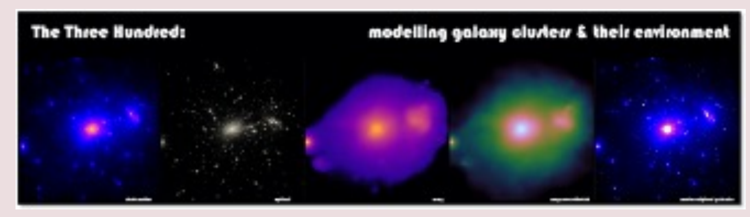
Inference of galaxy clusters mass radial profiles from y-Compton maps with Deep Learning techniques.

A. Sbriglio[1], W. Cui[2,3], D. De Andrès[2], M. De Petris[1], A. Ferragamo[1], G. Yepés[2]

[1]Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro, 5-00185 Roma, Italy
[2]Departamento de Fisica Teorica and CIAFF,Modulo 8 Universidad Autonoma de Madrid, 28049, Madrid, Spain
[3]Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK

MACHINE LEARNING FOR ASTROPHYSICS
CATANIA, 30 MAY - 1 JUNE, 2022

SAPIENZA UNIVERSITÀ DI ROMA



Abstract: Clusters of galaxies are the largest gravitationally bound systems in the Universe resulting from the natural evolution of cosmic structures. They are crucial tracers of the structure formation history and their mass function at different epochs is of key importance to constraining cosmological parameters. Therefore, it is essential to infer the mass of the observed clusters, which unfortunately is not directly observable and is affected by different biases related to the applied observational estimates. To overcome these obstacles we exploit a modern method, provided by machine learning algorithms, that turn out to outperform conventional statistical methods. In previous work, Convolutional Neural Networks (CNNs) were applied on full-sky y-Compton maps, to estimate the masses of clusters defined at a fixed aperture radius corresponding to 500 times critical overdensity. We now extend this study to estimate the radial profiles of the clusters' total mass. We make use of a deep learning architecture based on Autoencoders neural network to find the most efficient and compact representation of the input data. The training of the architecture is performed on mock images of the thermal Sunyaev-Zel'dovich signal generated by a large set of hydrodynamical simulated galaxy clusters from the "The Three Hundred" project.

Scientific topic

Clusters of galaxies are a powerful way to provide cosmological information. Their abundance in the Universe, as a function of total mass and redshift, is strictly related to the mean matter density, Ω_m, and the amplitude of matter perturbations at a scale of $8h^{-1}Mpc$, σ_8. High-precision, unsystematic estimation of the total mass of these objects becomes necessary.

