

# The three hundred project: mapping the matter distribution in galaxy clusters via deep learning from multiview simulated observations

Daniel de Andres<sup>1,2\*</sup>, Weiguang Cui<sup>1,2,3†</sup>, Gustavo Yepes<sup>1,2</sup>, Marco De Petris,<sup>4</sup>  
Antonio Ferragamo,<sup>4,5</sup> Federico De Luca<sup>4</sup>, Gianmarco Aversano<sup>6</sup> and Douglas Rennehan<sup>7</sup>

<sup>1</sup>Departamento de Física Teórica, Universidad Autónoma de Madrid, M-8, Cantoblanco, E-28049 Madrid, Spain

<sup>2</sup>Centro de Investigación Avanzada en Física Fundamental, (CIAFF), Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain

<sup>3</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

<sup>4</sup>Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro, I-5-00185 Roma, Italy

<sup>5</sup>Instituto de Astrofísica de Canarias (IAC), E-38205 La Laguna, Spain

<sup>6</sup>EURANOVA, Rue Emile Francqui 4, 1435 Mont-Saint-Guibert, Belgium

<sup>7</sup>Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA

Accepted 2024 January 3. Received 2024 January 3; in original form 2023 November 1

## ABSTRACT

A galaxy cluster as the most massive gravitationally bound object in the Universe, is dominated by dark matter, which unfortunately can only be investigated through its interaction with the luminous baryons with some simplified assumptions that introduce an un-preferred bias. In this work, we, *for the first time*, propose a deep learning method based on the U-Net architecture, to directly infer the projected total mass density map from idealized observations of simulated galaxy clusters at multiwavelengths. The model is trained with a large data set of simulated images from clusters of THE THREE HUNDRED PROJECT. Although machine learning (ML) models do not depend on the assumptions of the dynamics of the intracluster medium, our whole method relies on the choice of the physics implemented in the hydrodynamic simulations, which is a limitation of the method. Through different metrics to assess the fidelity of the inferred density map, we show that the predicted total mass distribution is in very good agreement with the true simulated cluster. Therefore, it is not surprising to see the integrated halo mass is almost unbiased, around 1 per cent for the best result from multiview, and the scatter is also very small, basically within 3 per cent. This result suggests that this ML method provides an alternative and more accessible approach to reconstructing the overall matter distribution in galaxy clusters, which can complement the lensing method.

**Key words:** methods: numerical – galaxies: clusters: general – cosmology: theory – galaxies: haloes – dark matter – large-scale structure of Universe.

## 1 INTRODUCTION

Estimating the matter content in galaxy clusters (see Kravtsov & Borgani 2012, for a review) is crucial for cosmological studies due to the fact that they are the biggest gravitationally bound objects originating from small density fluctuations in the early universe. Thus cosmological parameters can be constrained by studying the abundance of galaxy clusters as a function of the mass and redshift (e.g. Allen, Evrard & Mantz 2011; Planck Collaboration XXIV 2016b; Pratt et al. 2019; Salvati et al. 2022). Galaxy clusters are mainly composed of dark matter (DM), which is about 80 per cent of their total mass, diffused hot gas (about 12 per cent), i.e. intracluster medium (ICM), and stars, mainly in galaxies (the remaining 8 per cent).

Galaxies within clusters are normally observed at different optical bands. Both ground-based and space-based instruments have been used to measure these galaxy properties through their spectrum energy distributions. For example, the Sloan Digital Sky telescope (SDSS<sup>1</sup>) and the Hubble Space Telescope (HST<sup>2</sup>) have been crucial for the studies of galaxy clusters as well as cosmology. The recent photometric surveys, e.g. the Dark Energy Survey (DES<sup>3</sup>) using the Dark Energy Camera and the Javalambre Physics of the Accelerating Universe Astrophysical Survey J-PAS<sup>4</sup> from the Javalambre Survey Telescope, provide an unprecedented amount of data for understanding our Universe. Not to mention the recently launched space telescopes *James Webb Space Telescope (JWST<sup>5</sup>)* and *Euclid<sup>6</sup>*.

<sup>1</sup><https://www.sdss.org>

<sup>2</sup><https://science.nasa.gov/mission/hubble>

<sup>3</sup><https://www.darkenergysurvey.org>

<sup>4</sup><https://www.j-pas.org/survey>

<sup>5</sup><https://www.jwst.nasa.gov>

<sup>6</sup>[https://www.esa.int/Science\\_Exploration/Space\\_Science/Euclid](https://www.esa.int/Science_Exploration/Space_Science/Euclid)

\* E-mail: [daniel.deandres@uam.es](mailto:daniel.deandres@uam.es)

† Talento-CM fellow

While optical surveys explore galaxy members, the ICM is targeted through X-ray emission and the inverse-Compton scattering of the cosmic microwave background (CMB) photons, i.e. the Sunyaev–Zel’dovich (SZ) effect (Sunyaev & Zeldovich 1972). The SZ effect can be observed at millimetre wavelengths and different instruments have managed to detect more than a thousand clusters through it, such as the Planck satellite (PSZ2; Planck Collaboration XXVII 2016c), the Atacama Cosmology Telescope (ACT; Hilton et al. 2018), and the South Pole Telescope (SPT; Bleem et al. 2020). Nevertheless, Planck is the only all-sky survey for SZ clusters. On the contrary, clusters X-ray observations are more recently explored by eROSITA (Liu et al. 2022) and targeted *XMM–Newton* observations, such as the CHEX-MATE project (CHEX-MATE Collaboration 2021).

Although the total mass of a cluster of galaxies is not observable, it can be inferred by observing its baryonic components. It can be estimated from the dynamics (e.g. Biviano et al. 2006) or the abundance/richness (e.g. Rozo et al. 2009) of the member galaxies, or from the ICM radial profiles in X-ray and SZ observations with the assumption of the diffused gas is distributed following the hydrostatic equilibrium (HE) hypothesis (e.g. Gianfagna et al. 2021, 2023). Alternatively, their physical quantities, which are integrated inside a fixed aperture in the sky, strictly related to the mass of the object can be selected under the self-similarity assumption (Bryan & Norman 1998) as suitable observational proxies. However, these estimated masses rely upon theoretical assumptions, such as HE, and are therefore possibly affected by bias. Nevertheless, only weak gravitational lensing (WL) is sensitive to all matter along the line of sight and measures the total projected matter (Becker & Kravtsov 2011; Herbonnet et al. 2022). However, gravitational lensing observations are not numerous, with tens of images, in comparison to the hundreds or few thousands of observations of galaxy clusters available for SZ, X-ray, and stars. In addition, mass inference from WL also suffers from systematics due to the theoretical assumptions when reconstructing mass ( $\kappa$ ) maps from shear (Kaiser & Squires 1993), similar to the inference of the mass from SZ or X-ray. Recent data-driven approaches based on simulations and convolutional neural networks have proven to outperform conventional methods (Hong et al. 2021b) also for the task of WL mass reconstruction.

Machine learning (ML) has been applied in astronomy since quite a long ago (e.g. Odewahn et al. 1992), but only in past years we have witnessed an unprecedented increase of deep learning (DL) methods (for a review, see e.g. Huertas-Company & Lanusse 2023; Smith & Geach 2023), which implies a change in the paradigm from applying specific algorithms to fully data-driven science. We stress here that applying DL in real observations of galaxy clusters is inherently more challenging than in simulations and accordingly most of the literature is limited to only utilizing simulated data. For example, DL has also been used for identifying galaxy cluster members from HST images (Angora et al. 2020), increasing the resolution (super-resolution) and de-noising *XMM–Newton* images (Sweere et al. 2022), and deprojecting and deconvolving galaxy cluster X-ray temperature profiles within the CHEX-MATE collaboration (Iqbal et al. 2023). In our interests, recent studies have shown that galaxy cluster masses can be estimated using DL methods from catalogues of simulated galaxy clusters (Ho et al. 2019; Ntampaka et al. 2019; Gupta & Reichardt 2020; Kodi Ramanah et al. 2020; Yan et al. 2020; Gupta & Reichardt 2021; Ho et al. 2021; de Andres et al. 2022; Ferragamo et al. 2023; Ho et al. 2023), which outperform these classical methods. Fortunately, only recently these methods have also been applied to infer the mass of galaxy clusters of real surveys at different wavelengths, Kodi Ramanah, Wojtak & Arendse (2021)

applied DL to infer galaxy clusters masses from the SDSS Legacy Survey (Strauss et al. 2002), Ho et al. (2022) estimated the dynamical mass of the Coma cluster, de Andres et al. (2022) inferred the masses of the full-sky Planck satellite PSZ2 catalogue, Krippendorf et al. (2023) followed a ML approach to infer galaxy cluster masses from eROSITA X-ray images.

Notwithstanding, previous works on mass inference focus on the mass inside a sphere of a certain radius (i.e.  $R_{200}^7$  or  $R_{500}$ ). Recently, Ferragamo et al. (2023) pushed these ML applications even further by showing that ML can predict the mass radial profile using SZ idealized simulated images and thus, being able to theoretically obtain valuable information on internal properties of galaxy clusters as the concentration. As a general remark, all these studies suggest that the estimated masses are not affected by observational biases and assumptions due to the fact that they do not rely on hypotheses about the dynamics nor the hydrostatic equilibrium with spherical symmetry of the ICM, but rather on the quality of data set provided by cosmological simulations.

In this work, for the first time, we explore DL models to infer the projected matter density fields from simulated baryonic observations. To this end, we utilize the THE THREE HUNDRED (THE300; Cui et al. 2018): a set of ‘zoom-in’ hydrodynamical and cosmological simulations to generate the input idealized observations: SZ, X-ray, and star maps, and the output maps, mass maps. The data set created for this purpose corresponds to a set of idealized observations free from the observational impacts typical of real instruments, such as noise and point sources. Therefore, this work represents the needed theoretical proof-of-concept study where the limitations of DL are tested. For further details regarding our data set, we refer the reader to Section 2.

The manuscript is organized as follows: in Section 2, the creation of the data set is discussed, THE THREE HUNDRED set of hydrodynamic simulations and the particular selection of halo-sized objects that are optimal for training ML models and the characteristics of our input maps, Compton- $\gamma$  parameters maps (SZ), X-ray surface brightness (X-ray), and star density maps, and output mass density maps. In Section 3, we discuss the DL model and architectures considered for this project as well as how the models are trained. In Section 4, the results of the model applied to the test set are shown. To this end, we design a set of test metrics to assess the fidelity of the predicted mass density maps: pixel-wise statistics, cylindrical radial profiles, power spectra, and maximum mean discrepancy. Finally, in Section 5 the main conclusions of the work are drawn.

