mobility faced by researchers entails that contact information are very often outdated at best, while frequently the authors have simply left the field.

We tackle these issues through a python module named `CoReCon` (acronym for Collection of Reionization Constraint). The goal of `CoReCon` is twofold. First, it comprises a growing set of constraints on key physical quantities related to the EoR, homogenised in their format and units, lifting the busy researchers from the weight of searching, retrieving and formatting data. Second, and foremost, `CoReCon` provides a platform for the reionization and high-redshift research communities to collect and store, in an open way, such observational constraints. We do so by providing a Python infrastructure, which is able to load formatted data files and provides simple utility functions to deal with such data. The data files loaded by `CoReCon` are purposedly simple in their form and as complete as possible in their content, in order to collect all the relevant information in one place. Notably, they are required to contain a URL to the original publication and a short description of the methods used to retrieve the constraints from observed data. They are also allowed to contain any additional, unplanned-for data field, in order to reach the highest degree of flexibility and to allow the storage of all relevant information. In our view, once a new constraint is published, the author of the publication will update `CoReCon` with the relevant data. If this procedure becomes customary, `CoReCon` will serve as an up-to-date repository of easy-to-retrieve constraints.

To our knowledge, this is the first module of its kind - at least in the EoR community. With a similar spirit, there exist a collection of all the known quasars above a redshift of 5.7 (Bosman (2020)). Additionally, an effort toward openness of research in the EoR field recently

materialized into an open analysis pipeline for the reduction of spectra taken in most of the major telescopes in the world (Prochaska et al. (2020)).

Features

`CoReCon` is written as a Python module in order to provide portability, ease of installation and use, and to reach the large community of researcher using Python. Additionally, we put effort into building a template for entering new data to the module, which strives to be simultaneously easy to fill and complete.

The `CoReCon` module is able to read two different data layout, and internally transform them into the frontend data format exposed to the user. This choice is dictated to reduce the workload while entering data with different features. In particular, one of the two data format is devised to minimize the redundancy of information to be inserted by the researcher when the independent variables have a grid-like structure (potentially with holes).

The module also includes simple utility functions that can transform the available data in commonly-used ways. For instance, selecting only the constraint on a specific physical quantity, in a user-defined value range, or transforming their layout to be ready-to-plot using the matplotlib Python module.

`CoReCon` has been developed with openness in mind. For this reason, new constraints can be easily added by simply filling a form provided, and copying it into the directory tree of the module. Data entries are required to contain the reference and a link to the original publication, in order to ensure the original publication is acknowledged, a *quality* flag which specify if the data were explicitly available in the publication or has been retrieved in some indirect way (e.g. from a published plot, hence potentially introducing errors), and a short description of the constraints themselves and of the method employed to measure/compute them.

At the time of writing, `CoReCon` contains data for the following physical quantities: ionised fraction, IGM temperature at mean density, effective optical depth of the HI and HeII Lyman-α forest, flux power spectrum of the Lyman-α forest, cosmic microwave background optical depth, galaxy and quasar UV luminosity functions, column density ratio, mean free path of ionizing photons, star-formation-rate density, and correlation between the flux in Lyman-alpha spikes and galaxy position.

The `CoReCon` module can be easily installed via `pip` and is fully documented online.

Acknowledgements

We acknowledge in advance the community that – we are sure – will help making `CoReCon` a complete and valuable module. We are grateful to Benedetta Ciardi and Martin Glatzle for their comment on previous version of this manuscript. We are thankful to the community developing and maintaining the `numpy` (Walt, Colbert, & Varoquaux (2011)) software package, upon which `CoReCon` is built, and `matplotlib` (Hunter (2007)), employed for the data visualization in the `CoReCon` documentation.

References

Bosman, S. E. I. (2020, June). All z>5.7 quasars currently known. Zenodo. doi:10.5281/zenodo.3909340

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, *9*(3), 90–95.

Prochaska, J. X., Hennawi, J. F., Westfall, K. B., Cooke, R. J., Wang, F., Hsyu, T., & Farina, E. P. (2020). PypeIt: The Python Spectroscopic Data Reduction Pipeline. *arXiv e-prints*, arXiv:2005.06505. Retrieved from http://arxiv.org/abs/2005.06505

Walt, S. van der, Colbert, S. C., & Varoquaux, G. (2011). The numpy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22–30.