A Deep Learning Approach to Infer Galaxy Cluster Masses in Planck Compton parameter maps.

Daniel de Andrés

Universidad Autónoma de Madrid

daniel.deandres@uam.es

Collaborators: Machine Learning group in the THREEHUNDRED Project.

- Gustavo Yepes, Marco De Petris, Weiguang Cui, Florian Ruppin, Federico De Luca, Giulia Gianfagna, Jesús Vega Ferrero (+EURANOVA people: Mahmoud Jarraya, Gianmarco Aversano, Ichraf Lahouli, Romain Dupuis)

Introduction

- Galaxy clusters are the biggest gravitational bound objects in the universe and they can be observed through the inverse-Compton scattering of the cosmic microwave background (CMB)photons with electrons in the ICM, i.e. the Sunyaev-Zel'dovich (SZ) effect.

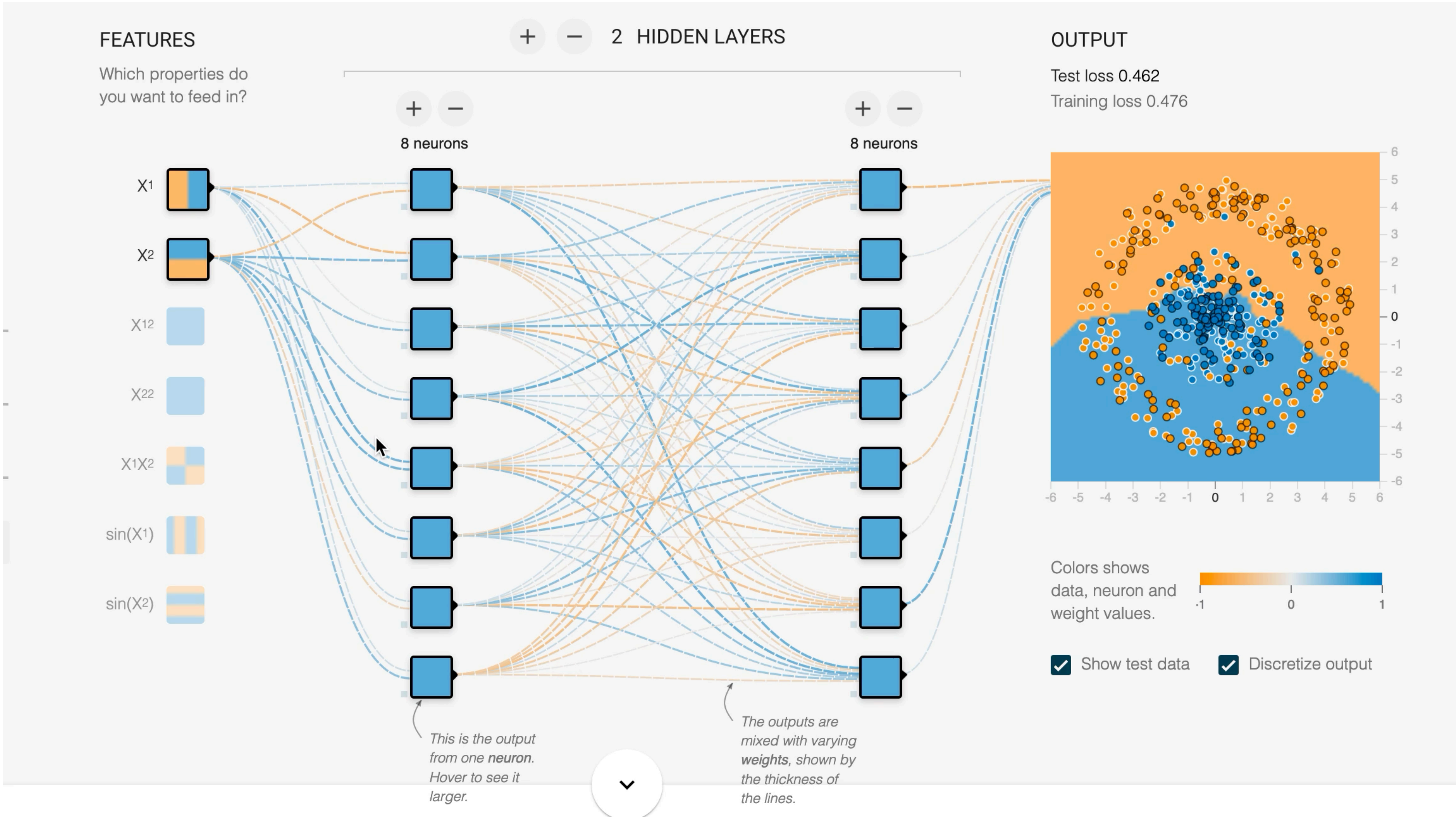
- The Planck Collaboration collected a full-sky survey of SZ galaxy cluster maps and estimated their masses through Y_{500}-M_{500} scaling relation. However, these masses are expected to be bias low due to the fact that hydrostatic equilibrium hypothesis is assumed.

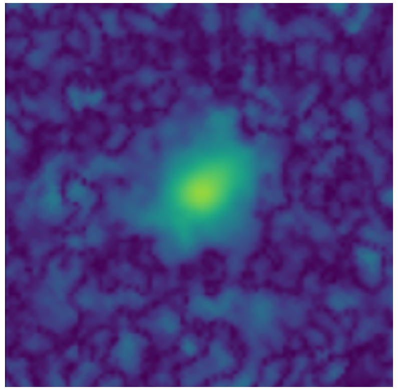
- We aim to address this issue by training a Convolutional Neural Network (CNN) on a large catalog of almost 200,000 simulated Planck-like SZ maps (with the same angular resolution and noise levels) from the 300th .

- This approach is based on finding a mapping between simulated SZ maps and the 3D dynamical mass M_{500} without assuming any apriori symmetry.