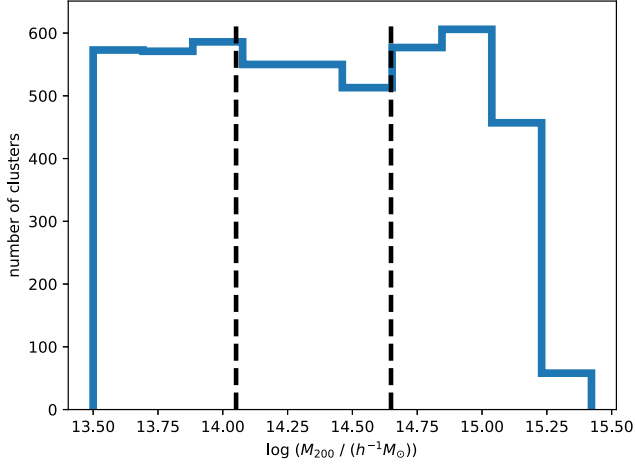
## 2 DATA SET

### 2.1 The three hundred simulations

In order to create the data set used for training our DL model, we use THE300<sup>8</sup> simulations (Cui et al. 2018). THE300 simulation project focuses on a set of 324 ‘zoom-in’ hydrodynamic simulations centred on the most massive haloes of the MULTIDARK simulation (MDPL2; Klypin et al. 2016) utilizing the cosmological parameters inferred by Planck (Planck Collaboration XIII 2016a). The full MDPL2 simulation contains  $3840^3$  particles whose mass is  $1.5 \times 10^9 h^{-1} M_{\odot}$ . Moreover, each cluster region covers  $15 h^{-1}$  Mpc and they have been simulated with different baryonic models: GADGET-MUSIC

<sup>7</sup> $R_{200}$  stands for the radius of the sphere at which the density is 200 times the critical density of the Universe at the corresponding redshift

<sup>8</sup><https://the300-project.org>



**Figure 1.** Mass distribution of the selected galaxy clusters sample. Note that the selected objects follow an approximately flat distribution in the logarithm of the mass. Dashed black vertical lines correspond to mass thresholds that divide the sample in 3 mass bins containing an equal number of clusters and they are at  $\log(M_{200}/(h^{-1} M_{\odot})) = 14.05$  and  $\log(M_{200}/(h^{-1} M_{\odot})) = 14.65$ .

(Sembolini et al. 2013), GADGET-X (Murante et al. 2010; Rasia et al. 2015), GIZMO-SIMBA (Davé et al. 2019; Cui et al. 2022). However, in this work, we only make use of GADGET-X (Cui et al. 2018) and GIZMO-SIMBA (Cui et al. 2022) runs.

In THE300 simulations haloes are identified by the AMIGA’S HALO FINDER package (AHF; Knollmann & Knebe 2009) and only haloes with  $M_{200} > 10^{13.5} h^{-1} M_{\odot}$ , at redshift  $z \simeq 0$  and free from contamination are selected. Note that  $M_{200}$  stands for the mass inside a sphere of density 200 times the critical density of the Universe at the corresponding redshift. Bear in mind that for THE300 simulations, the scaling relations do not depend on redshift at least up to  $z \simeq 1$  (de Andres Andres et al. 2023). Free from contamination means that heavy (or low resolution) DM particles entering from outside the re-simulated volume are not present inside our selected haloes. The selection is first done for GADGET-X, the counterpart haloes from GIZMO-SIMBA are selected to have the same number of objects and mass distribution as in GADGET-X.

Furthermore, to increase the statistics of underrepresented clusters we added objects from other snapshots ( $z = 0.022, 0.045, 0.069, 0.093, \text{ and } 0.117$ ) to account for a flat mass distribution as in Ferragamo et al. (2023), which is presented in Fig. 1. In that figure, the number of galaxy clusters as a function of mass is shown, and the whole sample is equally split into 3 bins with two vertical lines. Moreover, the THE300 simulations contain a rich variety of massive galaxy clusters with different dynamical states (De Luca et al. 2021). Note that at the end of the selection procedure, 2518 haloes are selected from GADGET-X and 2523 from GIZMO-SIMBA, with the same flat distribution in mass. We refer the reader to the Appendix A for more information.

## 2.2 Simulated multiview images

From the selected haloes, we further compute the Compton-y parameter – a dimensionless measure of the SZ effect, X-ray surface brightness, star mass density, and total mass density maps, see Table 1. Per each halo in our data set, 29 different line-of-sight projections are considered to create 146 189 maps in total, a large enough data set for training our ML model. To test the effectiveness of ML models and avoid observational biases, the generated images correspond to

a set of idealized maps in which the angular resolution is sufficiently high for all of them regardless of their masses. As such, we have created a data set in which the size of each map is  $2 \times R_{200}$ , all sampled with the same number of pixels  $N_{\text{pix}} = 640$  and therefore all maps are equally resolved.

Although all maps are initially created with a resolution of  $640 \times 640$  pixels, they are smoothed with a Gaussian kernel of  $\text{FWHM} \simeq 0.015 R_{200}$  and they are re-gridded to a lower resolution of  $80 \times 80$  pixels, which is needed for computational efficiency. Nevertheless, these ‘low-resolution’ final maps are still very well resolved, and the angular resolution for all maps is below 1 arcmin. The idea that we desire to convey is that this data set corresponds to an ideal theoretical data set from hydro-simulations which can be easily applied to the ML models for testing their performance. In detail, the SZ maps, X-ray surface brightness maps, star mass density maps, and total mass density maps are computed as follows:

**Compton-y parameter maps** ( $y$ ) correspond to the integrated pressure along the observer’s line of sight (l.o.s.)  $dl$ .

$$y = \frac{\sigma_{\text{T}} k_{\text{B}}}{m_e c^2} \int n_e T_e dl, \quad (1)$$

where  $\sigma_{\text{T}}$  is the Thomson cross-section,  $k_{\text{B}}$  is the Boltzmann constant,  $c$  the speed of light,  $m_e$  the electron rest-mass,  $n_e$  the electron number density, and  $T_e$  the electron temperature. These maps are computed using the publicly available package PYMSZ<sup>9</sup> (Cui et al. 2018).

**X-ray surface brightness maps** (X-ray) are estimated by computing the X-ray emission by thermal bremsstrahlung in the hot intracluster medium using a wrapper of PyAtomDB<sup>10</sup> package to compute X-ray luminosity.<sup>11</sup> First, using the atomic data base, we estimated the bolometric X-ray luminosity of gas particles in the simulation. Then we projected the sum of the luminosity values along the observer’s l.o.s. The projected luminosity  $L_{ij}$ , is then divided by the surface area  $\Sigma$  of the pixel  $i, j$ :

$$\text{X-ray}_{ij} = L_{ij} / \Sigma, \quad (2)$$

which is

$$\Sigma = (R_{200}/40)^2. \quad (3)$$

**Stellar density maps** are generated by projecting the sum of the masses of the star particles  $\sum_p M_{*,p}$  along the observer’s line of sight, that represent the pixel  $(i, j)$ . This value is then divided by the surface area of the pixel  $\Sigma$ :

$$\text{star}_{i,j} = \sum_p M_{*,p} / \Sigma. \quad (4)$$

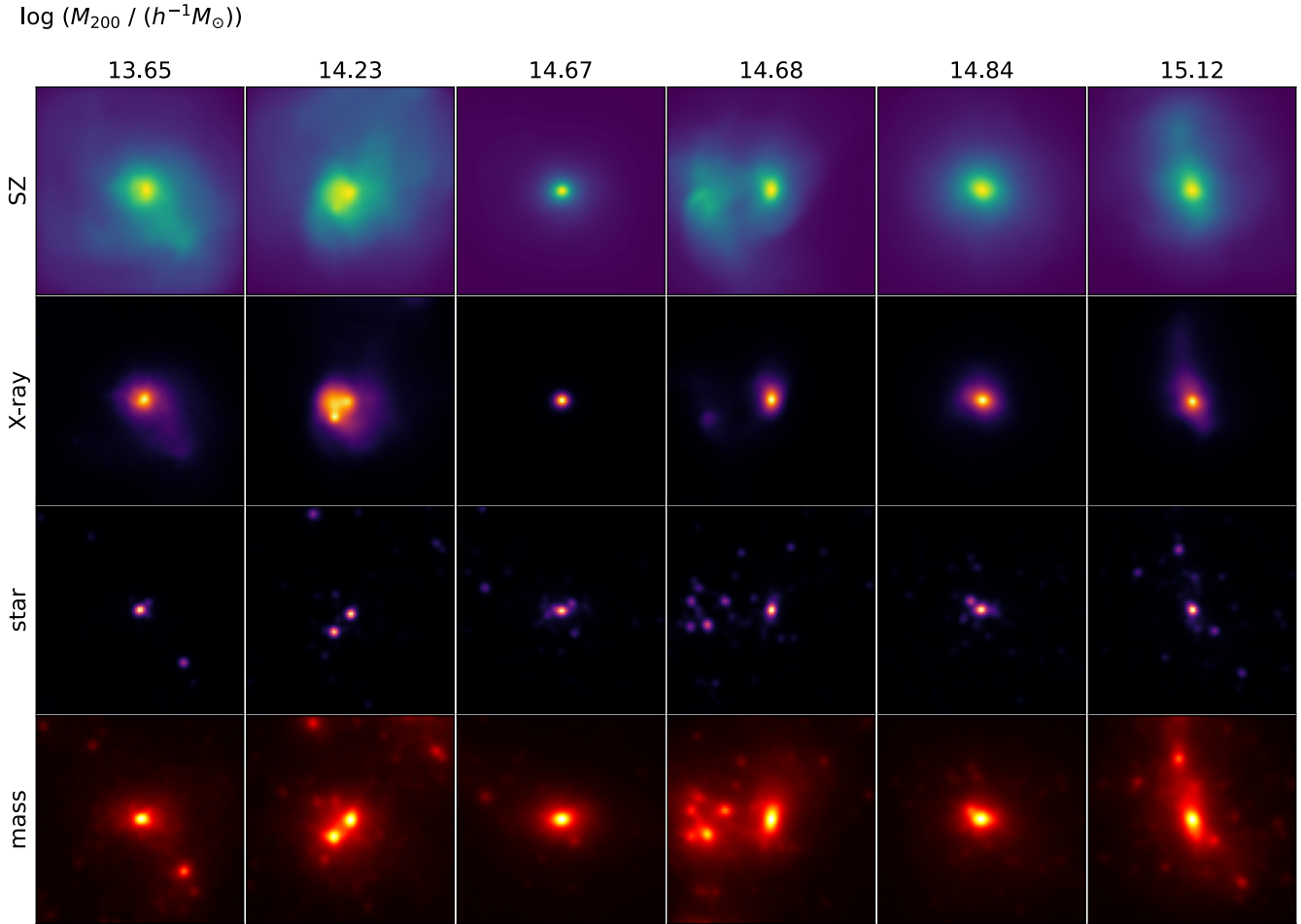
We note that opting for stellar density as the stars simulated maps requires some clarifications: (1) First, stellar observations in surveys are all presented as multiband observations, e.g. SDSS.<sup>12</sup> As a first proof-of-concept approximation, the density field of the stars reflects their spatial distribution and thus, for this purpose, as shown in Yan et al. (2020)  $M_{500}$  could be inferred from star density maps. (2) We can, for sure, compute luminosity maps from stars in the simulation by using Stellar Population Synthesis models, such as SPSM (Devriendt, Guiderdoni & Sadat 1999; Bruzual & Charlot 2003). This approach is closer to the real observation images. However, we argue that (a) the mass density map from observations can be derived by

<sup>9</sup><https://github.com/weiguangcui/pymsz>

<sup>10</sup><https://atomdb.readthedocs.io/en/master/>

<sup>11</sup><https://github.com/rennehan/xraylum>

<sup>12</sup><https://www.sdss.org>



**Figure 2.** From top to bottom, we show some examples of maps used for training the U-Net model: SZ, X-ray, star, and mass. The examples' mass  $M_{200}$  is shown at the top of each column. The size of all maps is  $2 \times R_{200}$  regardless of the mass.

fitting the different colour bands to a Stellar Energy Distribution, using the same SPSM, albeit a little bit higher uncertainties and complexities due to several hypotheses on the metallicity, stellar mass function, etc... (b) there should be not much difference in practice with either luminosity or stellar mass maps. This is because the galaxies' luminosity-to-mass ratio is approximately constant for The300 simulation, see fig. 8 of Cui et al. (2018). (c) to apply our ML models to real observation images, there are still multiple steps to take care of, such as instruments, and background noise. Therefore, we adopt the stellar mass maps for simplicity in this concept-proofing paper and leave the proper reproduction of mock observation images in the following work.

