

# Deep learning calibration of galaxy models within a cosmological context

Ángel Chandro Gómez

Directora

**Violeta González Pérez**

Tutora

**Violeta González Pérez**

Trabajo de Fin de Master en Física Teórica  
Especialidad Astrofísica

Año académico 2021-2022

## Abstract

The enormous observable Universe requires of large simulations to attempt to interpret it. In our current best cosmological model, dark matter (DM), which only interacts gravitationally, makes over 80% of the total matter content of the Universe. Therefore, we need a connection between observables (as for example galaxies) and the DM modelled on large scale state-of-the-art simulations. Here we make use of a semi-analytic model (SAM) of galaxy formation and evolution to make this connection, GALFORM run on the dark matter only UNIT simulation. These type of models encode the physical processes that affect the evolution of normal matter or baryons into sets of differential equations with a total of  $\sim 20$  free parameters. In this work we attempt to calibrate the model only varying 3 of the parameters that control the gas cycle. Our aim is to do this in a systematic way in order to measure the probability distribution of the varying parameters, being able to study potential degeneracies. For this purpose we train a deep learning technique that emulates, at least, the SAM luminosity functions in the K and  $b_J$  bands at  $z = 0$ . The main conclusion reached is that the degeneracies among the selected parameters cannot be broken by adding the star formation rate function and the passive fraction to the training set. We thus conclude that the relevant physical processes affecting these two observables were already included in the luminosity functions.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methods I: simulation</b>	<b>2</b>
2.1	UNIT . . . . .	3
2.2	ROCKSTAR . . . . .	3
2.3	CONSISTENTTREES . . . . .	4
2.4	DHALOS . . . . .	5
2.5	GALFORM . . . . .	7
2.5.1	Parameters set free . . . . .	8
<b>3</b>	<b>Methods II: calibration</b>	<b>10</b>
3.1	Emulation theory . . . . .	10
3.1.1	Emulator configuration . . . . .	10
3.1.2	Emulator training . . . . .	12
3.2	Emulating galaxy model observables . . . . .	14
3.2.1	Parameter space exploration . . . . .	16
<b>4</b>	<b>Results and discussion</b>	<b>16</b>
4.1	Emulator training using luminosity functions . . . . .	16
4.2	Emulator training using luminosity functions and star formation rate function .	17
4.3	Emulator training using luminosity functions, star formation rate function and passive fraction . . . . .	17
4.4	Comparison of emulators performance . . . . .	19
4.5	Fitting the observations with the emulators . . . . .	20
4.6	The best fit models . . . . .	22
<b>5</b>	<b>Summary and conclusions</b>	<b>24</b>
<b>6</b>	<b>Bibliography</b>	<b>26</b>

# 1 Introduction

The observable Universe is vast and currently it is plausible to examine it thanks to the large scale surveys. The interpretation of observations from these large surveys requires the use of mock catalogues, based on computational simulations. These simulations are the only laboratory available for astrophysicists and they describe the evolution of the Universe through physically motivated equations. Computational simulations can be used to explore unconstrained aspects of Astrophysics and Cosmology, and they can also serve to quantify the systematic errors that affect a certain type of observations. In the first case, a direct and careful comparison with observations can determine the validity of a model. In this work we are focusing on aspects related to the formation and evolution of galaxies, which involve the evolution of gas, the formation of stars and the growth of black holes. We will refer to all these components as "baryons", as opposed to dark matter (DM).

In our best cosmological model, the  $\Lambda$ CDM (Planck Collaboration et al., 2020), dark matter makes over 80% of the total matter content of the Universe and the remaining percentage accounts for the baryons. Simulations that only track the evolution of dark matter are less time consuming since DM is easier to be described as it only interacts gravitationally. In such a way, these simulations produce huge volumes of the Universe compared to the one of large surveys. Even though the DM-only simulations are an adequate way to have an overview of the large scale pattern of the Universe since baryons in the end fall into the potential wells generated by DM in what are named as DM halos (gravitationally bound structures that are decoupled from the expansion of the Universe), large scale surveys just detect baryons and not DM. Therefore, a connection between both the observables as e.g. galaxies and DM is vital to study the physical processes that play a role in galaxy formation and evolution.

Galaxies are formed in the gravitational potential overdensities due to dark matter as mentioned before, which we identify as dark matter halos. However, unlike DM-only simulations galaxies are challenging to model since all the physics (the formation of stars, mergers, gas outflows and cooling and growth of black holes, among others) involved in them are stochastic processes in which a large dynamical range of scales intervines in a non trivial way. In addition, the different physical processes interplay with each other in non-linear ways, producing degeneracies among them. There are various possibilities to describe galaxies computationally, ranging from more empirical models (data constrains a specific set of flexible non-physical parameters) as for example the halo occupation models to more physical ones (they directly simulates the physics of galaxy formation) as the hydrodynamical simulations (Wechsler and Tinker, 2018).