Deep learning

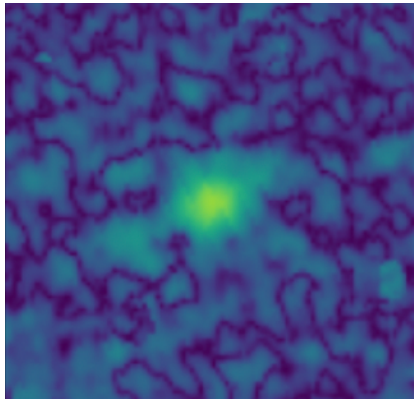
Why Neural Networks? https://playground.tensorflow.org/

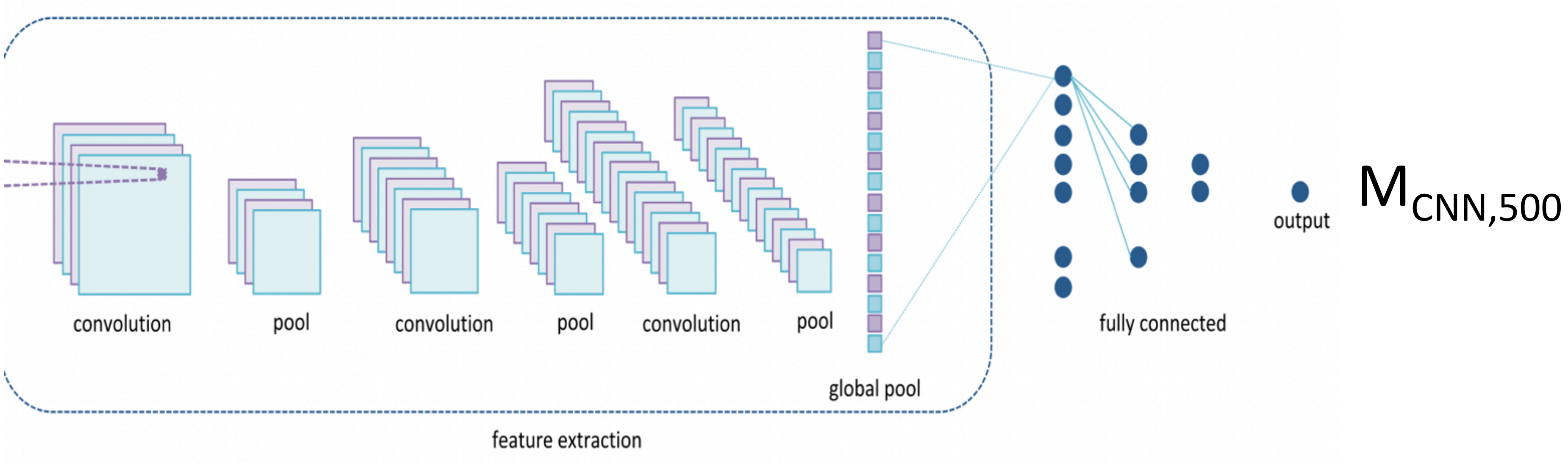




Train on mock images

Predict the mass of Planck SZ clusters





We used a version of the VGG network (Simonyan & Zisserman 2014) that has been successfully used in simulated clusters by Ntampaka et al. (2019) and Yan et al. (2020) .

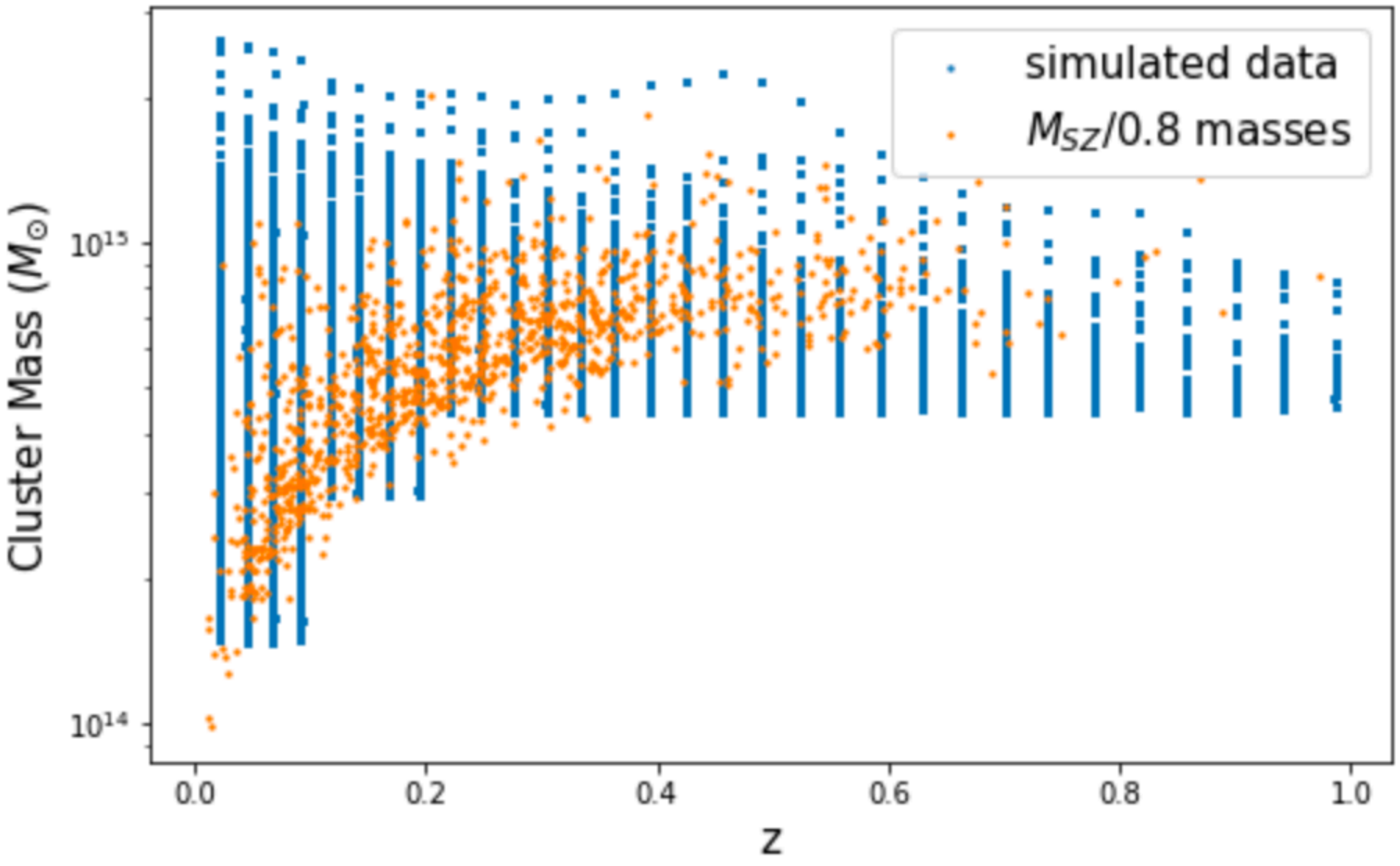
Training data set

We extract 7106 clusters from the 324 re-simulated regions. Furthermore, we take 27 different projections of every single cluster amounting to a total of 191,862 simulated maps.