**Mass density maps** are generated by projecting the sum of the masses of all particles, i.e. gas, star, dark matter, and black hole particles in the observer's line of sight. Similarly, this value is divided by the surface area of a pixel.

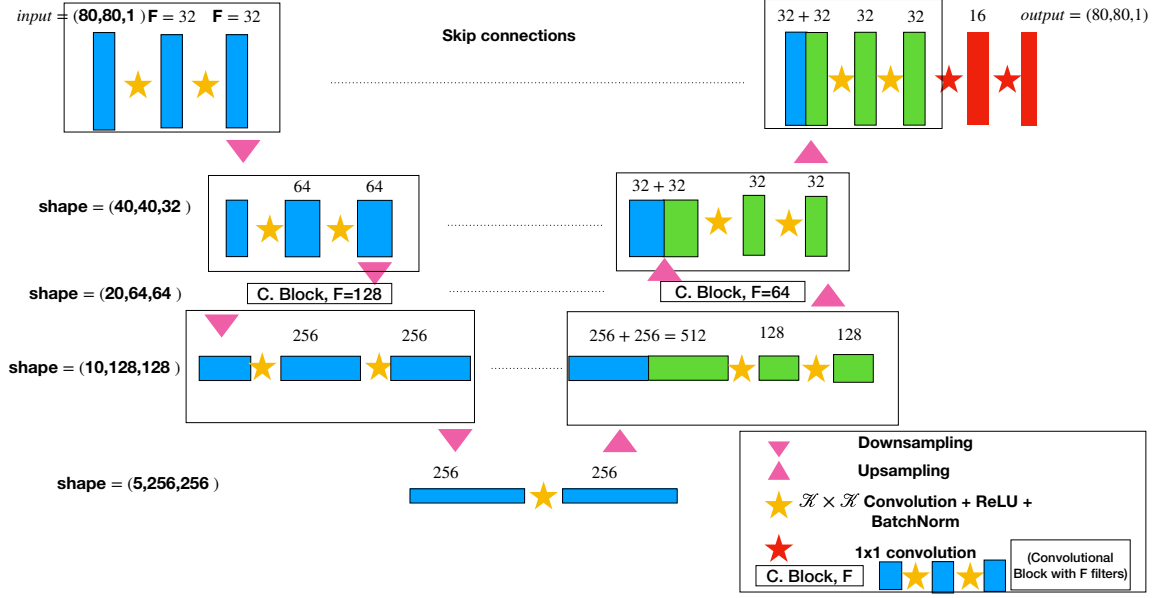
In Fig. 2, examples of our simulated maps are displayed for different cluster masses. With Compton- $\gamma$  maps, X-ray surface brightness maps and star density maps, we aim to infer total mass density maps shown at the bottom by using a DL model that is able to capture the non-linear relations between input (SZ, X-ray, and star) and output (total mass) maps. This model can end-to-end translate one input map into mass density. The choice of the particular model used and the training of that model are described in the next section.

### 3 METHODS

#### 3.1 The deep learning model

The chosen model and its architecture used for the generator of mass density maps is based on convolutional neural networks following a U-Net architecture. This model was originally developed for image segmentation in the field of biological imaging (Ronneberger, Fischer & Brox 2015). In astronomy, some applications followed the original purpose of segmentation such as removing radio frequency interference (Akeret et al. 2017), but also the U-Net has been successfully considered for various and different tasks (Milletari, Navab & Ahmadi 2016; Aragon-Calvo 2019; Berger & Stein 2019; Hausen & Robertson 2020; Hong et al. 2021a; Lauritsen et al. 2021). The U-Net is characterized by a set of two paths: the down-sampling path (or encoder) and the up-sampling path (or decoder). For our application, the considered U-Net is described below and represented in Fig. 3.

(i) The **down-sampling path** consists of a succession of convolutional blocks, each of these applies two  $\mathcal{K} \times \mathcal{K}$  convolutions with  $F$  filters ( $C_{\mathcal{K} \times \mathcal{K}}^F$ ), being  $\mathcal{K}$  the size of the receptive field or the kernel size. After the convolutions, a rectified linear unit operation is applied (ReLU; Nair & Hinton 2010) to ensure non-linearity, and a



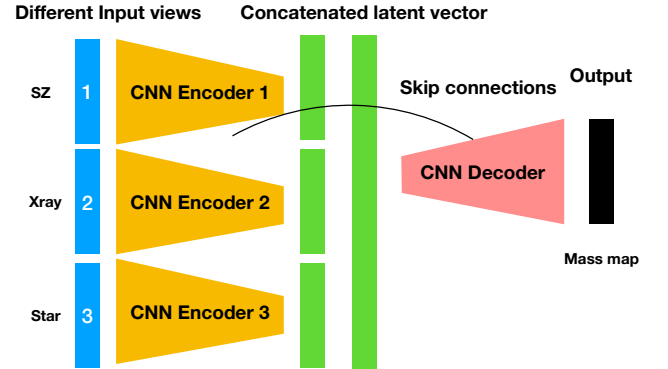
**Figure 3.** U-Net architecture for input dimensions = (80 pixels, 80 pixels, 1 channel) and  $F = 32$ . The input image (star, SZ, or X-ray) is down-sampled 4 times, once per layer through a set of convolution of kernel size  $\mathcal{K} \times \mathcal{K}$ . At the very bottom, the down-sampled representation is up-sampled using a similar convolutional block to generate the output mass map. Skip connections are used to ensure that the information is not totally lost during the down-sampling operations. We remind that dropout is applied to all convolutional layers only in the decoder.

batch normalization operation (BatchNorm; Ioffe & Szegedy 2015) is applied after each of the convolutions to improve the stability and performance of the network. Later, a  $2 \times 2$  max-pooling (e.g. Scherer, Müller & Behnke 2010) operation down-samples the tensor shape so that this is reduced by a factor of  $1/2$ . The subsequent convolutional blocks are built with the same architecture, maintaining the kernel size of the convolutions but increasing the number of filters by a factor of 2, i.e. for the  $n$  layer, the filters are  $F \times 2^{n-1}$ , where  $n = 1, \dots, N$  is the layer and  $N$  is the maximum number of layers. After down-sampling  $N$  times, a final convolution  $C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^{N-1}}$  is applied. Therefore, the down-sampling path can be written as a series of convolutions as follows:

$$\text{Encoder} = C_{\mathcal{K} \times \mathcal{K}}^F C_{\mathcal{K} \times \mathcal{K}}^F \xrightarrow{\text{down sampling}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2} \xrightarrow{\text{down sampling}} \dots \xrightarrow{\text{down sampling}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^{N-1}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^{N-1}} \xrightarrow{\text{down sampling}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^N} \quad (5)$$

(ii) The **up-sampling path** consists of a succession of similar convolutional blocks to infer the output mass density map from the latent (central layer) representation given by the down-sampling path. From this encoded reduced representation, which is a tensor whose shape has been decreased by a factor of  $1/2^N$  times and has  $F \times 2^{N-1}$  channels, up-sampling operations are applied until the shape of the output mass density map is recovered. This up-sampling operation consists of repeating the nearest points to increase the shape of the data by a factor of 2. Then, skip connections are used for the same row layers, and thus, information is not bottlenecked in the latent representation. At the final decoder step,  $1 \times 1$  convolutions are applied to recover the filter dimensions of the output mass density  $F = 1$ . For preventing overfitting, random ‘dropout’ (Hinton et al. 2012) is considered only in the convolutional blocks of the decoder.

$$\text{Decoder} = \xrightarrow{\text{up sampling}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^{N-2}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^{N-2}} \xrightarrow{\text{up sampling}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^{N-3}} C_{\mathcal{K} \times \mathcal{K}}^{F \times 2^{N-3}} \dots \xrightarrow{\text{up sampling}} C_{\mathcal{K} \times \mathcal{K}}^F C_{\mathcal{K} \times \mathcal{K}}^F C_{1 \times 1}^{F/2} C_{1 \times 1}^1 \quad (6)$$



**Figure 4.** Multiview 3 approach. Three different encoders are considered, with one down-sampling per each input view. Then this information is concatenated in the internal layer. From this concatenated latent vector, one decoder is in charge of the inference of the output mass density map.

### 3.1.1 Multiview approaches

This model has the advantage that it can be easily generalized to extract information simultaneously from multiple input views. To do that two approaches are studied:

(i) The three different views are combined in a single input tensor whose shape is increased as if it were an RGB image. This approach is labelled as ‘multi-1’; and

(ii) Three different encoders are used, one encoder for each input view. Subsequently, the three latent vectors are concatenated in the internal latent space. One decoder, the generator, is devoted to creating mass density maps from the information of this concatenated latent space. This approach is labelled as ‘multi-3’ (see Fig. 4).

A summary of all the models used in this work can be found in Table 2.

**Table 1.** Data set of simulated maps from THE THREE HUNDRED simulations.

| Map                                 | Short name | Units                              |
|-------------------------------------|------------|------------------------------------|
| Compton-y parameter                 | SZ         |                                    |
| Bolometric X-ray surface brightness | X-ray      | $\text{ergs}^{-1} \text{kpc}^{-2}$ |
| Star density maps                   | Star       | $h^{-1} M_{\odot} \text{kpc}^{-2}$ |
| Mass density maps                   | Mass       | $h^{-1} M_{\odot} \text{kpc}^{-2}$ |

**Table 2.** Models considered in this work. Each is a variation of the U-Net architecture presented in Fig. 3 to account for different inputs.

| Model name | Input maps          | Number of encoders |
|------------|---------------------|--------------------|
| Star       | Star                | 1                  |
| SZ         | SZ                  | 1                  |
| X-ray      | X-ray               | 1                  |
| Multi-1    | Star, SZ, and X-ray | 1                  |
| Multi-3    | Star, SZ, and X-ray | 3                  |

**Table 3.** The considered possible values of hyperparameters of our U-Net model.

| Hyperparameter           | Values        |
|--------------------------|---------------|
| Filters ( $F$ )          | 8, 9, ..., 32 |
| Kernel ( $K$ )           | 1, 2, ..., 6  |
| Number of layers ( $N$ ) | 1, ..., 4     |
| Dropout                  | [0, 0.5]      |

### 3.2 Training and validation

According to the description given in the previous section, we consider the following hyperparameters for our model: the number of channels or filters ( $F$ ), the size of the kernel in the convolutions ( $K$ ), the number of layers in both the encoder and decoder architectures ( $N$ ), and the fraction of neurons that are randomly omitted, i.e. the ‘dropout’. The considered possible values of these hyperparameters are shown in Table 3 and the optimal set of hyperparameters is found by TREE OF PARZEN ESTIMATORS algorithm (TPE; Bergstra et al. 2011). This algorithm is implemented and freely available at HYPEROPT<sup>13</sup>; DISTRIBUTED ASYNCHRONOUS HYPER-PARAMETER OPTIMISATION (Bergstra, Yamins & Cox 2013). The targeted loss function is the mean absolute error  $L_1$  between the predicted pixel values and the ground-truth pixels.