One of these models to analyse galaxy properties at large scales are semi-analytic models (SAMs) of galaxy formation and evolution (see Somerville and Davé, 2015; Baugh, 2006, for a review of these models), which allow for studying in detail different physical processes without being computationally expensive. SAMs are a set of coupled differential equations encapsulating all the processes we think relevant for the galaxies. As a result, a SAM is able to describe the evolution of baryons in an approximate way using analytic expressions for all those physical processes important in galaxies. SAMs need to be run onto the DM haloes already modeled by N-body DM only simulations in order to properly track the evolution of gas, stars, BHs and in some cases dust. SAMs are really helpful to compare global properties related to the distribution and statistics of galaxy properties (luminosity, stellar mass, star formation, galaxy sizes, spatial position, velocities, ...) with the results coming from observations so as to test physical models that describe processes taking place in galaxies. Nevertheless, the problem with the SAMs is that the coupled differential equations that encapsulate the baryonic physics

depend on numerous free parameters. SAMs typically have about 20 free parameters for those equations describing the astrophysical processes that affect galaxies. These parameters are a testimony of our ignorance about these processes. Plus, the non-linear interplay among the physical processes produces degeneracies among the parameters giving rise to a very complex multi-dimensional parameter space.

Artificial intelligence techniques such as machine learning and deep learning give us the opportunity to be capable of understanding the probability distribution of the free parameters and explore their possible degeneracies. This systematic analysis of the parameter space is what we have conducted in this project and it has been structured as follows: in Section 2 we introduce the required steps to run a SAM, Section 3 is focused on the SAM parameter space exploration through the development of an emulator. In Section 4 we discuss the most important emulating results; while in the final Section 5 we summarize the project drawing conclusions.

## 2 Methods I: simulation

As explained before, semi-analytic models (SAMs) of galaxy formation and evolution are a set of coupled differential equations that need to be solved on top of a DM-only simulation. But the DM-only simulation itself is not enough and DM gravitational potentials, or haloes, need to be identified across time. The steps needed to have galaxies from a SAM are displayed in Figure 1 and can be summarised as follows:

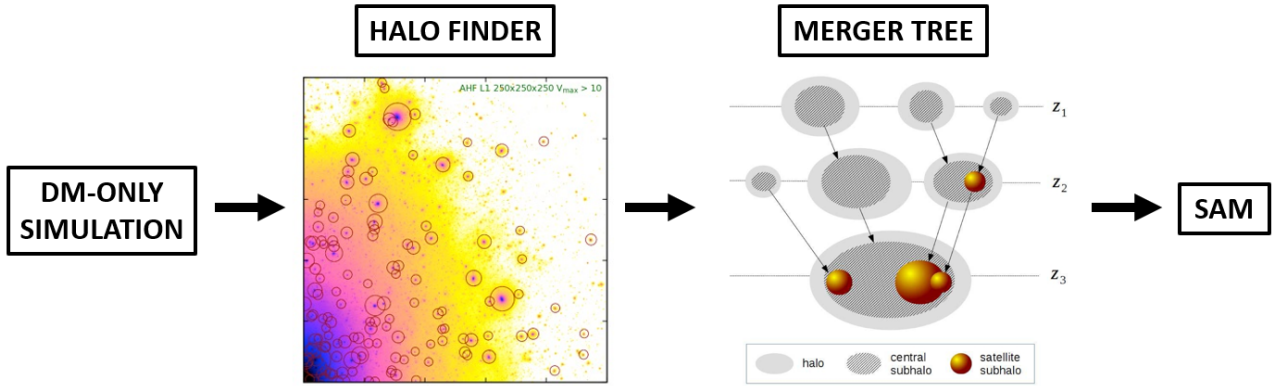


Figure 1: Diagram showing the necessary steps to run a SAM over a DM-only simulation, in chronological order from left to right: run a DM-only simulation, identify DM haloes, establish their evolution through merger trees and running the SAM.

1. Modelling the distribution of DM with a N-body DM-only simulation. In our case we used the UNIT simulation (Chuang et al., 2019).
2. Identifying dark matter haloes or bounded groups of DM particles, using a halo finder code. Here we used ROCKSTAR (Behroozi et al., 2013a).
3. Building the mass accretion history of each halo, using a merger tree code. CONSISTENT-TREES (Behroozi et al., 2013b) and DHALOS (Jiang et al., 2014) were used here.

4. Running a galaxy model using the merger trees. In our case, we have used the SAM GALFORM (Gonzalez-Perez et al., 2020).

In what follows, we explain each of these steps.

## 2.1 UNIT

A DM-only simulation is the only available tool that allow for studying the non linear gravitational collapse in the Universe. DM is driven by the gravitational force, generating overdensities in the Universe due to its collapse. These overdensities become larger and larger with time in a hierarchical way (bottom-up scenario) such that smaller overdense regions merge together generating larger ones. At the beginning of this process the gravitational collapse of matter is in a linear regime which can be described analytically, but when overdensities are large enough the collapse enters the non-linear regime.

DM-only simulations describe the DM density field numerically, so it has to be sampled with a huge number of discrete DM particles/markers inside a finite box. Using comoving coordinates the expansion of the Universe is taken into account, while with periodic boundary conditions the cosmological principle is satisfied. Initially the simulation distributes these DM particles in agreement with the anisotropies from the CMB observations where the gravitational collapse was still linear. After setting the initial conditions, the simulation just traces the DM particle position and velocity evolution with time when the Universe is at the non-linear gravitational regime. The output generated are the DM particle positions and velocities at different timesteps from a certain initial redshift in the past  $z > 0$  to the present time at  $z = 0$ . The larger the number of DM particles, the higher the resolution in mass and it is possible to trace reliably less massive objects. The best would be to describe the largest possible volume to reproduce the volumes observed by large scale surveys, and the largest number of particles to have the highest mass resolution. In our case we have worked with the UNIT simulation (Chuang et al., 2019).