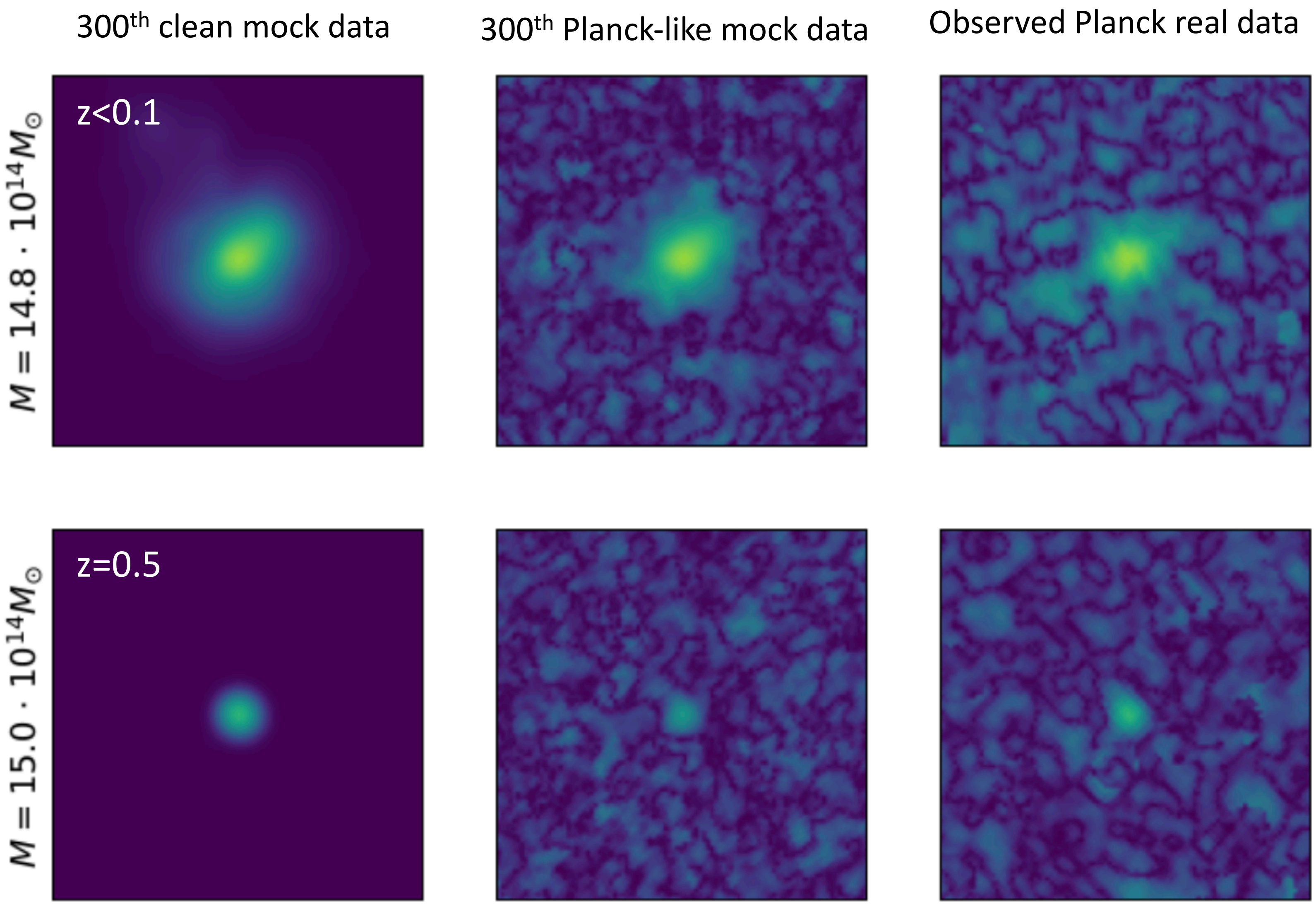
Particularly, we take this selection of clusters for covering the whole 1094 Planck sample of clusters in redshift 0<z<1 and mass $10^{14} < M_\odot < 2\times10^{15}$.

Furthermore , to improve the performance, we train four CNN models for different redshift bins:

- 0<z<0.1
- 0.1<z<0.2
- 0.2<z<0.4
- 0.4<z<1



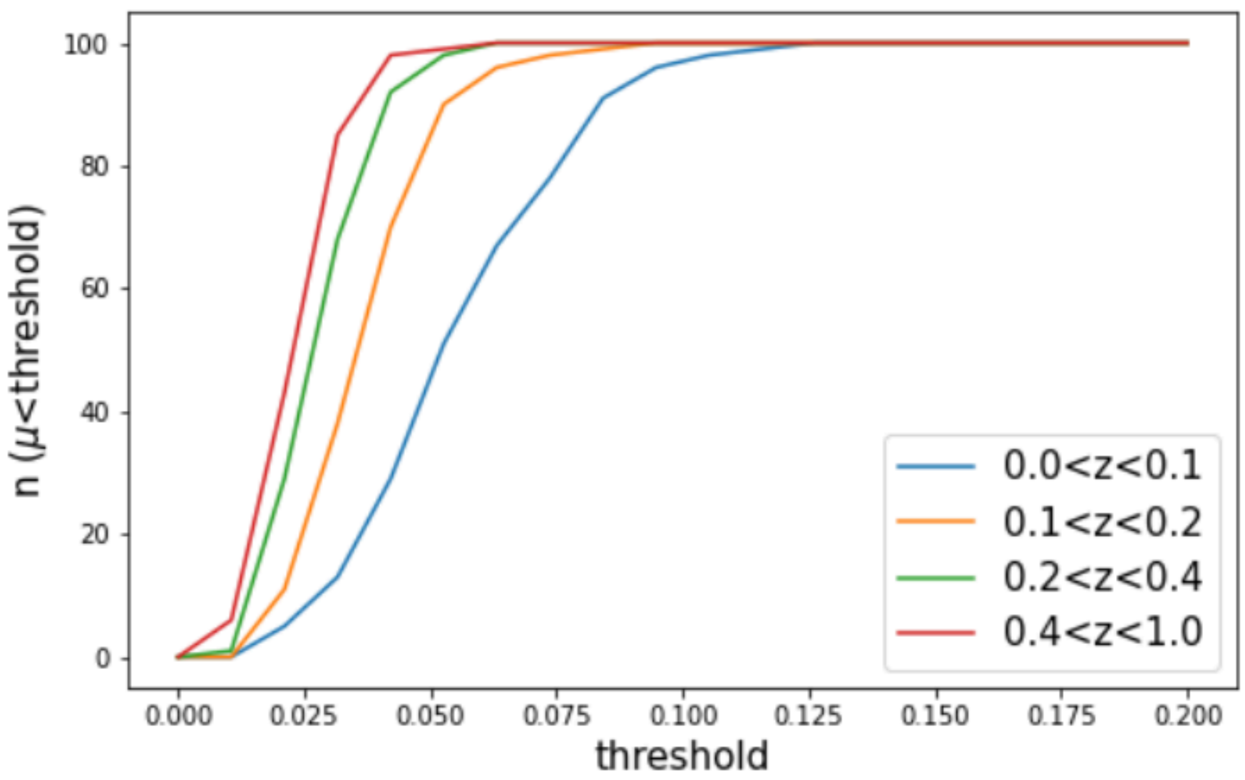
- **300^{th} clean mock data:** we simulated the SZ signal with the same angular resolution as Planck data.
- **300^{th} Planck-like mock data:** We add instrumental noise with the same power spectrum as the Planck sky.
- **Observed Planck real data:** SZ clusters measured by Planck .