Furthermore, the data is randomly split into 3 subsets: the training data set composed of 80 per cent of the sample, the validation data set with 10 per cent of data and the remaining 10 per cent belongs to the test data set. The training data set is used for tuning the U-Net parameters, the TREE OF PARZEN ESTIMATORS algorithm is fed with validation data, and the test set is only used for displaying the final results.

We have taken 100 evaluations when performing the HYPEROPT optimization, which means that the U-Net needs to be fitted 100 times for each of the 5 different U-Nets. The models are trained using the ADAM optimizer (Kingma & Ba 2014) with a learning rate of  $10^{-4}$ , which is reduced to  $10^{-5}$  after 50 epochs. The training stops if the validation loss does not improve after 5 epochs, a process commonly known as ‘early stopping’, or if a model has trained 100 epochs. The hardware used for training is an NVIDIA A100 GPU, which

translates to 1 epoch and takes about 2 min to complete. Before fitting the models, the maps are normalized following the common standard normalization, see e.g. de Andres et al. (2022) for further details.

## 4 RESULTS

In this section, our aim is to assess the quality of predicted density maps by our model. To do that, we compare the ground-truth density maps with their corresponding predicted maps in the test set with simple visualization as the first metric for assessing the quality of the predictions. First, we show one example of our predicted maps in Fig. 5 for our different U-Nets accounting for different input views. The first column on the left shows the three input views corresponding to the same galaxy cluster and the ground-truth density map is located at the top left. The second column shows the predictions for several input bands: SZ, X-ray, star, and multiview. The last column corresponds to the residuals, i.e. the difference between the prediction and the ground truth.

As a general result, we observe that the predicted mass density maps from SZ and X-ray inputs are smoother and do not contain most of the substructures that can be appreciated in the ground-truth map. This is clear in the last column where the residuals mostly contain all the missing substructures and they are underestimated as shown in the figure in blue colour. Conversely, predictions from stars and from the multiview models contain most of the substructures. Their residuals are generally closer to zero than the others. We only show here the multi-3 approach, given the similarity among multi-1 and multi-3 predictions in a human eye test.

This visualization of density-map residuals in Fig. 5 apparently is not sufficient to numerical quantify the similarity between predict and true maps. Therefore, we are compelled to utilize a set of additional metrics to measure the discrepancies between the two. The considered metrics are the pixel-wise statistics, cylindrical radial mass profiles, power spectrum, and maximum mean discrepancy which are studied in the following sections.

### 4.1 Pixel-wise statistics

One interesting metric is considering the pixel-value differences between ground truth and predicted maps. We calculate the relative difference  $D_{ij}$  between predicted maps  $\hat{I}_{ij}$  and true maps  $I_{ij}$  as

$$D_{ij,k} = \frac{\hat{I}_{ij,k} - I_{ij,k}}{I_{ij,k}}. \quad (7)$$

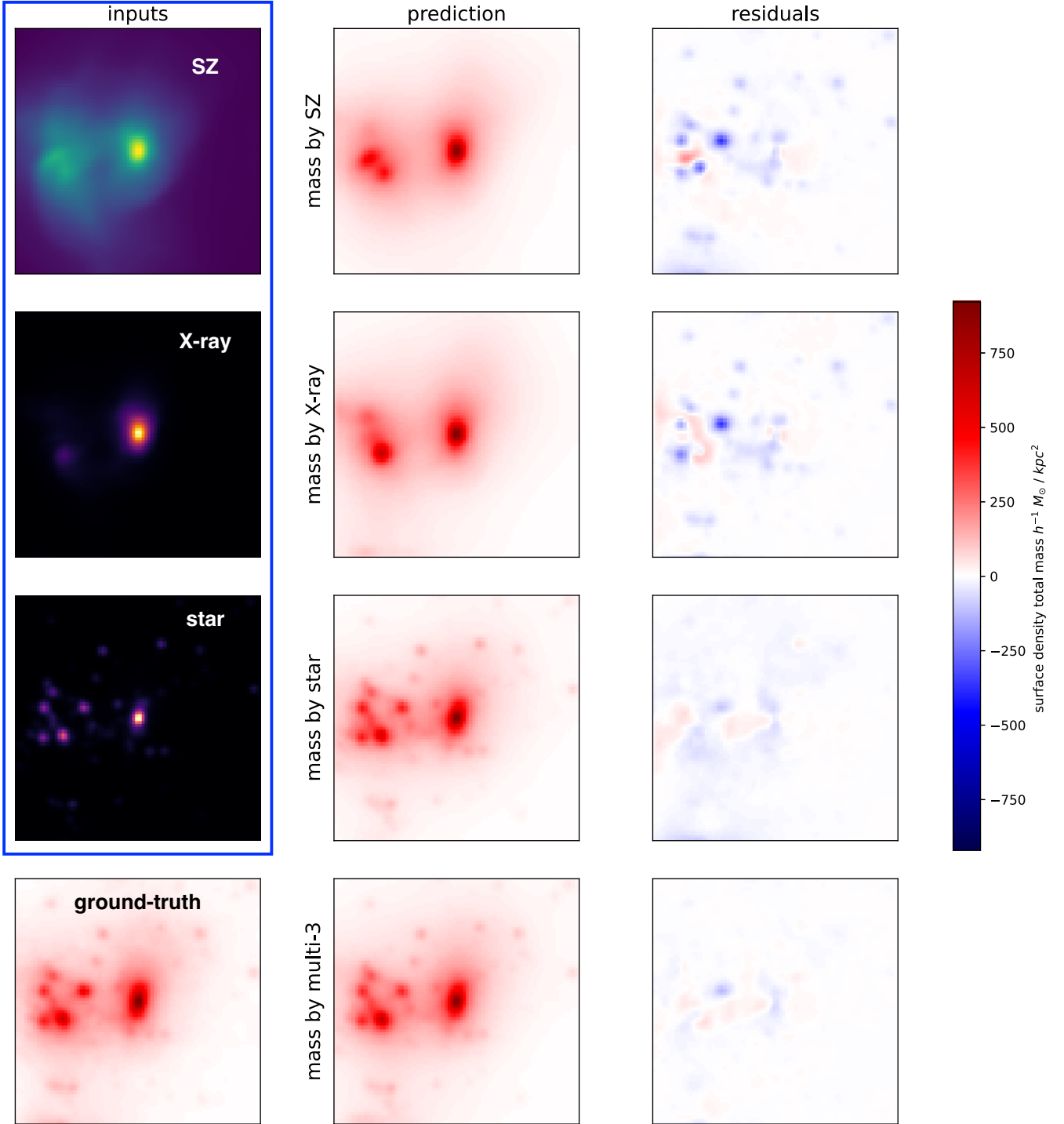
Therefore, the tensor  $D_{ij,k}$  computes the pixel-wise similarity between true images and predicted images. The index  $k$  here runs over maps, i.e. all the maps in the test set. Subsequently, the tensor  $D_{ij,k}$  is flattening to a vector of dimensions  $i \times j \times k$  and its histogram as a probability density is represented in Fig. 6. We drop the values of the tensor  $D_{ij,k}$  where the ground-truth signal  $I_{ij,k}$  equals 0 unless  $\hat{I}_{ij,k} - I_{ij,k} = 0$ .

The particular values of the median, 16th and 84th percentiles of the distributions are also displayed in the legend of Fig. 6 as

$$\text{value}_{-error}^{+error} = \text{median}_{-|\text{median}-16^{\text{th}}|}^{+|\text{median}-84^{\text{th}}|}. \quad (8)$$

In this part, we also study the performance of each input map to infer the predicted mass map at different apertures around the cluster. In the left-hand panel, pixel values inside the whole map are considered, in the middle only the values  $D_{ij,k}$  inside  $R_{500}$  are displayed and in the right-hand panel only values inside  $R_{500}/2$ .

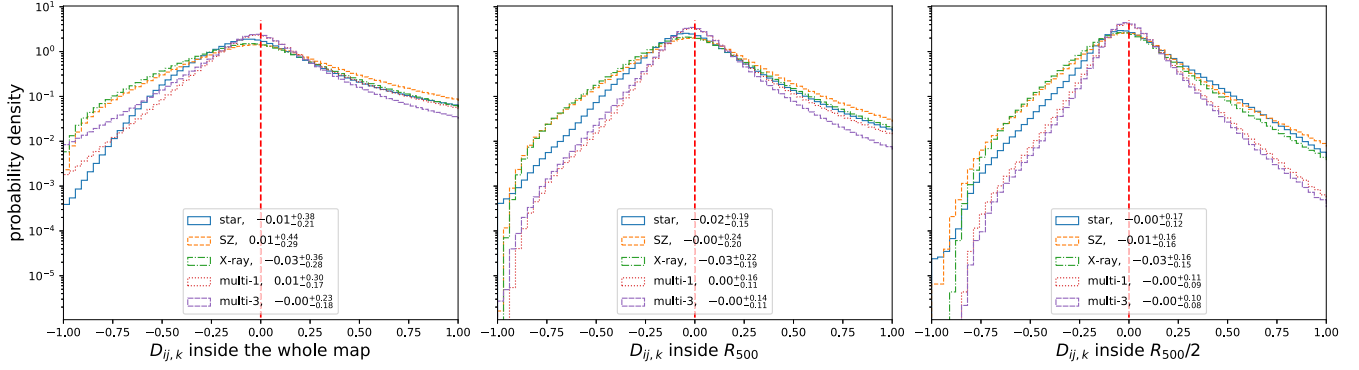
<sup>13</sup><http://hyperopt.github.io/hyperopt/>



**Figure 5.** The first column on the left corresponds to the input maps (SZ, X-ray, and stars) and the ground-truth mass density map. In the second column from top to bottom, we show the mass map predictions of our U-Net when training with SZ, X-ray, star or multiview (multi-3). The residuals are defined as the difference between the prediction and the ground-truth maps. The size of all maps is  $2 \times R_{200}$ .