UNIT is a large-scale DM-only simulation of 1Gpc/h as boxsize, which evolves from  $z = 99$  to the present time at  $z = 0$  in 128 timesteps or snapshots using the 2015 Planck Collaboration cosmology (Table 4 in Planck Collaboration et al. (2016)):  $\Omega_m = 0.3089$ ,  $\Omega_\Lambda = 0.6911$ ,  $\Omega_b = 0.0486$ ,  $h = H_0/100 = 0.6774$ ,  $n_s = 0.9667$  and  $\sigma_8 = 0.8147$ . The number of DM particles is  $4096^3$ , thus the mass resolution reached is  $m_{part} = 1.24718 \cdot 10^9 M_\odot/h$ .

## 2.2 ROCKSTAR

The DM particle positions and velocities at different timesteps of the DM-only simulation are not enough to run the SAM, since DM halos need to be located to know where galaxy formation may occur at the different timesteps. The gravitationally DM bound structures formed in a DM-only simulation, which are decoupled from the Hubble expansion thanks to the gravitational collapse, are so-called DM halos. Due to the bottom-up scenario produced by gravity, it is very common to deal with very complex halo structures discovered in DM-only simulations such as gravitationally bound smaller halos inside larger halos. For simplicity, it is usual to handle only a 2-level hierarchy model: structures that are not part of a larger structure are considered simply as host halos (or parent halos), while otherwise if they are part of a larger structure, we refer to them as subhalos (or satellite halos). Halo finder codes, such as ROCKSTAR (Behroozi et al., 2013a), which has been the one we have worked with, allow for locating all of these structures plus their associated properties.

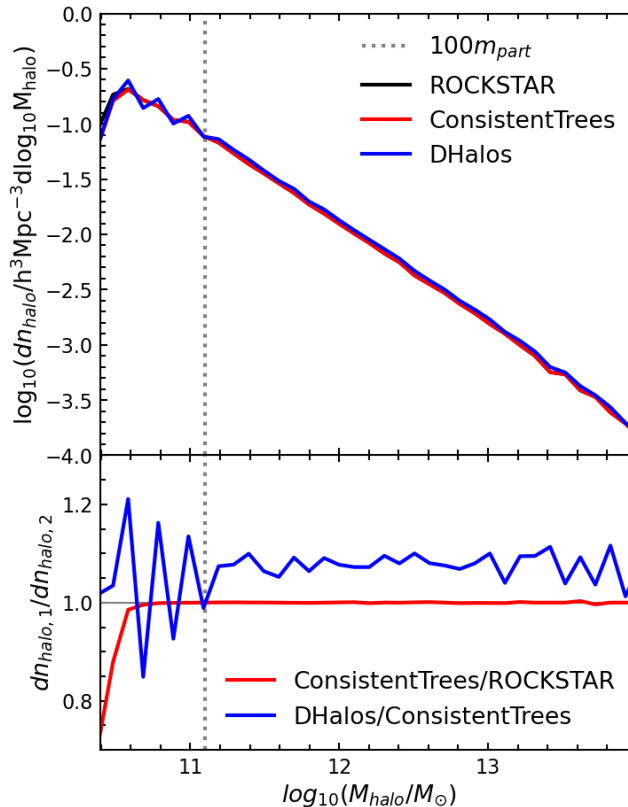


Figure 2: Halo mass function (number density of DM host halos per halo mass bin) at  $z=0$  of a 200 Mpc/h as boxsize subvolume of the UNIT DM-only simulation for the different codes used in this project: the ROCKSTAR halo finder (black line), the CONSISTENTTREES (red line) and the DHALOS (blue curve) merger tree codes. At the bottom panel, it is shown the ratio between the CONSISTENTTREES and ROCKSTAR functions (red curve) and between the DHALOS and CONSISTENTTREES functions (blue curve) to notice the variations produced by each code. The halos considered are composed of more than 20 DM particles, while the dotted grey vertical line indicates halos composed of 100 particles.

### 2.3 CONSISTENTTREES

The connection of DM halos across timesteps also need to be tracked to have not just independent galaxies at different timesteps but, in turn, consistent modeled galaxies with their evolution across timesteps. Following again that DM only interacts gravitationally, the different structures (host haloes and subhaloes) tend to merge with each other into larger ones along time. And not only that, since other situations are possible: a halo may be disrupted and not survived from one timestep to the following one; or a new halo could be formed between timesteps. Therefore, merger tree codes track the evolution of halo structures located by the halo finder assigning progenitors (the halo associated at the previous timestep) and descendants (the halo associated at the next timestep) to the different structures. The merger tree code we have introduced is CONSISTENTTREES (Behroozi et al., 2013b).

CONSISTENTTREES, while building the merger tree for the different halo structures, makes slight variations of the halo catalogues so that every halo at a certain timestep has a progenitor and a descendant as it is expected in the bottom-up scenario of the  $\Lambda$ CDM cosmology, except in the cases of spurious detections of halos or at earlier timesteps when the halos are still small enough to be detectable by the halo finder (there is a DM particle number threshold below which halos are not detected). This very slight variation is visible in Figure 2, where it is observed the halo

mass function (HMF), which describes the number density of host halos per halo mass bin, at  $z=0$  from the ROCKSTAR and the CONSISTENTTREES halo catalogues of the UNIT simulation we have worked with. In particular a subvolume of 200Gpc/h of the simulation. The variation is more prominent for host halos whose particle number approaches the limit of 20 particles in the red curve of the bottom panel that represents the ratio. As it is logical, the larger the host halo, the lower the probability to find a host halo in this specific mass bin, since the halo mass function is equivalent to a probability distribution of the host halo masses. To the left of the vertical line pointing out halos of 100 particles resolution limits start to be noticeable.