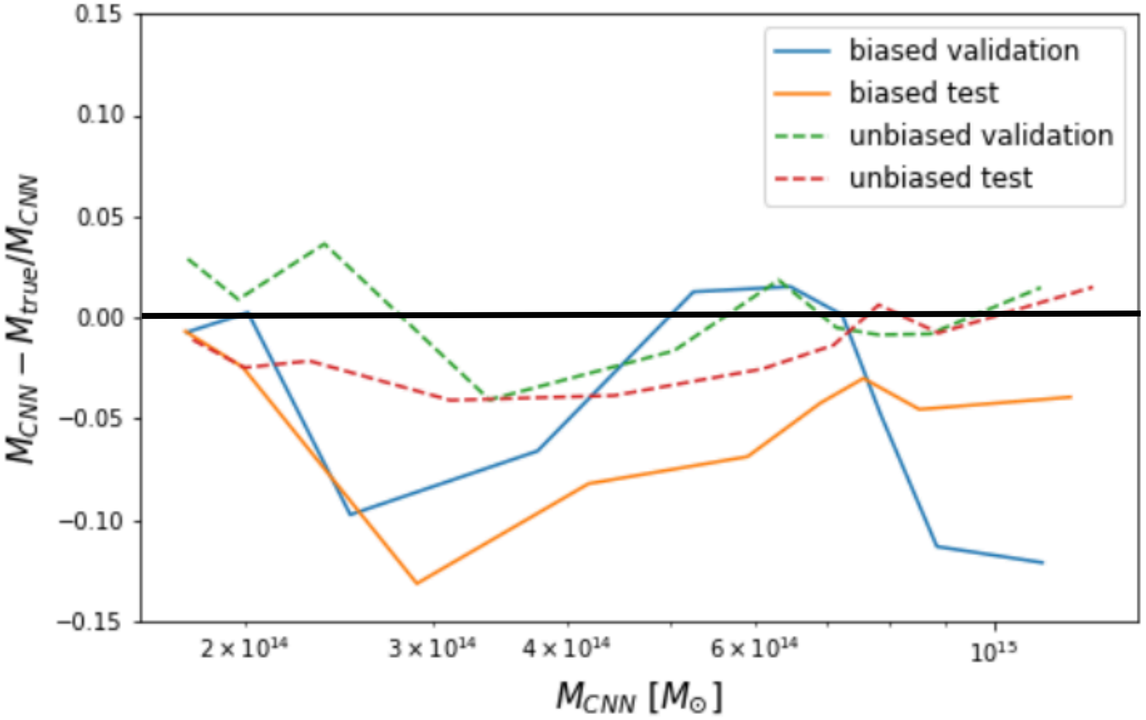
Training

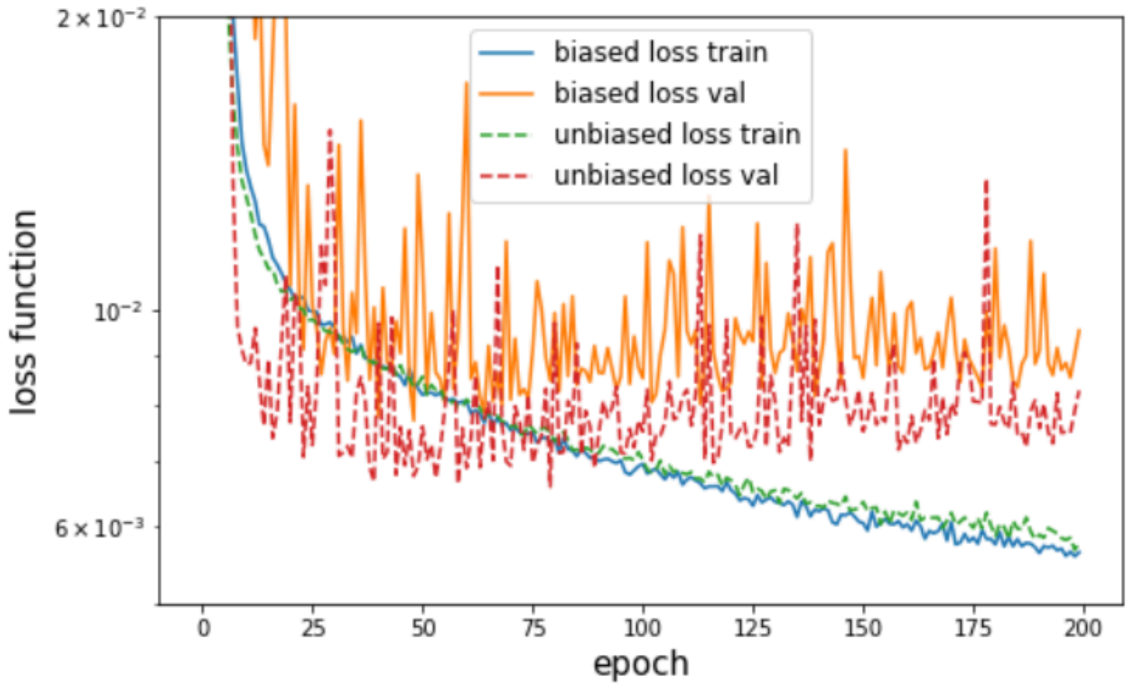
- We train 4 CNN models corresponding to different redshift bins.
- Due to the fact that the training is random, we train 100 models and discriminate the best among them through 100 runs where we split our data set in 80% training, 10% validation and 10% test.
- Validation set it is used for identifying the best model. Test set is used for showing the final results.
- We focus on obtaining an unbiased model with respect to the 3D dynamical mass.
- Loss function is the logarithm mean squared error

$$\mathcal{L} = \frac{1}{n}\sum_{n}(\log M_{\text{true}} - \log M_{\text{CNN}})^2$$