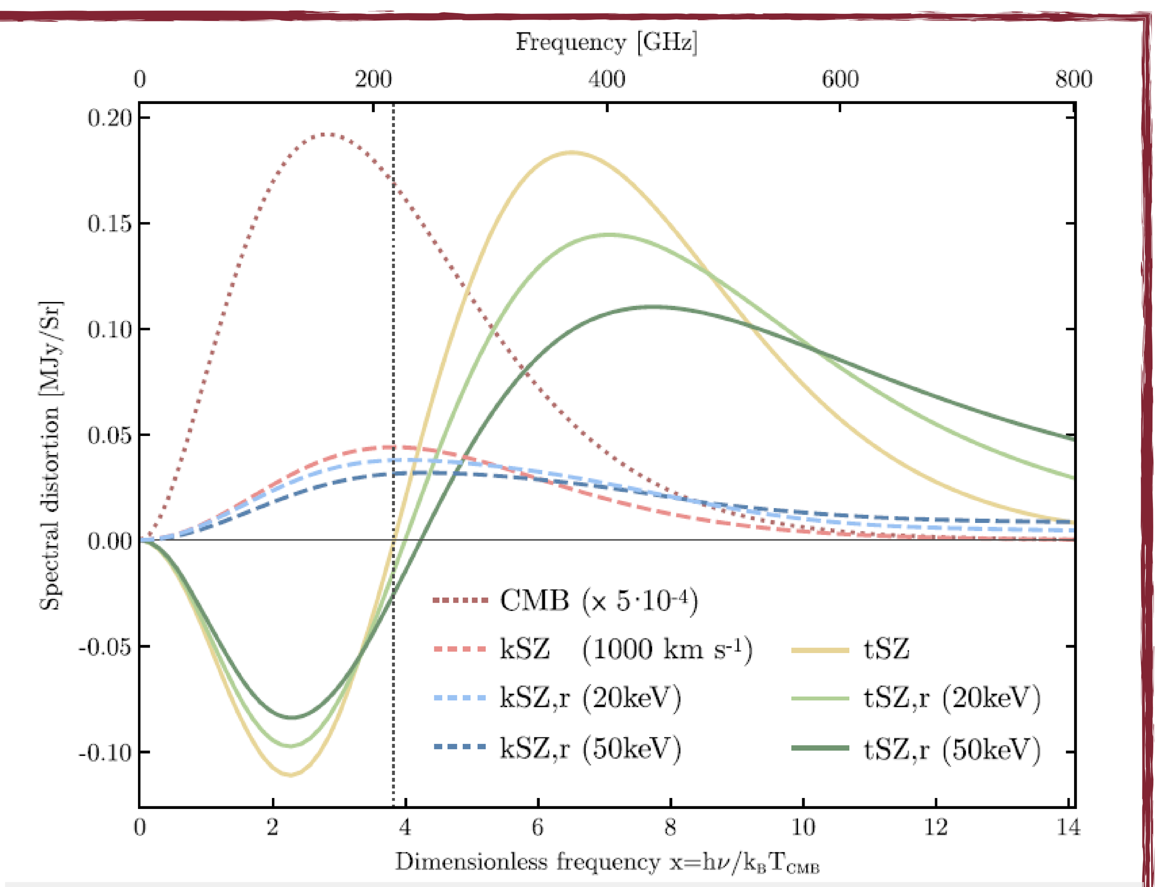


Fig.1: *Spectra of the components of the SZ signal (tSZ-solid lines, kSZ-dashed lines) with relativistic corrections) assuming different ICM temperatures. The dotted, dark red curve illustrates the shape of the unscattered CMB spectrum. [1]*

The thermal component of SZ (tSZ, **Fig.1**) is a suitable probe that infers the overall cluster mass. The tSZ observations allow us to carefully map ICM inhomogeneous distribution and detect point-like sources as contaminants.

Simulations and the *Three Hundred Project*

After recombination, the linear theoretical treatment of structure formation is no longer satisfied. Perturbative growths make it impossible to formulate a complete theoretical model: this kind of process can be treated with approximations. Numerical simulations make it possible to explore the biases that assumptions included in other methods might introduce into the final cluster mass estimate and give access to a comprehensive set of information about the ICM and its effect on CMB photons.

The300 simulation consists of a large set of cosmological hydrodynamic radiative simulations of more than 300 enlarged spherical regions of size $15h^{-1}Mpc$, centred on the most massive clusters tracked by the Amiga's halo finder (AHF) [3] on Multidark Planck 2, a $1h^{-1}Gpc$ volume simulation (**Fig.2**). From the results of these simulations with multiple line-of-sight projections, 2D maps of the mock tSZ signal were produced. The total number of images used to train the deep learning algorithms is ~ 20000