## 2.4 DHALOS

Besides the previously listed steps, the merger trees need to be written in a format that is compatible with the SAM code. In the case of GALFORM, to achieve it DHALOS (Jiang et al., 2014) needs to be run. We have executed DHALOS giving it as input the CONSISTENTTREES halo catalogues and generating as output the precise format expected by GALFORM (halo catalogues divided in 64 subvolumes, which has allowed us afterwards to have employed parallel techniques in the HPC system at the UAM to use GALFORM more efficiently). However, the DHALOS code again modifies the halo catalogues e.g. not merging halos with tenuous bridges of particles in common or with an overlap of their outer diffuse haloes, as it can also be seen in Figure 2. Here the cleaning is noticeable with a 10% change in the blue line of the bottom panel between the DHALOS and the CONSISTENTTREES HMFs.

Properties	ROCKSTAR/CONSISTENTTREES
Halo position	average particle location minimizing Poisson error [Mpc/h]
Halo velocity	average particle velocity 10% of the halo radius (peculiar) [km/s]
Halo virial mass	$M_{vir}[\text{CT}] = \frac{4}{3}\pi\Delta_c(z)\frac{\Omega_{m,0}}{\Omega_m(z)}\rho_{crit,0}R_{vir}^3[\text{CT}]$ (only bound particles) [ $M_\odot/h$ ]
Halo half mass radius	$R_{vir}[\text{CT}]/2$ (comoving) [Mpc/h]
Halo particle number	$M_{vir}[\text{CT}]/m_{part}$ (only bound particles)

Table 1: ROCKSTAR/CONSISTENTTREES properties used as DHALOS input.

In order to use DHALOS a connection of halo properties has been carried out between ROCKSTAR/CONSISTENTTREES and DHALOS codes due to the non identical definitions of each code. The most important properties are listed in Table 1 together with the correspondence used. Both positions and velocities are associated directly with the ones used by ROCKSTAR or CONSISTENTTREES, whereas the halo particle number supplied considers only bound particles since the virial mass in CONSISTENTTREES is defined in that way.



With respect to the halo mass, the input for DHalo has been the virial mass in the ROCKSTAR/CONSISTENTTREES catalogues. However, we have needed to understand accurately the possibilities of mass definitions before making a final decision. Commonly the halo masses are defined as the mass of a sphere at which the mean density is  $\Delta_c(z)$  times the critical density at that redshift  $z$ :

$$M_{halo} = \frac{4}{3}\pi\Delta_c(z)\rho_{crit}(z)R_{halo}^3 \quad (1)$$

where  $\rho_{crit}(z)$  is the critical density in the Universe at redshift  $z$  and the usual overdensity  $\Delta_c(z)$  definitions are:

1. A redshift-dependant overdensity computed using the approximation of Bryan and Norman (1998), considering a negligible radiation content in the Universe  $\Delta_c(z) = \Delta_{Bryan}(z) = 18\pi^2 + 82(\Omega_m(z) - 1) - 39(\Omega_m(z) - 1)^2$  where  $\Omega_m(z)$  is the matter content at a certain redshift. This is commonly referred as virial mass  $M_{vir}$ .
2. A constant overdensity of  $\Delta_c(z) = \Delta_{200} = 200$  at whatever redshift. This is usually named  $M_{200c}$ .

However, the ROCKSTAR halo finder and the subsequent CONSISTENTTREES merger tree codes work with particular mass definitions, which have not been found in the literature, but instead have been checked by hand, and they have the following form:

$$M_{vir} = \frac{4}{3}\pi\Delta_{Bryan}(z)\frac{\Omega_{m,0}}{\Omega_m(z)}\rho_{crit,0}R_{vir}^3 \quad (2)$$

$$M_{200c} = \frac{4}{3}\pi\Delta_{200}\frac{\Omega_{m,0}}{\Omega_m(z)}\rho_{crit,0}R_{200c}^3 \quad (3)$$

where  $\rho_{crit,0}$  and  $\Omega_{m,0}$  are the current critical density and matter content in the Universe, while  $\Omega_m(z)$  is the matter content at a certain redshift neglecting again radiation, and the equivalence  $\frac{\Omega_{m,0}}{\Omega_m(z)}\rho_{crit,0} = \frac{\rho_{crit}(z)}{(1+z)^3}$  is valid.

Finally, the most complicated connection has been the half mass radius, which is the radius at which the halo has half of the virial mass, due to the fact that CONSISTENTTREES output does not provide this property and ROCKSTAR only provides it until snapshot 80 of 128 ( $z > 1.8$ ). The half mass radius is decisive for the DHALOS code because it determines which subhalos are within which other host halos. There are no available catalogues with the positions and velocities of the DM particles to compute this property by hand, either. Therefore, we have had to use an approximation as half mass radius. There are various possible candidates to approximate the half mass radius since there are several definitions concerning the radius of DM halos found in DM-only simulations:

1. Virial radius: radius at which the mass is the virial mass following expression (2).
2. 200c radius: radius at which the mass is the 200c mass following expression (3).
3. Klypin scale radius (Klypin et al., 2011): it is the  $R_s$  radius considering a NFW density profile (Navarro et al., 1997) for the halo of the form:

$$\rho(R) = \frac{\rho_s}{x(1+x)^2} \quad (4)$$

where  $x = R/R_s$ , whereas  $\rho_s$  and  $R_s$  vary from halo to halo. The Klypin scale radius is computed starting with the rotational curve for the NFW halo:

$$v_c^2 = \frac{GM_{vir}cf(x)}{r_{vir}xf(c)} \quad (5)$$

where  $c = r_{vir}/r_s$  and  $f(x) = \ln(1+x) - \frac{x}{1+x}$ . It is known that the maximum of the rotational curve is found at  $R_{max} = 2.1626R_s$ . Therefore solving numerically the maximum velocity expression  $v_{max}^2 = v_c^2(x_{max})$  the final  $R_s$  value is reached.