We show the number of models whose bin<threshold for all bins. Different colors represent the different redshifts intervals



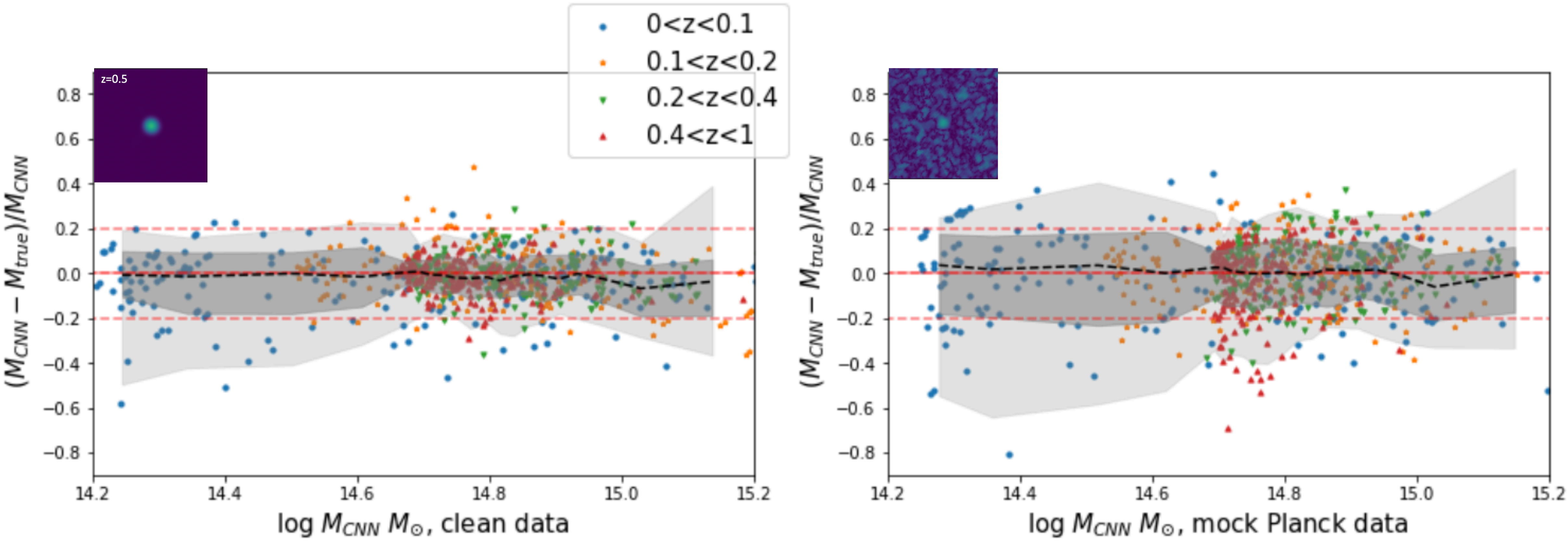


We show the mean value of the relative error as a function of the predicted mass M_{CNN}. The continuous lines represent a selected biased model while the dashed lines represent a selected unbiased model.

We show the loss function as a function of the number of epochs for the biased model and the unbiased model. Note that the performance is similar for the training set. However, the biased model's performance is slightly worse on the validation set.

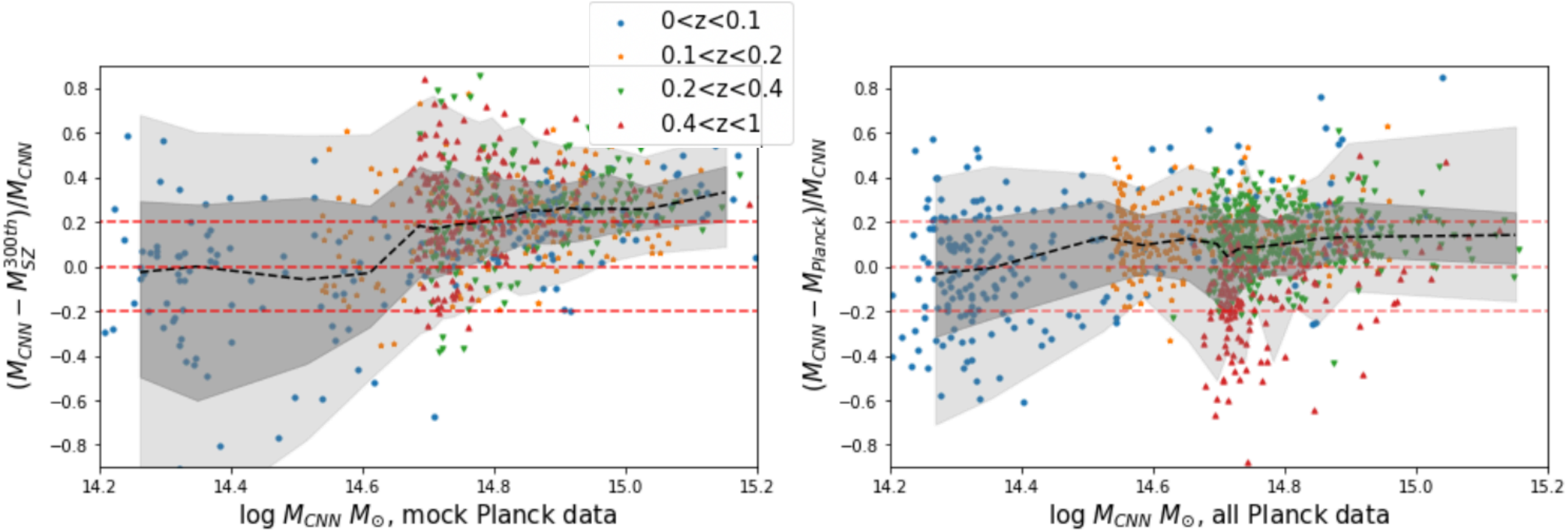
Results

Results on simulations



We show the median (dashed black) and the 68% and 95% percentile intervals (shaded regions). As required in the validation procedure, the predictions are unbiased with respect to the 3D dynamical mass. The points plotted are a random sample of 200 points but the statistics is compute using all the simulated clusters.