First, for the three panels in Fig. 6, the same trend is observed: the multiview-3 is the most accurate, followed very closely by the multiview-1, using star density maps as input corresponds to the third most accurate model while considering X-ray and SZ as inputs results in a higher relative error, which is consistent to the expectation from Fig. 5. Furthermore, this can be appreciated in

the legend of Fig. 6 that all the models are slightly biased mostly towards negative values, besides the multiview-3 with median  $\simeq 0.00$  in all three panels. While the model trained with only X-ray images yields the worst with median =  $-0.03$  for all three panels. Moreover, the scatter is smaller for the two multiview approaches, i.e.  $+0.23$ ,  $-0.18$  with only slightly larger in multiview-1 than in multiview-



**Figure 6.** Pixel-wise relative difference  $D_{ij,k}$  (see equation 7) between the predicted mass density maps  $\hat{I}_{ij,k}$  and the ground-truth mass density maps  $I_{ij,k}$ . Different lines represent the use of different input views to predict the mass density map: star, SZ, X-ray, multi-1, and multi-3. The legend also shows the median value of the distributions with the 16th and 84th percentiles as median  $_{-|median-16^{\text{th}}|}^{+|84^{\text{th}}-median|}$ . From left to right,  $D_{ij,k}$  is computed for pixels inside various circular apertures: the whole map, inside  $R_{500}$  and inside  $R_{500}/2$ .

3. In contrast, that scatter from the model trained with only SZ images is the largest,  $+0.44$ ,  $-0.29$ , followed by the X-ray model,  $+0.36$ ,  $-0.28$ , which is also closer to the values from the star model,  $+0.38$ ,  $-0.21$ . As it is depicted in Fig. 5, models whose input is either SZ or X-ray fail at predicting small-scale substructures and the signal vanishes in the outskirts of the clusters. Therefore, all those not predicted substructures contribute to negative values in the  $D_{ij,k}$  distribution. Subsequently, these substructures in the density mass maps are, allegedly, wrongly estimated. This will be quantified in the subsequent sections.

By comparing with the middle and right-hand panels of Fig. 6, we aim to test whether the asymmetry in the distributions of the relative difference tensor  $D_{ij,k}$  is mainly affected by pixels outside the central region of the maps and whether the model behaves worse in the cluster centre or not. As depicted in that figure, the distributions progressively become less spread and more symmetrical when considering only pixels inside a smaller aperture,  $R_{500}$  and  $R_{500}/2$ . This means that the central regions, which we care the most, actually have the best result. Nevertheless, the multiview models provide the most accurate result in all cases, albeit a slight improvement with multiview-3 than multiview-1.

## 4.2 Cylindrical radial profiles

One can assess the quality of the predictions by comparing the mass radial profiles of the predicted and the ground-truth mass density maps. The mass profiles  $M_\rho$  are defined by integrating the mass density, or pixel values, of the maps in different circular regions  $\Omega_\rho$  with radius  $\rho$ :

$$M_\rho = \sum_{i,j \in \Omega_\rho} I_{ij} \Sigma. \quad (9)$$

Note that all circular regions include pixels from the very centre of the maps, i.e. radius  $\in [0, \rho]$ . Here,  $I_{ij}$  are the pixel values and  $\Sigma$  is the pixel size  $-\Sigma = (R_{200}/40)^2$ . Note that equation (9) provides the total 3D mass of a galaxy cluster integrated over a cylindrical volume of  $2 \times R_{200} \times \Omega_\rho$ . We then have computed  $M_\rho$  for ten equally spaced circular regions of radii  $\rho = R_{200}/10, R_{200}/9, \dots, R_{200}$ . Nevertheless, the difference between the ground-truth profile  $M_\rho$  and the predicted profile  $\hat{M}_\rho$  can be estimated as the mass bias  $b_\rho$  as it is computed

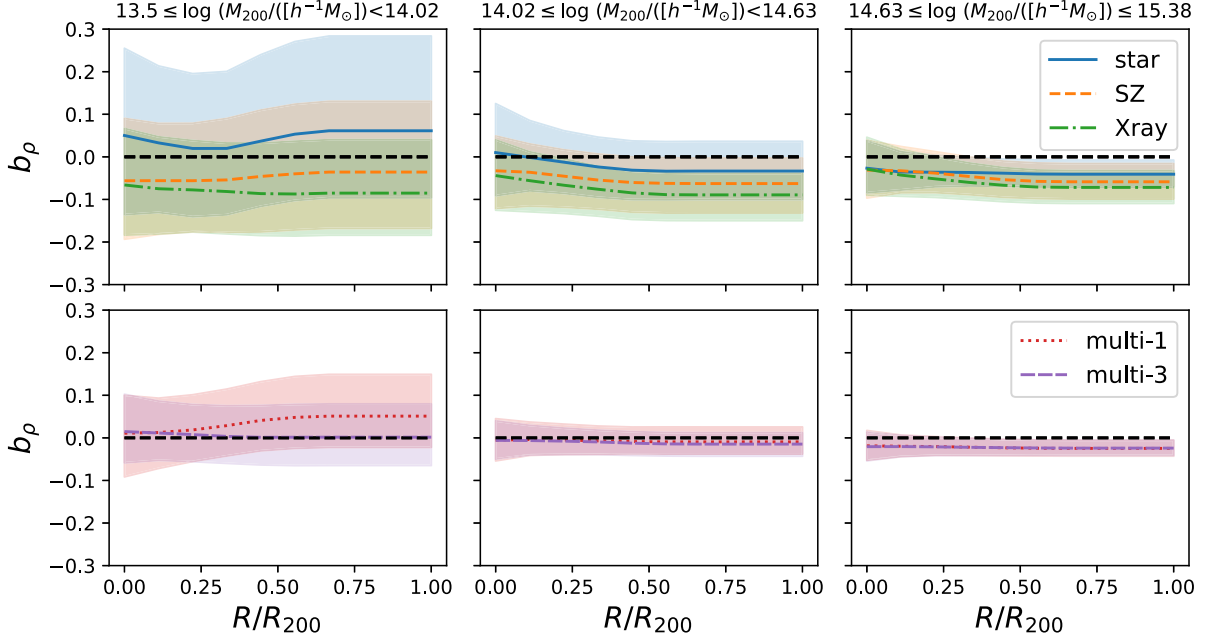
also in de Andres et al. (2022) and Ferragamo et al. (2023):

$$b_\rho = \frac{\hat{M}_\rho - M_\rho}{M_\rho}. \quad (10)$$

In Fig. 7, we show the bias  $b_\rho$  as a function of the normalized radius  $R/R_{200}$ . From left to right, we show the bias that corresponds to three different mass intervals, which divide the data set in a manner that each interval contains roughly the same number of galaxy clusters, see Fig. 1. Therefore, interval 1 corresponds to  $13.5 \leq \log M_{200}/h^{-1}M_\odot < 14.02$ , the interval 2 range is  $14.02 \leq \log M_{200}/h^{-1}M_\odot < 14.63$ , and the interval 3 is  $14.63 \leq \log M_{200}/h^{-1}M_\odot \leq 15.38$ . Moreover, in the top panels, we show the 3 single-view approaches (star, SZ, and X-ray) and the 2 multiview models are shown in the bottom panels. From Fig. 7, we see that all models have negative biases, apart from the low mass interval in which the biases for SZ and X-ray models generally are negative, but positive for star and multi-1 models. For the single view result in the top panels of Fig. 7, negative biases for the results from SZ and X-ray inputs are possibly caused by the fact that they cannot provide any information of substructures, especially relatively low mass ones. The biases for the stellar inputs decrease from positive at the low halo mass bin to negative at the high mass bin. We suspect that is because the contribution of substructures is more significant in haloes with higher mass. Since we train the model regardless of the sample's halo mass, there is also a possibility that ML has to balance the predictions due to different contributions of the sample's halo mass. Nevertheless, it is surprising to see the improvements from multiview inputs shown at the bottom panels of Fig. 7: the median bias is reduced by roughly a factor of  $\sim 1/3$ , while the scatter significantly shirked by almost  $\sim 1/2$  in all three mass bins. For the reader's interest, the particular values for the bias at  $R_{200}$  for different models and mass intervals are shown in Table 4.

The data in Fig. 7 and Table 4 show a general trend that the scatter of the bias decreases with mass regardless of the input images. As can be seen in the table, for interval 1 considering the multiview-3 model, the results are  $b_{200} = 0.001^{+0.078}_{-0.066}$  and for interval 3 is  $0.024^{+0.019}_{-0.017}$ , that is, the width is reduced by a factor of 4. This translates into the effect that the predictions for massive clusters are more precise than those predictions for less massive clusters with smaller scatter. This problem could be caused by different factors:





**Figure 7.** Radial profiles of the mass bias, see equation (10). From left to right, we show the bias corresponding to different mass intervals as indicated in Fig. 1: Interval 1 corresponds to  $13.5 \leq \log M_{200}/h^{-1}M_{\odot} < 14.02$ , the interval 2 range is  $14.02 \leq \log M_{200}/h^{-1}M_{\odot} < 14.63$ , and the interval 3 is  $14.63 \leq \log M_{200}/h^{-1}M_{\odot} \leq 15.38$ . The top panels show the single-input view models (star, SZ, and X-ray) and the bottom panels the multiview-1 and -3 models. The lines represent the median values per bin and the shaded regions cover the 16th and 84th percentiles. Furthermore, the particular enclosure mass bias at  $R_{200}$  is presented in Table 4.

**Table 4.** Values for the bias  $b_{200}$  defined in of equation (10) at  $R_{200}$  for our different models. These values are also displayed in Fig. 7 and the intervals 1, 2, and 3 are defined to separate the data set in roughly equal number of clusters, as shown in Fig. 1. Consequently, interval 1 corresponds to  $13.5 \leq \log M_{200}/h^{-1}M_{\odot} < 14.02$ , the interval 2 range is  $14.02 \leq \log M_{200}/h^{-1}M_{\odot} < 14.63$ , and the interval 3 is  $14.63 \leq \log M_{200}/h^{-1}M_{\odot} \leq 15.38$ .

| Model   | Interval 1                 | Interval 2                 | Interval 3                 |
|---------|----------------------------|----------------------------|----------------------------|
| Star    | $-0.061^{+0.222}_{-0.158}$ | $-0.033^{+0.069}_{-0.066}$ | $-0.041^{+0.033}_{-0.030}$ |
| SZ      | $-0.036^{+0.166}_{-0.132}$ | $-0.063^{+0.060}_{-0.069}$ | $-0.059^{+0.042}_{-0.040}$ |
| X-ray   | $0.085^{+0.125}_{-0.098}$  | $-0.089^{+0.050}_{-0.060}$ | $-0.071^{+0.039}_{-0.038}$ |
| Multi-1 | $0.051^{+0.098}_{-0.073}$  | $-0.009^{+0.034}_{-0.028}$ | $-0.025^{+0.019}_{-0.018}$ |
| Multi-3 | $0.001^{+0.078}_{-0.066}$  | $-0.015^{+0.028}_{-0.026}$ | $-0.024^{+0.019}_{-0.017}$ |

(i) The loss function utilized during training, which is the mean absolute error  $L_1$ , could be inducing the weights of the neural network to be updated to fit only massive clusters. This can be understood by noticing that the gradients of the loss function could be higher for massive clusters. According to this hypothesis, the presence of high-mass clusters in our data set hinders the predictions at low masses. We claim here that this is false as demonstrated in the Appendix Section A, in which the DL model is trained with a different configuration where there is only data corresponding to low masses in the training set.

(ii) ICM and stars follow gravity at higher masses, but at lower masses astrophysical effects are relevant and ICM and star components differ between GADGET-X and GIZMO-SIMBA simulations, increasing the irreducible scatter in the observable–mass relation. Since the training uses both simulation haloes together, this larger difference at lower halo mass introduced such a large scatter.