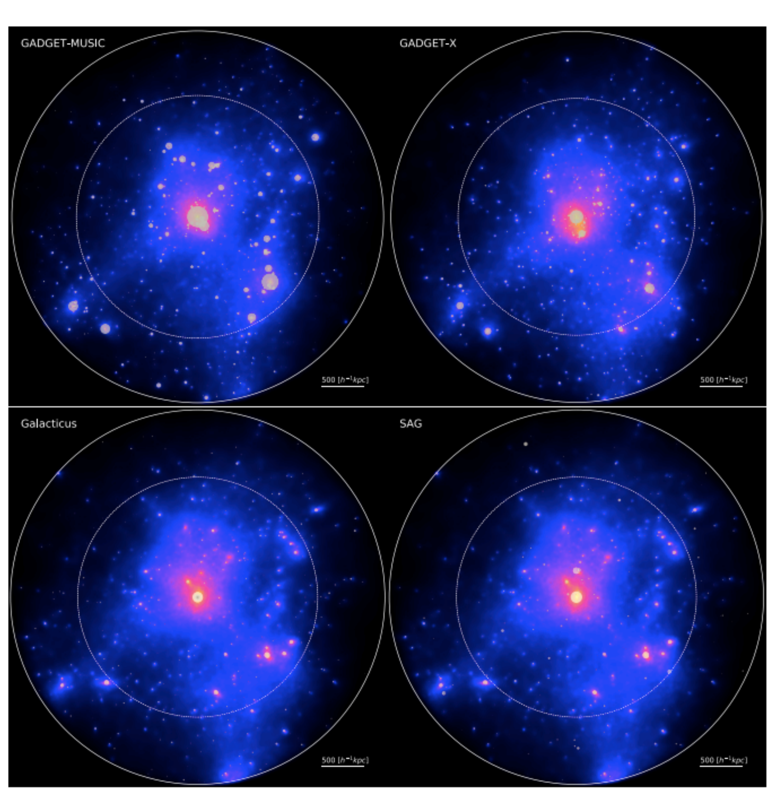


Fig.2: *The distribution of galaxies, ICM and DM within R200 of the most massive cluster in a re-simulated region with different methods.[4]*

Results

The inferred cluster mass radial profiles show median values close ($\sim 1\,\%$) to the real ones with a scatter of $\sim \pm 10\,\%$ (**Fig.5**). Segregation of the cluster populations into relaxed and disturbed objects showed that the algorithm is influenced by the dynamic state of the clusters (**Fig.6**).

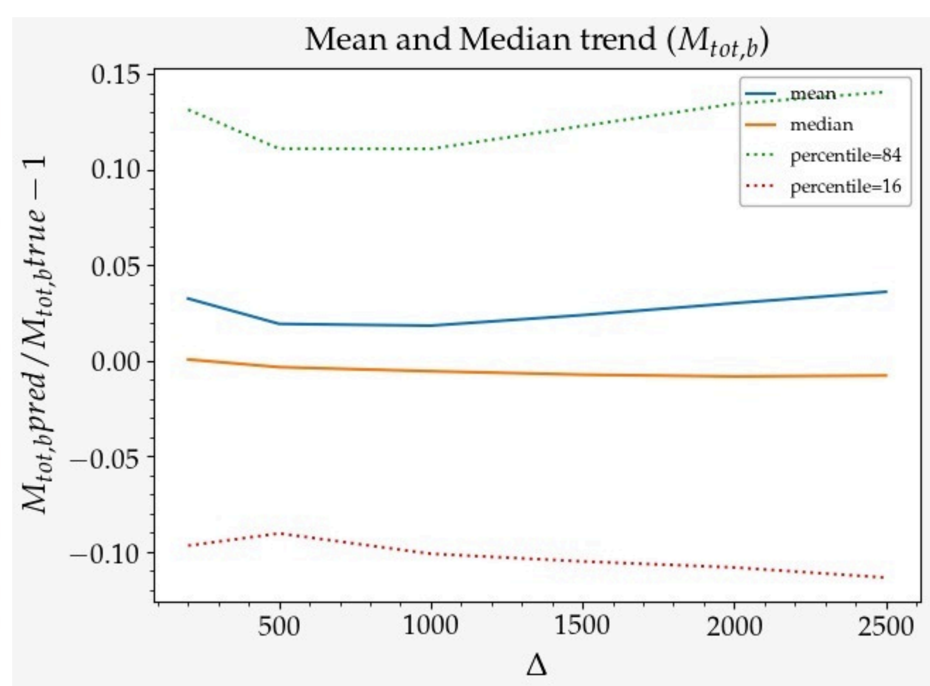


Fig.5: *Bias (mean and median values) with scatters (16% and 84% percentiles) of the inferred mass radial profiles at different clusters apertures sampled in terms of* $\Delta = \rho_{cl}/\rho_c$.