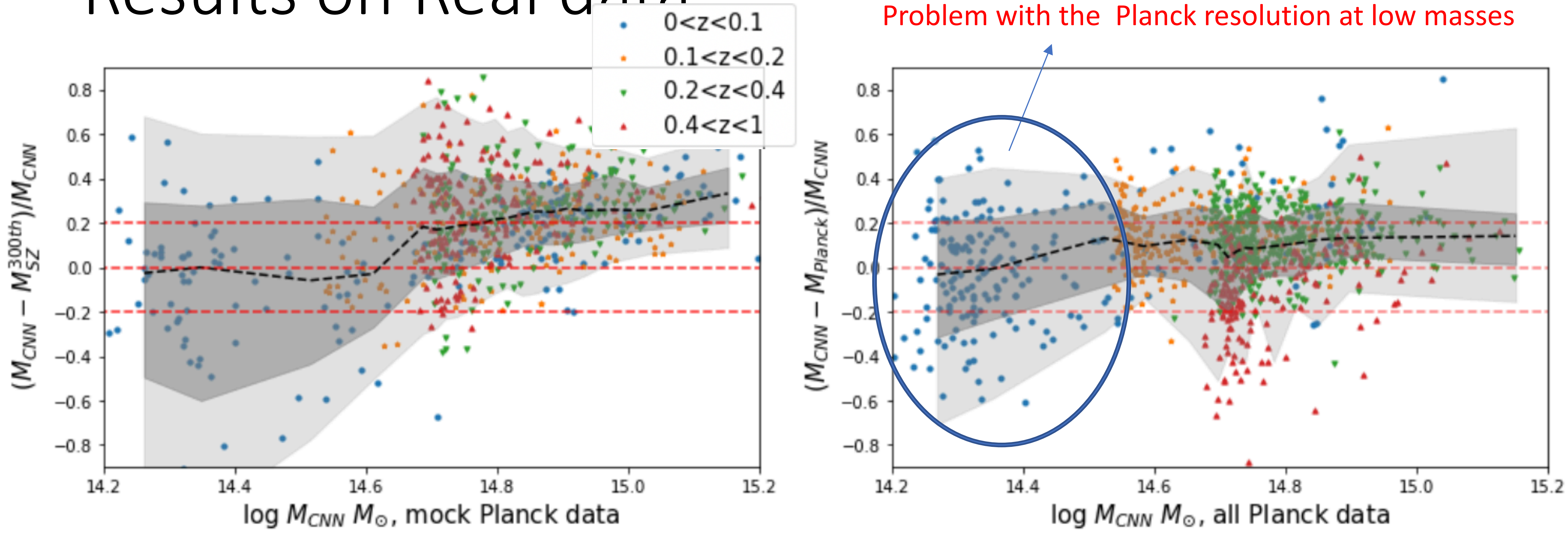
Results on Real data



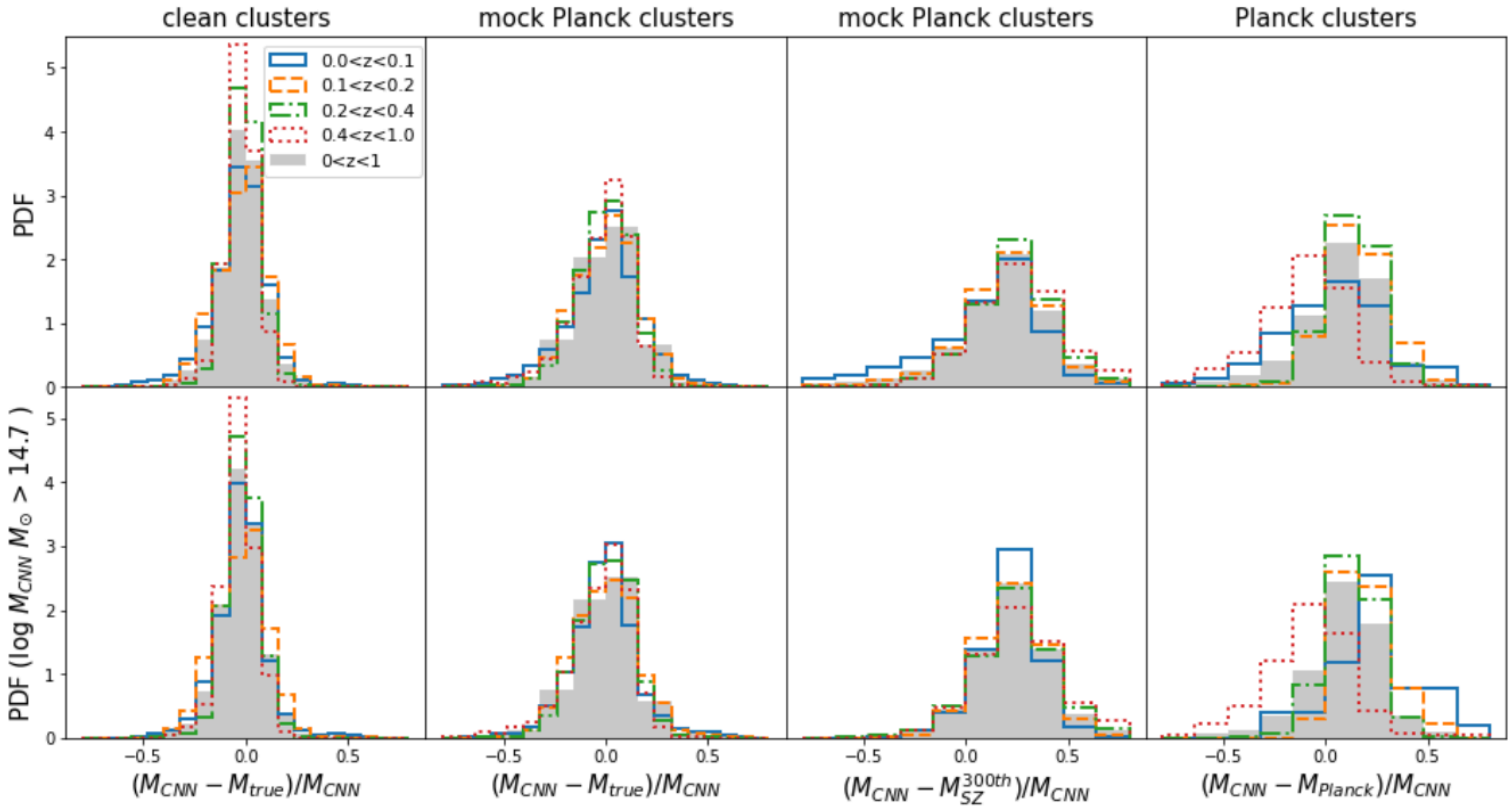
Left: We show the results for the 300th galaxy clusters. We compare the predicted mass M_{CNN} with M_{SZ}^{300th},which is the mass computed using Planck SZ scaling relations on the 300th simulated clusters (Arnaud et al. 2010). **Right:** We show the relative error of the predicted mass M_{CNN} and the Planck mass M_{Planck} . Note that M_{Planck} is calculated using the same SZ scaling relation and no bias correction is taking into account. $M_{planck} = M_{SZ}^{Planck}$.