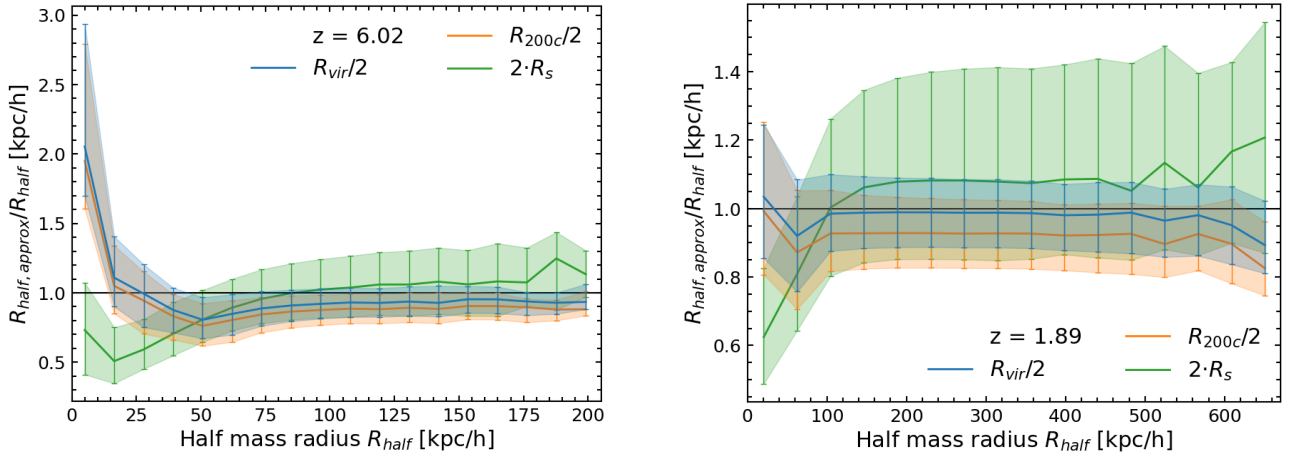


Figure 3: Percentiles of the ratio between the different proposed approximations of the half mass radius ( $R_{vir}/2$  in blue,  $R_{200c}/2$  in orange and  $2R_s$  in green) and the real half mass radius for 2 different snapshots/timesteps: the left panel corresponds to snapshot 40 and redshift  $z=6.06$  while the right panel to snapshot 80 and  $z=1.89$ . The solid central lines represent the median values, while the upper and lower limit of the errorbars are the 84% and 16% respectively.

Knowing every radius definition, the most promising approximations for the half mass radius are:  $R_{vir}/2$  (previously used in Knebe et al. (2015) as half mass radius) and other attempts like  $R_{200c}/2$  or  $2R_s$ . We have computed the percentiles of the ratio between these possible approximations and the real half mass radius (found at some ROCKSTAR snapshots) at different redshifts with some of the results presented in Figure 3. The left and right panels shown are evenly distributed snapshots (snapshot 40 and 80) of the simulation because the objective is to have a proper definition for the whole simulation timesteps. We realised that  $R_{vir}/2$  was the best approximation for at least  $z=1.8$  or greater since the median values (dotted line) are closer to 1 and the dispersion is not reasonable. Although for  $z<1.8$  was not possible to make this comparison, we could extrapolate the conclusions reached.

It is important to say that both GALFORM and DHALOS codes have had to be built at the UAM HPC system overcoming technical difficulties. In addition, the initial idea was to run DHALOS over the whole UNIT simulation of 1Gpc/h. But after various attempts to make it possible, we realised DHALOS code was not adjusted to run over such a huge volume. Instead a subvolume of 200Mpc/h as boxsize had to be run. A treatment of the whole UNIT volume is left as a future improvement, because we would have a much better statistics of the galactic properties of the SAM results using the whole volume. Plus different tests trying to reach the most efficient configuration (nodes, number of cpus and tasks,...) of the UAM HPC system have been developed to run this particular code.

## 2.5 GALFORM

In this study we use the GALFORM semi-analytical model of galaxy formation and evolution that can assign in a physically motivated way galaxies to DM haloes. For instance in Figure 4 we can notice a clear correspondence on how GALFORM galaxies are formed inside the DM

halos from the UNIT simulation, due to the gravity exerted by these DM bound structures. As a starting point for this work, we use a flavour of GALFORM (VGP20) explained in Gonzalez-Perez et al. (2020). The galaxy model described in VGP20 was calibrated using a DM-only simulation with WMAP7 cosmology (Komatsu et al., 2011), which is different from the Planck cosmology assumed in UNIT. The model was calibrated against the following observables at  $z=0$ : the K-band and  $b_J$  band luminosity functions and the super massive black hole mass-bulge stellar mass relation. In this work all the parameters are set equal to those in VGP20, except those 3 we are setting free as described in the next section § 2.5.1. A complete list of the GALFORM free parameters can be found in Lacey et al. (2016). Choosing the values of these free parameters is what we call calibrating the model. What it is usually do is to calculate the specific values of the parameters that reproduce certain observational constraints in order to describe properly the baryonic processes.

