



ID de Contribution: 82

Type: Oral presentation

## Exploring high dimensional datasets for an enhanced interactivity and reproducibility: the case of the Euclid Morphology Challenge

*mercredi 23 mars 2022 12:15 (10 minutes)*

With the new generations of surveys, in astrophysics and cosmology, datasets available to the community become more complex. It is often stated that surveys such as Euclid or LSST will generate a volume of data which needs new methods of analysis. The increase in data volume generally implies more complexity which is translated into an increase of the dimensionality of the parameter space to be explored. With the scientific toolbox becoming more complex, we can extract from those huge datasets an increasing number of parameters, in different filters, at different redshifts or SNRs, coming from different instruments or telescopes. The dimensionality in which the important scientific information lies can become quickly overwhelming.

In this scenario, any researcher can face what we called the « curse of plotting », where the amount of plots to explore all the interesting behavior of a dataset become too large to fit into a scientific paper with a reasonable size. Indeed, a figure being limited to 2D, sometimes 3D plots, the number of combinations between the different parameters of interests need to be restricted, sometimes in spite of scientifically meaningful and important (cor)relations. In addition, some of the extracted data are often a function of some hyper parameters or threshold, that one must fix for the particular context of a paper, sometimes arbitrary and which is not adapted to different science cases. One can try to explore the impact of changing this hyper parameter to the result of the study, but here again, the freedom in the exploration of the parameter space remains limited.

To tackle this curse of plotting, we are developing an interactive interface using the Streamlit python package, which allows the reader of our paper analyzing the Euclid Morphology Challenge to explore in a more extended way the parameter space presented in the paper. With a « user friendly » (and developer friendly) interface, one can reproduce all the figures of the paper, but also go significantly further.

For example, we plot in the paper some metrics as a function of the galaxy magnitude. In the interface, one can plot the same metrics but as a function of radius, redshift, or other physical parameters. The range of parameters and number of bins are also let free, for example to limit the study on nearby galaxies.

Another example is that those metrics are dependent on a threshold defining what is considered as a bad fit. One can continuously change this threshold and see the impact on his/her science. We also present 3D plots, where the x and y parameters can be changed to explore correlations.

In this talk, we propose to show and discuss the power of this tool, and the ease of its implementation. If we are talking about a specific case regarding a paper, one can adapt this idea to any high dimensional dataset, which we think is very important for any scientific paper, both for reproducibility but also completeness of the study. Developing this kind of tools would be a major interest for the next generation surveys, to fully embrace their complexity and amount of information.

### Field

Not in the above

### Day constraints

**Author:** BRETONNIÈRE, Hubert (Institut d'Astrophysique Spatiale)

**Orateur:** BRETONNIÈRE, Hubert (Institut d'Astrophysique Spatiale)

**Classification de Session:** Talk

**Classification de thématique:** Astrophysics