Results on Real data

Problem with the Planck resolution at low masses



Left: We show the results for the 300th galaxy clusters. We compare the predicted mass M_{CNN} with M_{SZ}^{300th},which is the mass computed using Planck SZ scaling relations on the 300th simulated clusters (Arnaud et al. 2010). **Right:** We show the relative error of the predicted mass M_{CNN} and the Planck mass M_{Planck} . Note that M_{Planck} is calculated using the same SZ scaling relation and no bias correction is taking into account. $M_{planck} = M_{SZ}^{Planck}$.



We can also estimate a probability distribution function PDF for the errors (scatter). The quantitative values of these PDFs are shown in the table.

Statistics for PDFS

redshift	clean data $(M_{\rm CNN} - M_{\rm true})/M_{\rm CNN}$	Mock Planck data $(M_{\rm CNN} - M_{\rm true})/M_{\rm CNN}$	Mock Planck data $(M_{\rm CNN} - M_{\rm SZ}^{300th})/M_{\rm CNN}$	Planck data $(M_{\rm CNN} - M_{\rm Planck})/M_{\rm CNN}$	N
$0 < z < 0.1$	-0.03 ± 0.14	-0.02 ± 0.21	0.08 ± 0.37	0.03 ± 0.29	228
$0.1 < z < 0.2$	-0.01 ± 0.12	0.01 ± 0.15	0.2 ± 0.23	0.15 ± 0.14	245
$0.2 < z < 0.4$	-0.01 ± 0.08	-0.01 ± 0.13	0.25 ± 0.22	0.12 ± 0.13	443
$0.4 < z < 1$	-0.02 ± 0.08	-0.02 ± 0.15	0.29 ± 0.26	-0.09 ± 0.22	178
$0 < z < 1$	-0.02 ± 0.11	-0.02 ± 0.17	0.19 ± 0.30	0.08 ± 0.21	1094
			$\log M_{\rm CNN}\,[M_\odot] > 14.7$		
$0 < z < 0.1$	-0.02 ± 0.13	-0.01 ± 0.16	0.21 ± 0.15	0.26 ± 0.25	33
$0.1 < z < 0.2$	-0.02 ± 0.13	0.01 ± 0.16	0.23 ± 0.17	0.18 ± 0.13	82
$0.2 < z < 0.4$	-0.01 ± 0.08	-0.01 ± 0.13	0.25 ± 0.22	0.13 ± 0.13	403
$0.4 < z < 1$	-0.02 ± 0.08	-0.02 ± 0.15	0.29 ± 0.25	-0.08 ± 0.21	166
$0 < z < 1$	-0.02 ± 0.11	-0.01 ± 0.17	0.25 ± 0.20	0.09 ± 0.19	684

Table 1. We show the mean μ and the standard deviation σ which correspond to the PDFs in figure 7 for different redshift bins and different data sets. The results are display in the following format: $\mu \pm \sigma$. The last column corresponds to the number of maps in a particular redshift bin for the real Planck data set.

Statistics for PDFS

redshift	clean data $(M_{\mathrm{CNN}} - M_{\mathrm{true}})/M_{\mathrm{CNN}}$	Mock Planck data $(M_{\mathrm{CNN}} - M_{\mathrm{true}})/M_{\mathrm{CNN}}$	Mock Planck data $(M_{\mathrm{CNN}} - M_{\mathrm{SZ}}^{300th})/M_{\mathrm{CNN}}$	Planck data $(M_{\mathrm{CNN}} - M_{\mathrm{Planck}})/M_{\mathrm{CNN}}$	N
$0 < z < 0.1$	-0.03 ± 0.14	-0.02 ± 0.21	0.08 ± 0.37	0.03 ± 0.29	228
$0.1 < z < 0.2$	-0.01 ± 0.12	0.01 ± 0.15	0.2 ± 0.23	0.15 ± 0.14	245
$0.2 < z < 0.4$	-0.01 ± 0.08	-0.01 ± 0.13	0.25 ± 0.22	0.12 ± 0.13	443
$0.4 < z < 1$	-0.02 ± 0.08	-0.02 ± 0.15	0.29 ± 0.26	-0.09 ± 0.22	178
$0 < z < 1$	-0.02 ± 0.11	-0.02 ± 0.17	0.19 ± 0.30	0.08 ± 0.21	1094
$\log M_{\mathrm{CNN}}\,[M_{\odot}] > 14.7$					
$0 < z < 0.1$	-0.02 ± 0.13	-0.01 ± 0.16	0.21 ± 0.15	0.26 ± 0.25	33
$0.1 < z < 0.2$	-0.02 ± 0.13	0.01 ± 0.16	0.23 ± 0.17	0.18 ± 0.13	82
$0.2 < z < 0.4$	-0.01 ± 0.08	-0.01 ± 0.13	0.25 ± 0.22	0.13 ± 0.13	403
$0.4 < z < 1$	-0.02 ± 0.08	-0.02 ± 0.15	0.29 ± 0.25	-0.08 ± 0.21	166
$0 < z < 1$	-0.02 ± 0.11	-0.01 ± 0.17	0.25 ± 0.20	0.09 ± 0.19	684

Table 1. We show the mean μ and the standard deviation σ which correspond to the PDFs in figure 7 for different redshift bins and different data sets. The results are display in the following format: $\mu \pm \sigma$. The last column corresponds to the number of maps in a particular redshift bin for the real Planck data set.