Furthermore, even with a single simulation, the scatter in halo baryon fractions increases as halo mass decreases which is caused by their intrinsic evolution history (see Cui et al. 2021, for example). This is the most relevant source that induces a mass-dependent scatter, as it is discussed in further detail in the Appendix A.

### 4.3 Power spectrum

Quantifying the information contained in our maps at different spatial frequencies is of great interest because the empirical evidence suggests that substructures tend to be underestimated as observed in Fig. 5, especially for the X-ray and SZ input images. This can be clearly viewed by investigating the power spectrum of the density mass map at small scales. For 2D images ( $I_{nm}$ ) of pixel size  $N \times M$ , the Fourier Transform  $\mathcal{F}_{k_n k_m}$  can be defined as

$$\mathcal{F}_{k_x k_y} = \sum_n^N \sum_m^M I_{ab} \exp \left[ -i2\pi \left( \frac{k_x n}{N} + \frac{k_y m}{M} \right) \right], \quad (11)$$

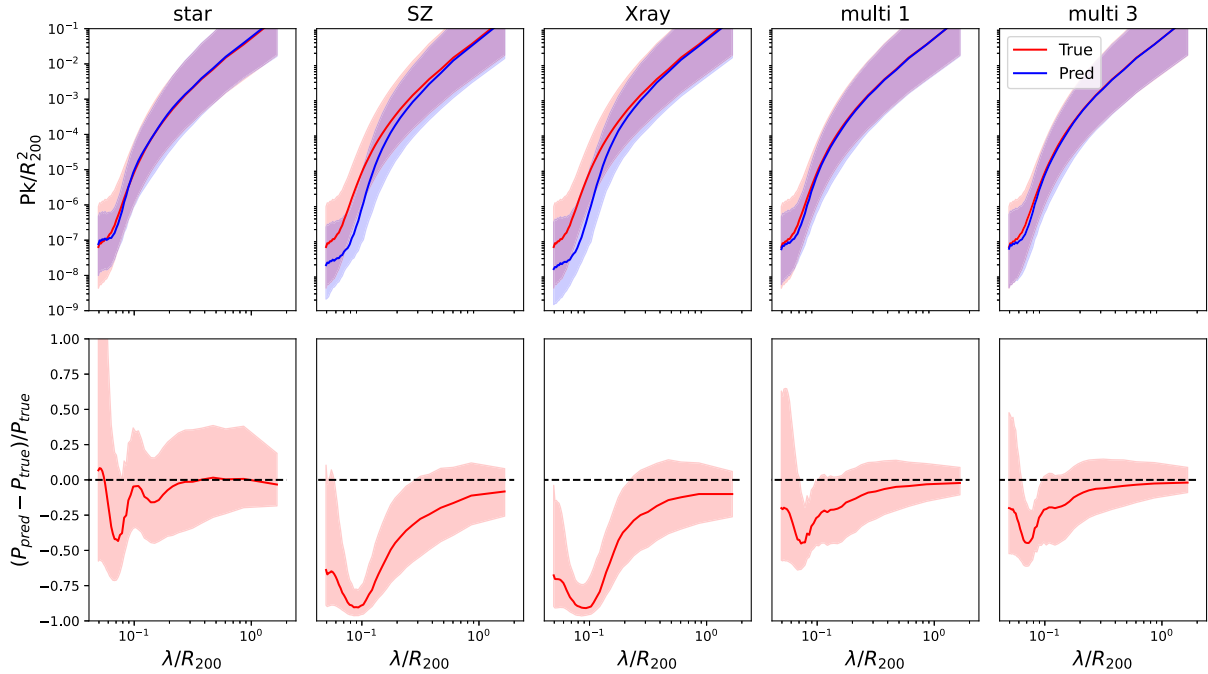
where  $i$  stands here for the imaginary unit. Note that for our images of  $80 \times 80$  pixels  $N = M = 80$ . Moreover, the power spectrum is computed as

$$P_k = |\mathcal{F}_k|^2 \left( \frac{2R_{200}}{80^2} \right)^2 \quad (12)$$

where  $k = \sqrt{k_x^2 + k_y^2}$ . Note that the possible values of  $k$  in pixels go from 1 to the one corresponding to the Nyquist frequency, i.e.  $\sqrt{2} \times N/2 \simeq 56$  pixels, or in physical units with a factor of  $\pi/R_{200}$ . From here we defined the spatial length  $\lambda$  as

$$\lambda = 2\pi/k. \quad (13)$$

Therefore, we further calculate the power spectra of the ground-truth maps and of the predicted maps for comparison, using the



**Figure 8.** **Top:** Power spectrum corresponding to our ground-truth (red) and the predicted (blue) mass density maps as a function of the spatial length  $\lambda = 2\pi/k$  for our different inputs: star, SZ, X-ray, multi-1, and multi-3, are displayed in different columns. **Bottom:** We show the relative difference  $(P_{\text{pred}} - P_{\text{true}})/P_{\text{true}}$  of the predicted power spectrum  $P_{\text{pred}}$  and the ground-truth power spectrum  $P_{\text{true}}$  of our mass density maps. The dashed black line depicts the perfect prediction where the difference is zero. The solid lines correspond to the median values while the shaded regions represent the 16th and 84<sup>th</sup> percentiles.

PYLIANS library (Python libraries for the analysis of numerical simulations; Villaescusa-Navarro 2018), and highlight the difference in Fig. 8. In this figure, the power spectra of the predicted maps and the ground-truth maps are shown in the top panels as a function of the spatial length which is normalized to  $R_{200}$ . In the lower panels, we show the relative difference of the spectrum defined as  $(P_{\text{pred}} - P_{\text{true}})/P_{\text{true}}$ . In this figure, several things can be noted. First, the power spectrum of the predicted mass density maps from star maps is consistent within the errors to the ground truth, where the median power spectrum starts deviating significantly at scales  $0.1 \times R_{200}$ . For ICM tracers, the mass density maps predicted from SZ and X-ray observables indicate that their power spectrum differs on average from the ground-truth one (lower by about 25 per cent at  $\lambda \sim 0.5R_{200}$ ), being only similar at high spatial wavelengths. In addition, in Fig. 8 the advantage of considering multiview models is appreciated, as shown in the fourth and fifth columns in the figure, where the multiview models effectively combine the information available from the different inputs. This means that the median power spectrum is similar to the power spectrum of the predicted mass density maps from stars, but with the clear advantage that the scatter is much smaller. To this end, SZ and X-ray views do not, as expected, improve the predictions of small structures but rather allow the model, when combined with data from stars, to better calibrate the overall signal, reducing the scatter.

#### 4.4 Maximum mean discrepancy

Given two distributions, maximum mean discrepancy (MMD) is a test that assesses whether the two images are the same (Gretton et al. 2008). Although MMD can be used for training generative adversarial neural networks (Karolina Dziugaite, Roy & Ghahramani

2015), it is used here for simply assessing the quality of the predicted mass maps.

The MMD can be defined by choosing a kernel function  $k$  and a pair of random variables of inputs  $X$  and outputs  $Y$ , so that one can compute the MMD as

$$\begin{aligned} \text{MMD}^2(X, Y) = & \frac{1}{m(m-1)} \sum_i \sum_{j \neq i} \mathcal{K}(x_i, x_j) \\ & - 2 \frac{1}{m^2} \sum_i \sum_j \mathcal{K}(x_i, y_j) \\ & + \frac{1}{m(m-1)} \sum_i \sum_{j \neq i} \mathcal{K}(y_i, y_j). \end{aligned} \quad (14)$$

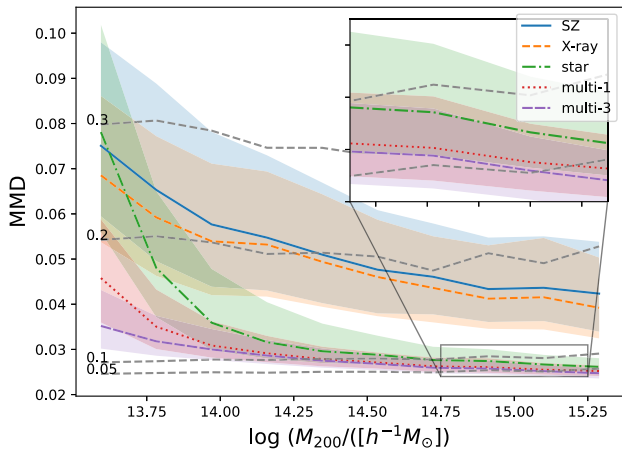
Here,  $x_i$ 's are the ground-truth data points and  $y_i$ 's are the predictions of our U-Net model. For our case, the kernel is Gaussian, and therefore the estimation of  $\mathcal{K}(x, y)$  can be written as

$$\mathcal{K}(x, y) = \exp\left(\frac{-\|x - y\|}{2\sigma^2}\right), \quad (15)$$

where  $\sigma$  corresponds to the band-width range. Typically, one chooses a range of values of  $\sigma$  to evaluate the MMD. In our case, we have computed the MMD as the maximum value estimated by equation (14) taking into consideration the following values of  $\sigma = [0.01, 0.1, 0.25, 0.5, 0.75, 1.0, 2.5, 5.0, 7.5, 10.0]$ , defined in equation (15).

Though the MMD might be very useful, it has little physical meaning and needs to be calibrated. To overcome this issue, we have computed the MMD between the ground-truth mass density map  $I_{ij}$  and a perturbed map  $P_{ij}$  that is created by adding a random Gaussian noise such that

$$P_{ij} = I_{ij} + \mathcal{N}(0, \max(I_{ij}) \cdot \alpha). \quad (16)$$



**Figure 9.** MMD, see equation (14), between predicted mass density maps and the ground-truth mass density maps as a function of the cluster mass  $M_{200}$ . Different colours represent the predictions when the model is trained with only SZ data (blue), X-ray data (orange), stellar data (green), multi-1 (red), and multi-2 (purple). Horizontal grey dashed lines represent the calibration of the MMD values, equation (16), and numbers written in black colour on the left of these lines correspond to the noise intensity  $\alpha$  used for the calibration of the MMD values. The inset highlights the results at higher halo mass where the multiview results are compatible with  $\alpha \sim 0.05$ .

Here  $\alpha$  corresponds to the parameter that quantifies the standard deviation of the noise map. In Fig. 9, we have used  $\alpha = 0.05, 0.1, 0.2, 0.3$  as shown in horizontal grey dashed lines, which represent the calibration of the MMD. This means that the standard deviation of the noise map is at most 30 per cent of the peak value.

Subsequently, we have calculated the MMD between the ground-truth maps  $I$  and the predicted maps  $\hat{I}$ . In Fig. 9, we show the median values of the MMDs as a function of the mass of the cluster that corresponds to our different models. As a general result, the MMD metric is in agreement with the other previous investigations in the sense that the multi-3 model is the closest to the ground truth. We also find a mass dependence of the MMDs values with respect to the total mass of the galaxy clusters, which is coherent with the results in Section 4.2 and it is explained in detail in the Appendix A. Moreover, the noise calibrations, which are represented in dashed grey lines, show that for small groups of clusters, the discrepancy is equivalent to having less than 30 per cent of noise in the maps. However, in the massive end of our data set around  $10^{15} h^{-1} M_{\odot}$ , MMDs are below 10 per cent for all the models (SZ, X-ray, star, multi-1, and multi-3) and very close to 5 per cent in case of the multiview models. This result suggests the use of multiview data to boost the statistical agreement of the pixel’s distribution in the generated total mass density maps.