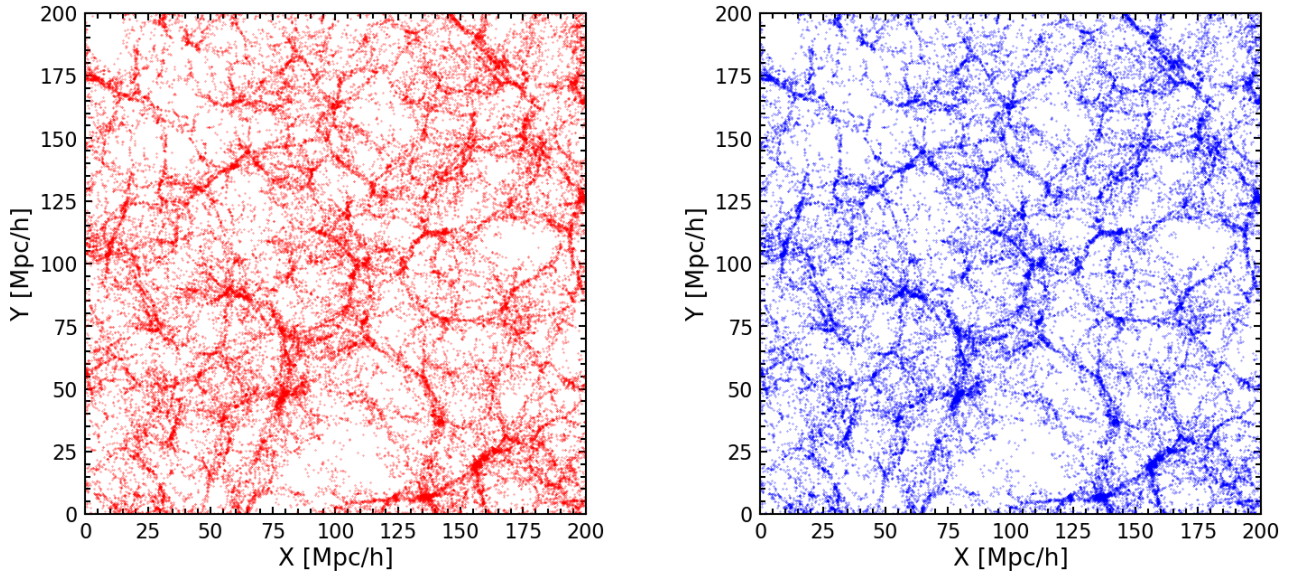


Figure 4: 2D projection (X,Y-directions) of the 200Mpc/h as boxsize subvolume we have worked with in this project (10Mpc/h sliced in the Z-direction). Left panel: distribution of halos composed of at least 20 DM particles obtained from the DHALOS code. Right panel: distribution of galaxies modeled by GALFORM once run over DHALOS whose DM halos are composed of at least 20 DM particles.

### 2.5.1 Parameters set free

The varying parameters chosen are those having a severe effect on the luminosity, because firstly we wanted to obtain reasonable luminosities as it is a fundamental galactic property. In Bower et al. (2010) the most constrained dominant parameters found in the  $b_J$  and K luminosity functions (LFs) are 3: one connected with supernova (SN) ( $V_{SN}$ ) feedback, other with active galactic nucleus (AGN) ( $\alpha_{cool}$ ) feedback and the remaining one with disk instabilities ( $F_{stab}$ ). Other references (e.g. Lacey et al., 2016; Elliott et al., 2021) also consider those 3, thus they are the ones whose degeneracies have been studied in this project.

On the other hand, our objective was to analyse observations at  $z=0$ . Looking for a systematic exploration of the parameter space and to break the degeneracies among those 3 free parameters studied, the first observables to be reproduced are the fundamental local K and  $b_J$  LFs at  $z=0$ ,

as they are affected by the parameters chosen and so as to have the correct galaxy statistics. The observations at these bands suppose the most robust and accurate: the K band ( $\lambda_{eff} = 2.2\mu m$ ) is at the near-infrared and it usually traces the stellar mass (Broadhurst et al., 1992), whereas the  $b_J$  ( $\lambda_{eff} = 4500\text{\AA}$ ) band tracks young stars where star formation is important.

Additional observables are introduced: the star formation rate function (SFRF) (number density of galaxies with a specific star formation rate (SFR), which is the total mass of stars that is formed in a galaxy per unit time) and the observed local fraction of passive galaxies or passive fraction (PF) at  $z=0$  are considered as well. The passive fraction (PF) is the number ratio of quenched or passive galaxies with respect to the total. Galaxies are separated into passive and star-forming depending on their specific star formation rate (sSFR), which is the ratio between the SFR and the stellar mass. Franx et al. (2008) proposed that passive galaxies (considered as quenched) satisfied the condition  $sSFR < 0.3/t_{Hubble}(z)$  ( $t_{Hubble}(z) = 1/H(z)$  is the Hubble time), while on the contrary they are star-forming galaxies.