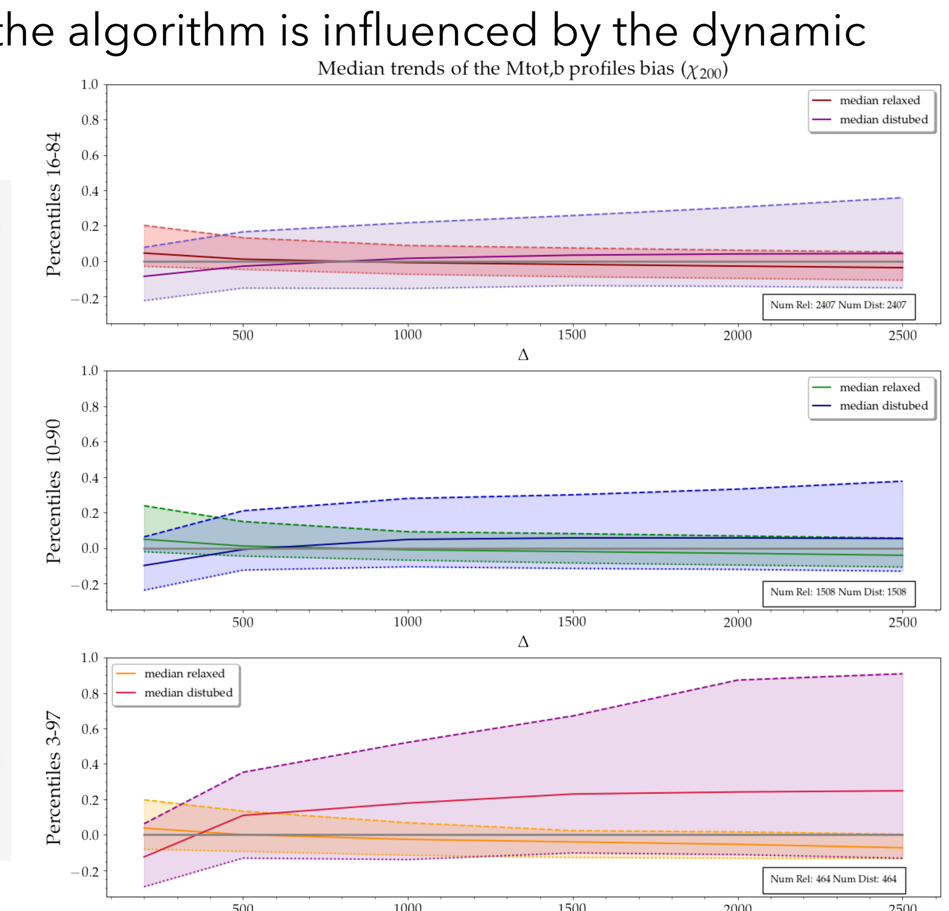


Fig.6: *Comparison of median and percentiles of bias (16% and 84%) for the relaxed and disturbed population. Segregation was performed by taking the most extreme tails of the cluster dynamic state parameter, in three different cases, with increasingly extreme selections.*

We verify that the Machine Learning inference method has a negligible bias differently from other methods, such as with the Hydrostatic Equilibrium approximation [2]. Even the scatter seems smaller than in the HE case (**Fig.7**).