## 5 CONCLUSIONS

DL models have been used for generating images (see Rothschild et al. 2022, for example for the application in astronomy). Building on the success in the development of ML in this field, we propose, for the first time, to expand its application to predicting the total matter density distributions using multiview simulated observational data: star, SZ, and bolometric X-ray. The DL architecture used on the project is based upon the U-Net, which was introduced in the context of biomedical imaging and it has been modified for accounting for multiview input as described in Section 3.1. As the first step along this research line, we validated the applicability of this ML model using a

data set of simplified maps, described in Section 2, from THE THREE HUNDRED set of hydrodynamic simulations. This data set is free from instrumental and observational effects so that noise, point sources, and telescopes’ impact are not considered. In this work, we tested different convolutional U-Net architectures for both single-input and multi-input models and applied different matrices to quantify the fidelity of the predicted matter density maps from this DL model. From the results, we can conclude that, although the ML output is much more complex – 2D versus 1D or a single data point ( $M_{200}$ ) – compared to previous ML models, its outcomes, for example the 1D mass bias profile, are much better and very promising, especially for the multiview inputs. In detail, we can summarize our results in the following:

- (i) The scatter in the estimation of the mass maps from multiview is reduced by a factor of  $\simeq 1/2$ . As depicted in Fig. 6, the pixel statistics show that the scatter can be reduced from, e.g.  $\pm \sim 16$  per cent when inferring the mass from SZ and  $\pm \sim 10$  per cent when using the combination of inputs maps using the multiview-3 model. Note that the scatter is defined as the deviation from the median value using the 16th and 84th percentiles as it is written in equation (8). This fact is also acknowledged when examining the mass profiles in Fig. 7 in which the scatter can be reduced from  $\pm \sim 4$  per cent to  $\pm \sim 2$  per cent, see Table 4 for further details.
- (ii) Each input view (see Fig. 5) distinctly correlates with the output mass map and therefore, the capability of predicting different spatial frequencies is supreme for the multi-input models, as it is manifested in Fig. 8.
- (iii) By computing the MMD values defined in equation (14), we have examined the statistical similarity of the pixel distributions between the predicted maps and the true maps. The results suggest again the multiview models are foremost and they can be below 5 per cent noise level as shown in Fig. 9.
- (iv) Overall, the multiview models provide the best predictions in all the matrices: very low bias with a small scatter, especially in the massive regime. Multi-3 shows a little improvement compared with multi-1 because using 3 different encoders tends to better capture the important features of each input image.

It is interesting to also assess what is the improvement in the mass reconstruction when combining only gas inputs, i.e. SZ and X-ray. For that purpose, we have trained a model with two encoders and the results are in accordance with single-input models, but with no observed improvement in the quality of the predicted mass. Quantitatively, this can be analysed by computing the mass bias in the mass profiles which has roughly the same values as considering SZ alone. The performance of the double-input gas model seems to be not comparable with the benefits of including information from both tracers, e.g. star and gas together. The star model trained with star density maps is more precise than the corresponding double-input one with SZ and X-ray.

The results of this work hint that observations of galaxy clusters at different spectral bands will translate into a better estimation of the underlying mass distribution. We have limited ourselves, as a first work on the topic, to 3 input tracers, but our DL approach can be generalized to account for many available inputs at different wavelengths, e.g. the several luminosity bands of the SDSS survey. The accessible computational resources are the only limitation, such as GPU memory.

Different DL models could also be considered as an additional job. Though not shown in this paper, we have considered modifying the loss function in the framework of generative adversarial networks (Goodfellow et al. 2014), and training the Wasserstein GAN (Arjovsky, Chintala & Bottou 2017) with no improvement over using

only the MAE loss function. Therefore, at the moment, the U-Net model presented in this work is, the most suitable for addressing the problem of the inference of the mass map from multiview simulated observations. Conversely, vision transformers (Dosovitskiy et al. 2020) might yield competitive results in comparison with convolutional networks.

These techniques that are based on image-to-image translation algorithms (Isola et al. 2016) can also be applied to painting baryons on to  $N$ -body simulations (Chadayammuri et al. 2023), which we will also consider them for generating baryon maps from DM-only simulations, increasing the number of maps already available in simulations.

The results of this project represent the required proof-of-concept step towards the estimation of mass density maps from real observational data. This constitutes a challenge in the sense that our data considered here for training the models could not completely match the characteristics of real data due to several regards. First, our models can be used in real data training on mock observational data where the instrumental impacts are considered in the framework of simulation-based inference. As achieved in our previous work, observational mock data was created by introducing directly the instrumental effects on our simulated clean maps (de Andres et al. 2022). Secondly, the physics implemented in our simulations could lead to biased results. To address this problem, DL models could be trained with data provided from simulations with various and plausible physical models (Villaescusa-Navarro et al. 2021).

## ACKNOWLEDGEMENTS

We would like to express our gratitude to the anonymous referee for the comments that greatly helped us improve the quality of the manuscript. D.d.A., W.C. and G.Y. would like to thank Ministerio de Ciencia e Innovación for financial support under project grant PID2021-122603NB-C21. WC is supported by the STFC AGP Grant ST/V000594/1 and the Atracción de Talento Contract no. 2020-T1/TIC-19882 granted by the Comunidad de Madrid in Spain. He also thanks the ERC: HORIZON-TMA-MSCA-SE for supporting the LACEGAL-III project with grant number 101086388 and the China Manned Space Project for its research grants. MDP and AF acknowledge support from Sapienza Università di Roma thanks to Progetti di Ricerca Medi 2022, RM1221816758ED4E. The authors acknowledge The Red Española de Supercomputación for granting computing time for running the hydrodynamic simulations of The300 galaxy cluster project in the Marenostrum supercomputer at the Barcelona Supercomputing Center.

## DATA AVAILABILITY

The results shown in this work use data from The THREE HUNDRED galaxy clusters sample. The data is freely available upon request following the guidelines of The Three Hundred collaboration, at <https://www.the300-project.org>.

## REFERENCES

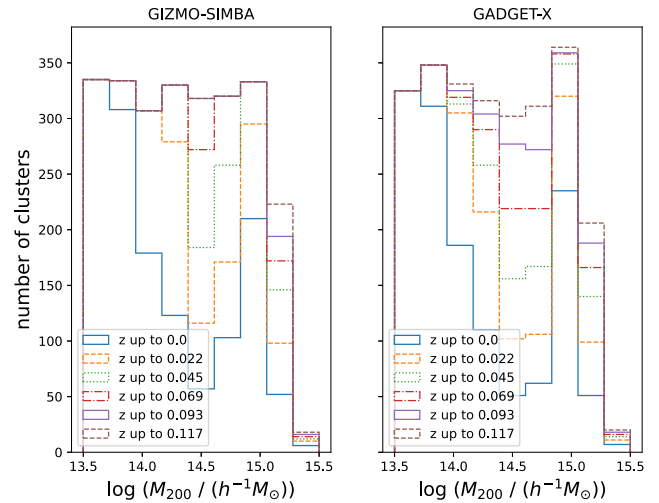
- Akeret J., Chang C., Lucchi A., Refregier A., 2017, *Astron. Comput.*, 18, 35
- Allen S. W., Evrard A. E., Mantz A. B., 2011, *ARA&A*, 49, 409
- Angora G. et al., 2020, *A&A*, 643, A177
- Aragón-Calvo M. A., 2019, *MNRAS*, 484, 5771
- Arjovsky M., Chintala S., Bottou L., 2017, preprint (arXiv:1701.07875)
- Becker M. R., Kravtsov A. V., 2011, *ApJ*, 740, 25
- Berger P., Stein G., 2019, *MNRAS*, 482, 2861
- Bergstra J., Bardenet R., Bengio Y., Kégl B., 2011, *Adv. Neural Inform. Process. Syst.*, 24
- Bergstra J., Yamins D., Cox D., 2013, *International Conference on Machine Learning*, p. 115
- Biviano A., Murante G., Borgani S., Diaferio A., Dolag K., Girardi M., 2006, *A&A*, 456, 23
- Bleem L. E. et al., 2020, *ApJS*, 247, 25
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- CHEX-MATE Collaboration, 2021, *A&A*, 650, A104
- Chadayammuri U., Ntampaka M., ZuHone J., Bogdán Á., Kraft R. P., 2023, *MNRAS*, 526, 2812
- Cui W. et al., 2018, *MNRAS*, 480, 2898
- Cui W., Davé R., Peacock J. A., Anglés-Alcázar D., Yang X., 2021, *Nat. Astron.*, 5, 1069
- Cui W. et al., 2022, *MNRAS*, 514, 977
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- de Andres D. et al., 2022, *Nat. Astron.*, 6, 1325
- de Andres D., Yepes G., Sembolini F., Martínez-Muñoz G., Cui W., Robledo F., Chuang C.-H., Rasia E., 2023, *MNRAS*, 518, 111
- De Luca F., De Petris M., Yepes G., Cui W., Knebe A., Rasia E., 2021, *MNRAS*, 504, 5383
- Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, *A&A*, 350, 381
- Dosovitskiy A. et al., 2020, preprint (arXiv:2010.11929)
- Ferragamo A. et al., 2023, *MNRAS*, 520, 4000
- Gianfagna G. et al., 2021, *MNRAS*, 502, 5115
- Gianfagna G., Rasia E., Cui W., De Petris M., Yepes G., Contreras-Santos A., Knebe A., 2023, *MNRAS*, 518, 4238
- Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, preprint (arXiv:1406.2661)
- Gretton A., Borgwardt K., Rasch M. J., Scholkopf B., Smola A. J., 2008, preprint (arXiv:0805.2368)
- Gupta N., Reichardt C. L., 2020, *ApJ*, 900, 110
- Gupta N., Reichardt C. L., 2021, *ApJ*, 923, 96
- Hausen R., Robertson B. E., 2020, *ApJS*, 248, 20
- Herbonnet R. et al., 2022, *MNRAS*, 513, 2178
- Hilton M. et al., 2018, *ApJS*, 235, 20
- Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R. R., 2012, preprint (arXiv:1207.0580)
- Ho M., Rau M. M., Ntampaka M., Farahi A., Trac H., Póczos B., 2019, *ApJ*, 887, 25
- Ho M., Farahi A., Rau M. M., Trac H., 2021, *ApJ*, 908, 204
- Ho M., Ntampaka M., Rau M. M., Chen M., Lansberry A., Rühle F., Trac H., 2022, *Nat. Astron.*, 6, 936
- Ho M., Soltis J., Farahi A., Nagai D., Evrard A., Ntampaka M., 2023, *MNRAS*, 524, 3289
- Hong S. E., Jeong D., Hwang H. S., Kim J., 2021a, *ApJ*, 913, 76
- Hong S. E., Park S., Jee M. J., Bak D., Cha S., 2021b, *ApJ*, 923, 266
- Huertas-Company M., Lanusse F., 2023, *PASA*, 40, e001
- Ioffe S., Szegedy C., 2015, preprint (arXiv:1502.03167)
- Iqbal A. et al., 2023, *A&A*, 679, A51
- Isola P., Zhu J.-Y., Zhou T., Efros A. A., 2016, preprint (arXiv:1611.07004)
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- Karolina Dziugaite G., Roy D. M., Ghahramani Z., 2015, preprint (arXiv:1505.03906)
- Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Knollmann S. R., Knebe A., 2009, *ApJS*, 182, 608
- Kodi Ramanah D., Wojtak R., Ansari Z., Gall C., Hjorth J., 2020, *MNRAS*, 499, 1985
- Kodi Ramanah D., Wojtak R., Arendse N., 2021, *MNRAS*, 501, 4080
- Kravtsov A. V., Borgani S., 2012, *ARA&A*, 50, 353
- Krippendorf S. et al., 2023, preprint (arXiv:2305.00016)
- Lauritsen L., Dickinson H., Bromley J., Serjeant S., Lim C.-F., Gao Z.-K., Wang W.-H., 2021, *MNRAS*, 507, 1546
- Liu A. et al., 2022, *A&A*, 661, A2