In Figure 5 the variations that produce different values of the 3 free parameters selected over these 4 observables (K KF,  $b_J$  LF, SFRF and PF) are displayed. Here we explain what is the role of these free parameters in GALFORM e.g. Lacey et al. (2016):

- SN feedback: SN explosions reheat the surrounding gas inhibiting star formation, but these explosions are able to eject the gas from galaxies. The energy injection to the medium is dominated by massive stars (Type II SN), so it can be considered proportional to the instantaneous SFR  $\psi$  with  $\beta(V_c)$  as proportionality constant depending as a power law on the galaxy circular velocity  $V_c$ . The cold gas is ejected to beyond the virial radius of the DM host halo, whose rate of ejection is given by:

$$\dot{M}_{eject} = \beta(V_c)\psi = \left(\frac{V_c}{V_{SN}}\right)^{-\gamma_{SN}} \psi \quad (6)$$

$V_{SN}$  is the normalization factor that regulates the rate of gas ejection due to SNs, while  $\gamma_{SN}$  corresponds to another adjustable parameter that is kept constant. The larger the  $V_{SN}$  parameter, the larger the SN feedback. The SN feedback affects more to the less massive galaxies since their gravitational potential is smaller and the gas is more easily ejected. In Figure 5 when the SN feedback increases luminosities decrease due to the heating of the gas that prevents star formation, which is visible in a lower SFRF and a higher ratio of passive galaxies.

- AGN feedback: supermassive black holes at the center of high massive galaxies accrete gas becoming into AGNs and releasing energy. This energy is able to reheat the surrounding gas, so there is again feedback in this case: in reality it is considered that the energy released by the AGN is ejected through relativistic jets which propagate into the halo and deposit thermal energy in the hot gas. Besides, the feedback is able to balance the radiative cooling if the following condition is satisfied:

$$\tau_{cool}(r_{cool})/\tau_{ff}(r_{cool}) > 1/\alpha_{cool} \quad (7)$$

which means the cooling time of halo gas  $\tau_{cool}$  has to be in the quasi-hydrostatic regime with the free-fall time  $\tau_{ff}$  for the energy released heats the gas effectively. The larger the  $\alpha_{cool}$  parameter, the larger the AGN feedback ( $\alpha_{cool} = 0$  is turning off the feedback); and more galaxies (massive galaxies which have associated AGNs) are affected by it: the star formation activity is inhibited, what involves luminosities decrease and more quenched galaxies appeared. All these effects are visible in Figure 5.

- Disk instabilities: galaxies are able to have morphological transformations if there are disk instabilities, making possible to transform the disk into a bulge. The condition to have instabilities in a disk is that the disk has to be enough self-gravitating, that is:

$$F_{disk} = \frac{V_c(r_{disk})}{(1.68GM_{disk}/r_{disk})^{1/2}} < F_{stab} \quad (8)$$

where  $M_{disk}$  is the total disk mass (stars plus gas) and  $r_{disk}$  the disk half-mass radius. The larger the  $F_{disk}$  value, the more stable the disk is with a lower probability to become a bulge. On the contrary, the larger the  $F_{stab}$  parameter, the less dynamically stable the disk is and more disks would end up as bulges, which have less star formation activity and suppose a decrease in luminosities. These tendencies are observed in Figure 5, where as well a majority of bulges is going to increase the ratio of quenched galaxies.

From Figure 5 it is clear that the variation of the 3 parameters produce the same effects on the observables chosen, which is an indication of the degeneracy among all these parameters. It was already known that the  $\alpha_{cool}$ - $V_{SN}$  pair was strongly constrained (Bower et al., 2010). Moreover, in the VGP20 flavour there are only BHs and, as a result, AGN feedback in bulges. For this reason, we expected that  $\alpha_{cool}$  and  $F_{stab}$  were also degenerate, since in the end they depend on each other.

## 3 Methods II: calibration

The calibration depends on the problem at hand. In general, the calibration is done to reproduce a minimal set of observations that then will allow to have a model to explore the physical problem we are interested in. In this work we would like to explore in detail the probability distribution of the previous 3 parameters, in particular to be able to explore systematically possible degeneracies among them. A huge number of models are to be run in order to do this. Instead of running the SAM model, here we use a deep learning technique to emulate the behaviour of GALFORM in order to get thousands, instead of hundred of models, with the same computational cost. In our case, the running of a single GALFORM model takes about 10 minutes thanks to taking advantage of parallel jobs run at the UAM HPC system. On the other hand, the running of a single emulator model takes of the order of seconds. Therefore, we have to develop an emulator, which helps us introducing a more efficient way of exploration and calibration of the SAM parameter space.

### 3.1 Emulation theory

The emulator is analogous to solving a regression problem of a non-linear function. The generic function takes an input and produces an output and the emulator tries to decode it. But to do it, the emulator has a specific configuration and it needs to follow certain steps. In our particular case, we have applied the emulator configuration and techniques described in Elliott et al. (2021) given the successful performance obtained there. Therefore, we have made use of the same Tensorflow code (Abadi) of Python for developing it computationally.