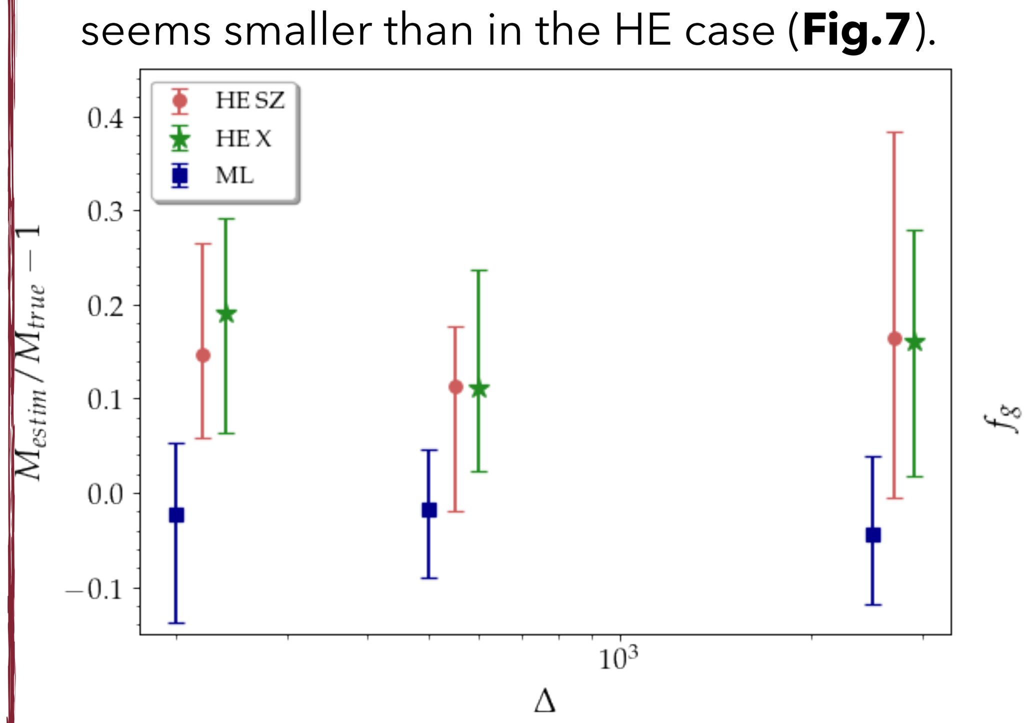


Fig.7: *The mass inferred with ML was compared with the one from HE approaches estimated in [2]. Red dots are in the HE case from SZ maps; green dots are in the HE case from X maps; blu dots are in this work case;*

The gas fraction, derived from the predicted gas and total masses, $\left(f_G = \frac{M_{gas}}{M_{tot}}\right)$, is consistent with previous observational results (**Fig.8**).