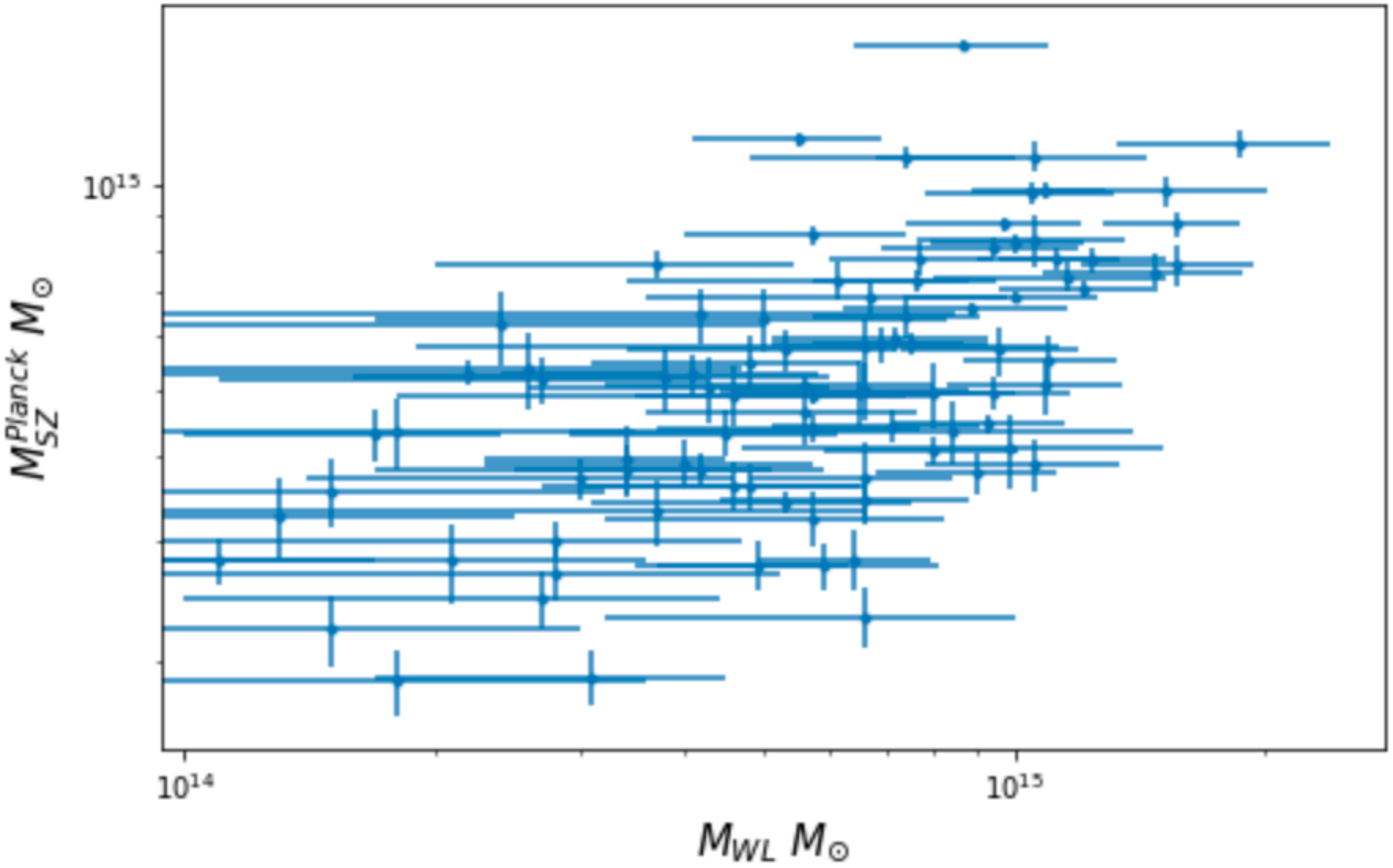
Statistics for PDFS

redshift	clean data $(M_{\rm CNN} - M_{\rm true})/M_{\rm CNN}$	Mock Planck data $(M_{\rm CNN} - M_{\rm true})/M_{\rm CNN}$	Mock Planck data $(M_{\rm CNN} - M_{\rm SZ}^{300th})/M_{\rm CNN}$	Planck data $(M_{\rm CNN} - M_{\rm Planck})/M_{\rm CNN}$	N
$0 < z < 0.1$	-0.03 ± 0.14	-0.02 ± 0.21	0.08 ± 0.37	0.03 ± 0.29	228
$0.1 < z < 0.2$	-0.01 ± 0.12	0.01 ± 0.15	0.2 ± 0.23	0.15 ± 0.14	245
$0.2 < z < 0.4$	-0.01 ± 0.08	-0.01 ± 0.13	0.25 ± 0.22	0.12 ± 0.13	443
$0.4 < z < 1$	-0.02 ± 0.08	-0.02 ± 0.15	0.29 ± 0.26	-0.09 ± 0.22	178
$0 < z < 1$	-0.02 ± 0.11	-0.02 ± 0.17	0.19 ± 0.30	0.08 ± 0.21	1094
			$\log M_{\rm CNN}\,[M_\odot] > 14.7$		
$0 < z < 0.1$	-0.02 ± 0.13	-0.01 ± 0.16	0.21 ± 0.15	0.26 ± 0.25	33
$0.1 < z < 0.2$	-0.02 ± 0.13	0.01 ± 0.16	0.23 ± 0.17	0.18 ± 0.13	82
$0.2 < z < 0.4$	-0.01 ± 0.08	-0.01 ± 0.13	0.25 ± 0.22	0.13 ± 0.13	403
$0.4 < z < 1$	-0.02 ± 0.08	-0.02 ± 0.15	0.29 ± 0.25	-0.08 ± 0.21	166
$0 < z < 1$	-0.02 ± 0.11	-0.01 ± 0.17	0.25 ± 0.20	0.09 ± 0.19	684

Table 1. We show the mean μ and the standard deviation σ which correspond to the PDFs in figure 7 for different redshift bins and different data sets. The results are display in the following format: $\mu \pm \sigma$. The last column corresponds to the number of maps in a particular redshift bin for the real Planck data set.

Comparison with weak lensing masses

- A sample of 91 Planck clusters whose WL masses have been already estimated (Umetsu et al. 2014; Sereno et al.2017; Medezinski et al. 2018; Herbonnet et al. 2020) is used to compare them with their CNN mass estimates.



Work in progress..

Ongoing ML projects in the 300th collaboration

The 300th data base of simulated clusters is an excellent tool for the training of other machine learning (ML) algorithms.

Ongoing ML projects :

- Inference of the projected total mass map from mock observations: SZ, X-ray, optical maps (with EURANOVA support).
- Inferring the 3D dynamical state parameter from 2D morphological estimators or from directly from SZ/X-ray maps using CNN approach.
- Regressions of baryon properties from dark matter halo catalogs. Application to large volume dark matter only simulations (to construct fast all sky cluster number counts in X-ray and SZ).

Conclusions

- We have managed to give an estimation to the total cluster mass without any prior assumption on the cluster dynamical state.
- We have found that Y_{500} might be overestimated for low mass clusters at low redshift in the Planck catalog due to resolution effects.
- However, the results on our simulated data and the results on real data are slightly different but nevertheless, statistically compatible and they both show an overall similar trend. The possible difference can be due to the physics modelled by the simulations.
- We tested our algorithm on the GIZMO simulation of the same clusters finding no significant difference with the results presented here.
- Nevertheless, as in other areas of image processing, the success of these techniques depends on the quality and accuracy of the training set. To this end, hydrodynamical numerical simulations are an indispensable tool to provide the mock observations on which we can train CNN architectures.

Thanks, questions?