Machine learning within the THREEHUNDERD simulation project

Daniel De Andres*[1], Gustavo Yepes[2], Weiguang Cui[3], Marco De Petris[4], and Florian Ruppin[5]

[1]Universidad Autonoma de Madrid – Spain
[2]Universidad Autonoma de Madrid – Spain
[3]Institute for Astronomy [Edinburgh] – United Kingdom
[4]Department of Physics, Sapienza Universit'a di Roma – Italy
[5]MIT Kavli Institute for Astrophysics and Space Research – United States

Abstract

The THREEHUNDRED project aims at collecting one of the largest set of hydrodynamical simulations of massive galaxy clusters. Since massive clusters are very scarce, the zooming technique is used to individually resimulate spherical regions around the 324 most massive objects found in the MULTIDARK, 1 giga-parsec volume N-body only simulation. From the results of these simulations we train a machine learning algorithm to find a mapping between dark matter halo properties derived by Rockstar halo finder and baryon properties, such us gas mass , stellar mass and the Compton Y parameter within R500 obtained from the hydrodynamical simulations. By doing so, we successfully populate the whole MultiDark cluster-size halos with baryonic properties similar to those from the THREEHUNDRED hydro simulations. We also prove that the trained models can also be applied to other N-body simulations even when the mass resolution is a factor 8 lower than the one used for training. Therefore, we can use these machine learning models to populate halos in much larger computational volumes so they can be used to produce mock light cone observations of full sky Xray surveys like e-ROSITA.
We have also collected a catalogue of almost 200,000 mock SZ maps from the THREEHUNDRED simulations with the same angular resolution and noise as the clusters observed by the Planck satellite. We then applied deep learning techniques to compute the mass of the real full-sky Planck maps by training our algorithms on the simulated mock images. We show that the deep learning mass estimate is compatible with the Planck mass derived by applying SZ scaling relation within a 20% bias, which is the expected bias assuming the hypothesis of hydrostatic equilibrium holds. The advantage of deep learning methods is that they do not rely on any assumption on the gas profiles to derive masses. The only requirement is that the training set is accurate enough to reproduce the physical and observing conditions of the objects of interest.

Slides: in PDF
Video: https://youtu.be/tytpmpp1978

Keywords: galaxies, cluster, mock, observation, deep learning, simulations, cosmology

*Speaker