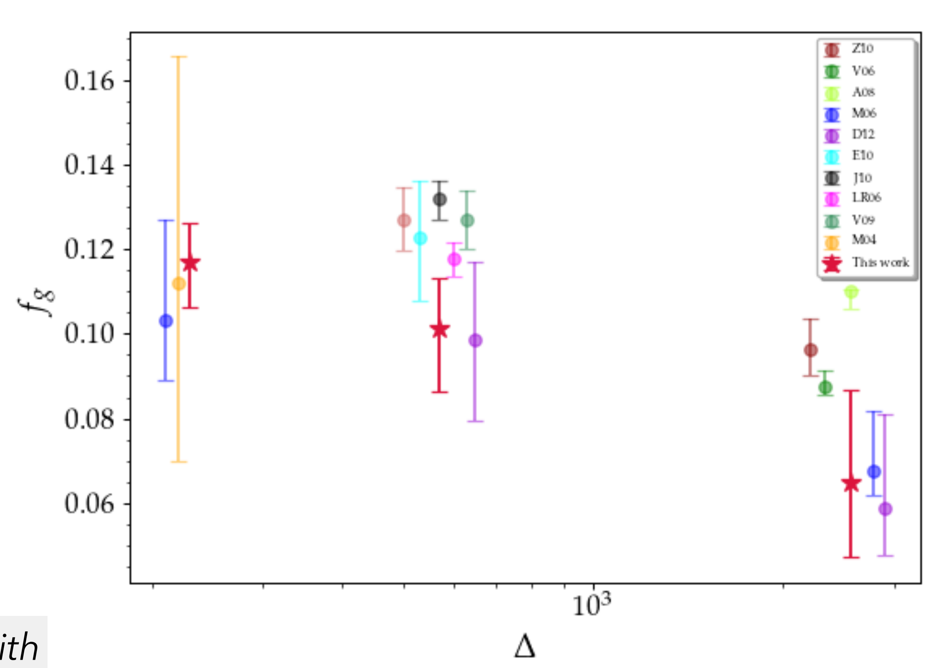


Fig.8: *Comparison of* f_g *estimated value at the 3 different overdensities, with results of previous observations [5].*

Deep Learning Approach

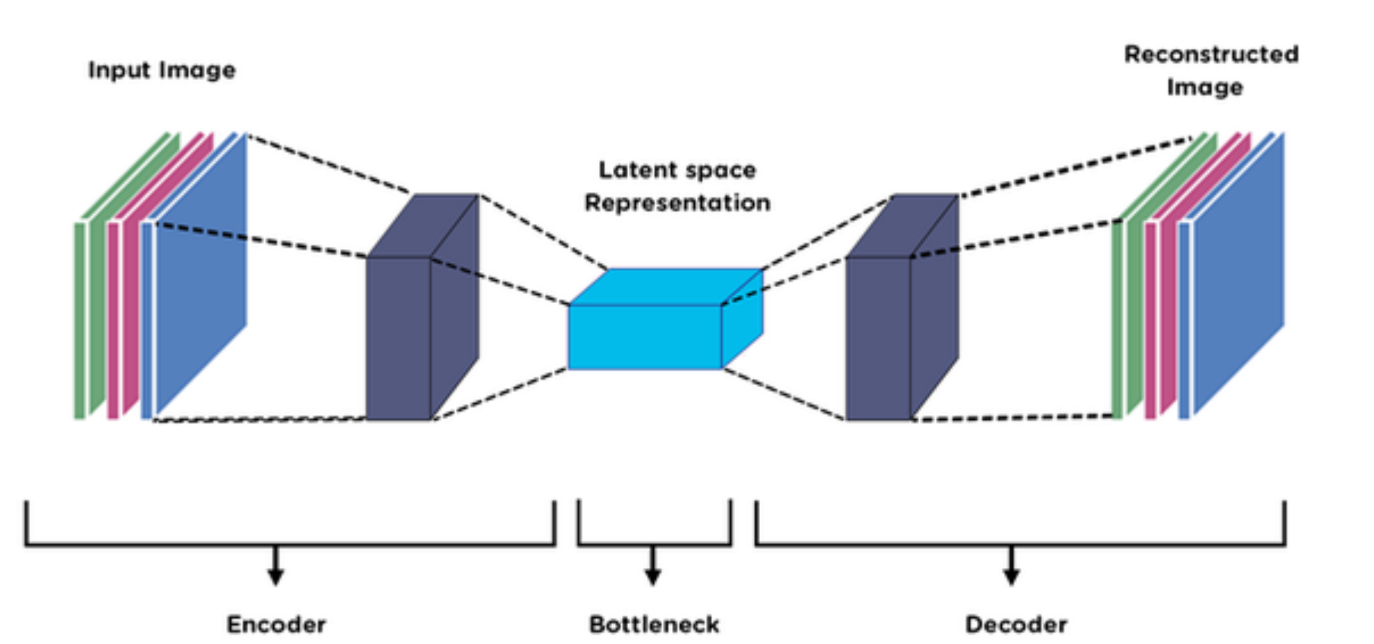


Fig.4: *Autoencoder architecture: The encoder part condenses the input (y-Compton parameter mock maps) in a Latent Vector (inside Latent Space) containing essential features for reconstructing profiles. The decoder layer produces a reconstruction of the input image.*

The Autoencoder (AE) condenses the information contained in the y-Compton parameter mock maps (**Fig.3**) into a vector representation called 'Latent Vector' (**Fig.4**). We apply a Random Forest (RF) algorithm to perform a multidimensional regression between the Latent Vector and the true values of the masses radial profiles of simulated clusters.