- Milletari F., Navab N., Ahmadi S.-A., 2016, preprint (arXiv:1606.04797)
- Murante G., Monaco P., Giovalli M., Borgani S., Diaferio A., 2010, *MNRAS*, 405, 1491
- Nair V., Hinton G. E., 2010, Wrobel S., et al., Proceedings of the 27th international conference on machine learning (ICML-10). Omnipress, Madison, WI, USA, p. 807
- Ntampaka M. et al., 2019, *ApJ*, 876, 82
- Odehahn S. C., Stockwell E. B., Pennington R. L., Humphreys R. M., Zumach W. A., 1992, *AJ*, 103, 318
- Planck Collaboration XIII, 2016a, *A&A*, 594, A13
- Planck Collaboration XXIV, 2016b, *A&A*, 594, A24
- Planck Collaboration XXVII, 2016c, *A&A*, 594, A27
- Pratt G. W., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T. H., 2019, *Space Sci. Rev.*, 215, 25
- Rasia E. et al., 2015, *ApJ*, 813, L17
- Ronneberger O., Fischer P., Brox T., 2015, in Navab N., Hornegger J., Wells W. M., Alejandro F. F., eds, Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. Springer-Verlag, Munich, Germany, p. 234
- Rothschild T., Nagai D., Aung H., Green S. B., Ntampaka M., ZuHone J., 2022, *MNRAS*, 513, 333
- Rozo E. et al., 2009, *ApJ*, 699, 768
- Salvati L. et al., 2022, *ApJ*, 934, 129
- Scherer D., Müller A., Behnke S., 2010, in Diamantaras K., Duch W., Iliadis L. S., eds, Artificial Neural Networks—ICANN 2010. Springer, Berlin, Heidelberg, p. 92
- Sembolini F., Yepes G., De Petris M., Gottlöber S., Lamagna L., Comis B., 2013, *MNRAS*, 429, 323
- Smith M. J., Geach J. E., 2023, *R. Soc. Open Sci.*, 10, 221454
- Strauss M. A. et al., 2002, *AJ*, 124, 1810
- Sunyaev R. A., Zeldovich Y. B., 1972, *Comment. Astrophys. Space Phys.*, 4, 173
- Sweere S. F. et al., 2022, *MNRAS*, 517, 4054
- Villaescusa-Navarro F., 2018, Astrophysics Source Code Library, record:1811.008
- Villaescusa-Navarro F. et al., 2021, *ApJ*, 915, 71
- Yan Z., Mead A. J., Van Waerbeke L., Hinshaw G., McCarthy I. G., 2020, *MNRAS*, 499, 3445

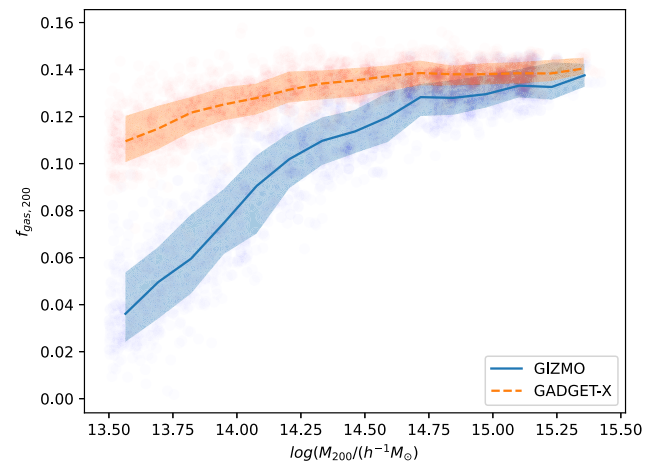
## APPENDIX A: THE DIFFERENCES BETWEEN GIZMO-SIMBA AND GADGET-X

Throughout this work, we have shown that the difference between the ground-truth maps from the predicted maps is dependent on the cluster mass  $M_{200}$  (see Fig. 7 and Fig. 9). We have claimed that this trend is primarily due to the fact that gravity is more relevant for massive galaxy clusters and astrophysical effects implemented in THE THREE HUNDRED simulations, which are different codes for star formation, supernovae and black hole feedbacks, become more relevant in small groups and thus, the scatter of the predicted mass density maps of these small objects increases.

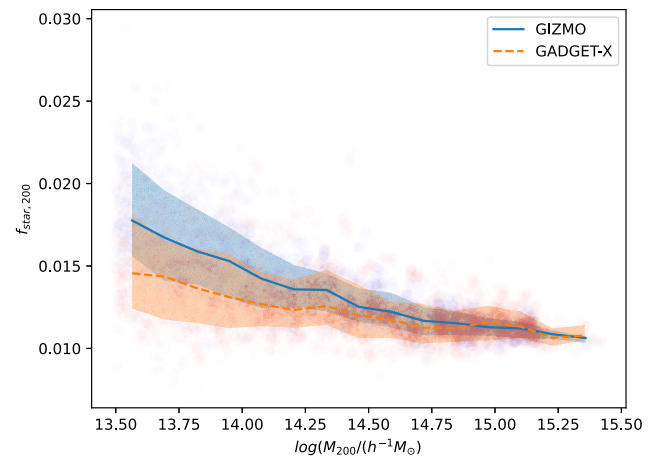
Furthermore, as mentioned in Section 2, we have trained with data from GADGET-X and GIZMO-SIMBA simulations and the CNN model has managed to learn from both simulations simultaneously. The particular distribution of selected objects is shown in Fig. A1. However, both codes generate very different types of galaxy clusters as far as the ICM and star properties are concerned. In Fig. A2, the gas fraction of our data sample is shown for GIZMO-SIMBA (blue) and GADGET-X (dashed orange). The figure implies that the gas fraction for GADGET-X is more constant, while for GIZMO-SIMBA decays faster in small groups. Nevertheless, the star fractions follow the opposite trend, as shown in Fig. A3. In summary, these two figures show that different astrophysical implementations converge in the massive end and therefore, one might expect that the ICM



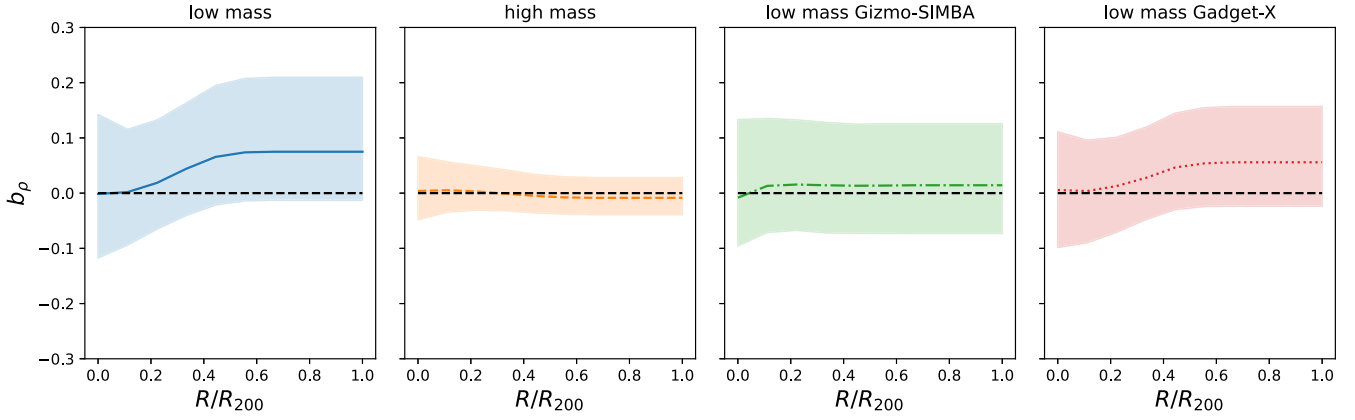
**Figure A1.** Number of selected clusters as a function of mass for our two simulations GIZMO-SIMBA and GADGET-X. The sample is increased when not only snapshots at  $z=0.0$  are considered and objects up to  $z=0.117$  are selected.



**Figure A2.** Median gas fraction as a function of the mass for the GIZMO-SIMBA and GADGET-X simulations. The shaded regions correspond to the 16<sup>th</sup> and 84<sup>th</sup> percentiles.



**Figure A3.** Median star fraction as a function of the mass for the GIZMO-SIMBA and GADGET-X simulations. The shaded regions correspond to the 16<sup>th</sup> and 84<sup>th</sup> percentiles.



**Figure A4.** Mass bias see equation(10) for the mass profiles as a function of the radius  $R/R_{200}$  for the experiments previously mentioned in this appendix. From left to right, we show the bias corresponding to the different experiments: Train with low-mass objects with data from both simulations, train only with high-mass objects with data from both simulations, train low-mass objects with data only from GIZMO-SIMBA and train with low-mass objects with data only from GADGET-X. The lines represent the medium values per bin and the shaded regions cover the 16<sup>th</sup> and 84<sup>th</sup> percentiles.

and star distributions follow gravity, and are more related to the total mass density of the galaxy clusters for massive galaxy clusters.

To show that the scatter does not depend on the ML model, but rather on the diversity of the data, we have run different experiments:

(i) **Low mass:** We have retrained the multi-3 model with galaxy clusters belonging only to the first mass interval, i.e.  $13.5 \leq \log M_{200}/h^{-1}M_{\odot} < 14.02$ .

(ii) **High mass:** We have also retrained the model with data belonging only to the most massive interval ( $14.63 \leq \log M_{200}/h^{-1}M_{\odot} \leq 15.38$ ).

(iii) **Low mass GIZMO-SIMBA:** We have retrained the model multi-3 with data belonging only to GIZMO-SIMBA only in the first mass interval.

(iv) **Low mass Gadget-X:** We have also retrained the multi-3 model with data belonging to GADGET-X only in the first mass interval.

The results of the experiments are shown in the Fig. A4. From the first and second experiments, we conclude that the scatter does not depend on the ML model and training procedure. We acknowledge that although these could bias the prediction to be more accurate at higher masses, that is not the case as the experiments suggest. From the third and fourth experiments, the results on GIZMO-SIMBA and GADGET-X simulations are similar and therefore, we conclude that the performance should not vary significantly for different hydrodynamic simulations.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.