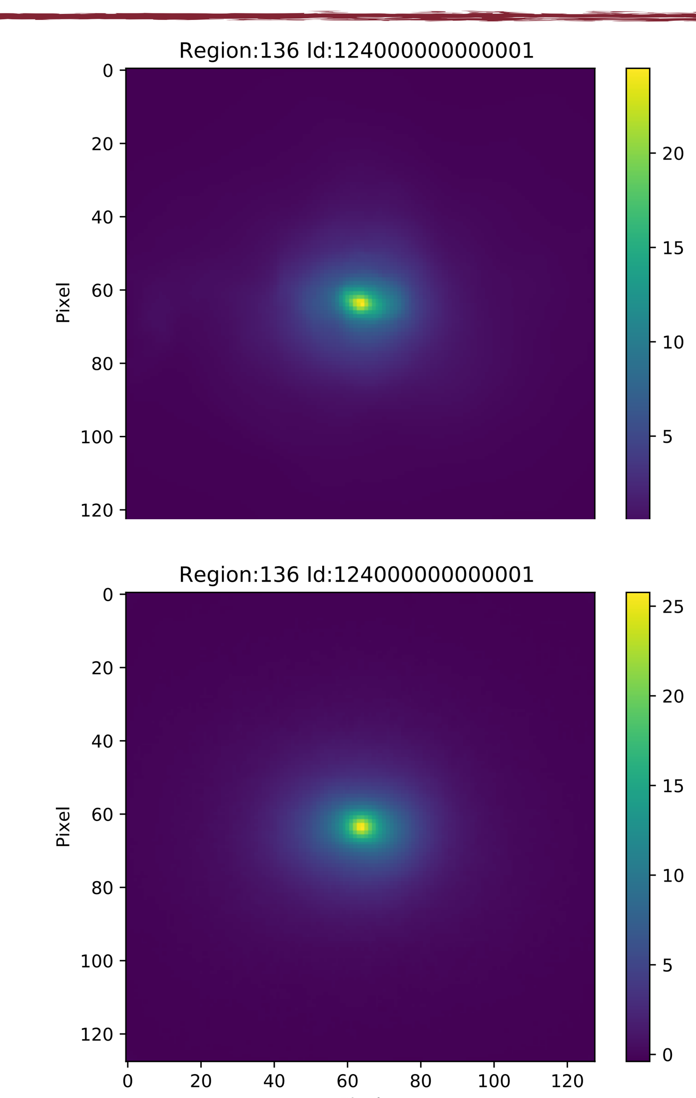


Fig.3: *The y-Compton mock maps for a cluster: original (upper side) and recovered with AE algorithm (lower side).*

References

[1]Mroczkowski T. et al., S. Sci Rev, 215, 17 (2019);
[2]Gianfagna G. et al., MNRAS **000**, 1–11 (2021), in press;
[3]Knollmann SR. et al., AJSS, 182, 608–624, (2009);
[4]Cui W. et al., MNRAS, 480, 2898 (2018);
[5]Sembolini F. et al., MNRAS 429, 323–343 (2013) ;

Useful website

Amiga's halo finder:
http://popia.ft.uam.es/AHF/Download.html

The Three Hundred:
www.nottingham.ac.uk/~ppzfrp/The300

Contact:
sbriglioalessia@gmail.com 3282290563
Viale Ippocrate 75, Roma,00161

Catania, June/30-1/2022