#### 3.1.1 Emulator configuration

To decode the non-linear function, the emulator strategy is to combine multiple functions to obtain it. Each function is named a "neuron" that has some inputs  $\mathbf{x}$  and produces an output  $z$  based on a commonly non-linear function called "activation function"  $\sigma(p)$  as  $z = \sigma(\mathbf{x} \cdot \mathbf{w} + b)$ ,

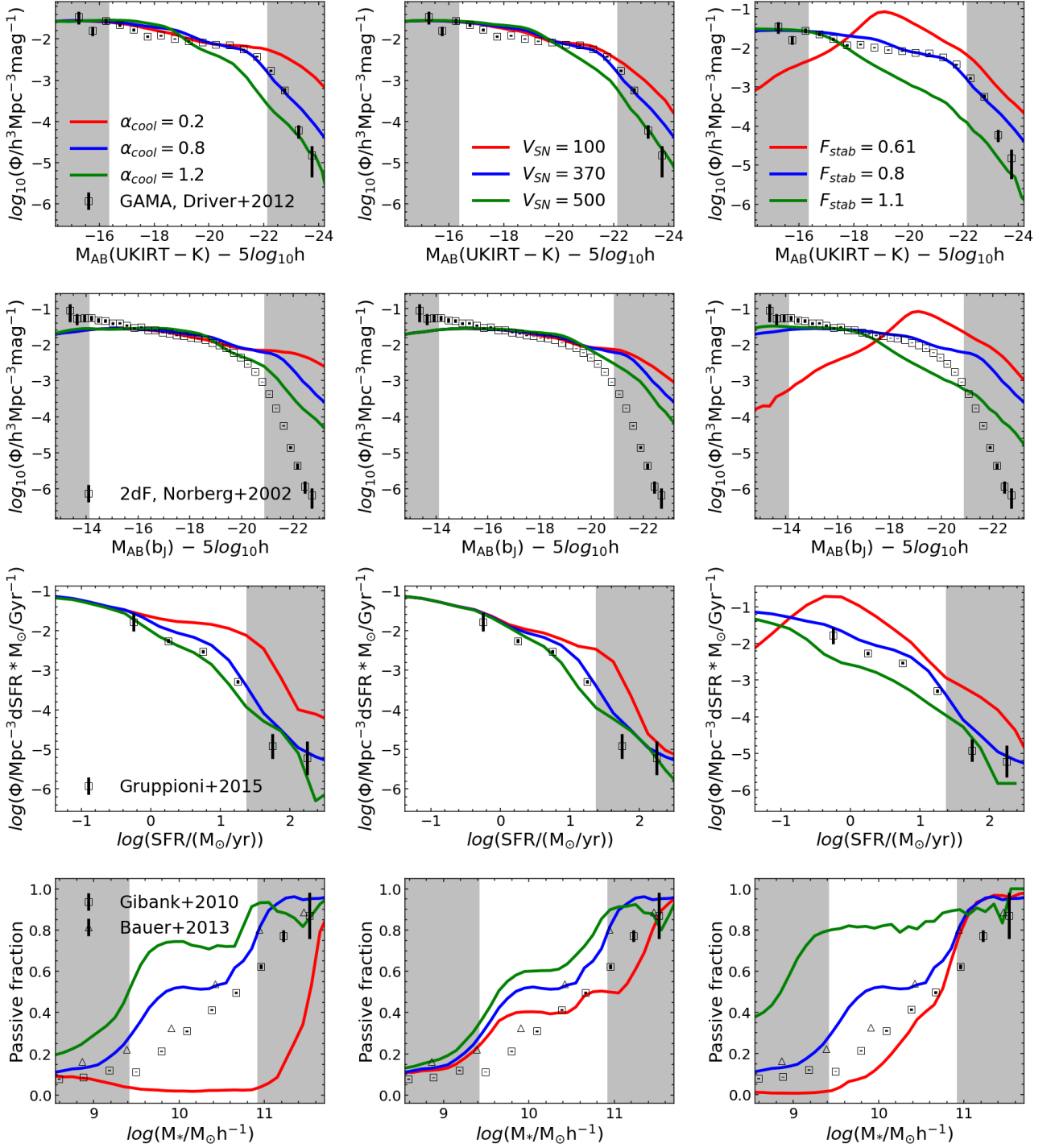


Figure 5: Observational functions (K LF: panels of the first row;  $b_J$  LF: second row panels; SFRF: third row panels; PF: fourth row panels) where we can compare the effect of varying one by one each of the parameters used in the emulators ( $\alpha_{cool}$  variation in the first column,  $V_{SN}$  in the second column,  $F_{stab}$  in the third column). The blue curves correspond in all the panels to the default VGP20 model. All of them with the observations employed to compute the  $\chi^2$  to know the parameter space probability distribution.

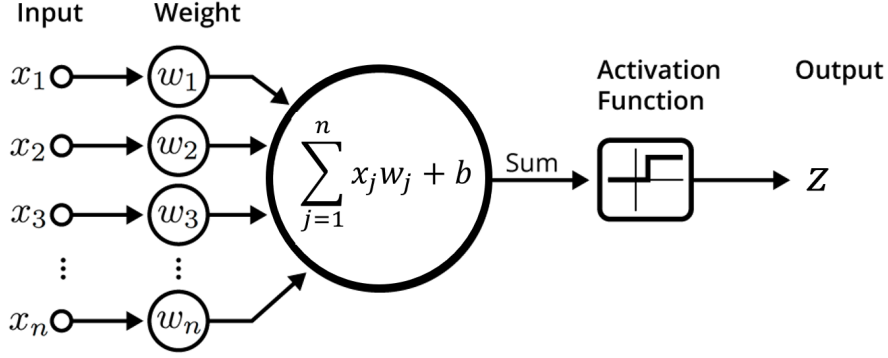


Figure 6: Diagram of the function that represents a neuron in the emulator.

and the variables  $\mathbf{w}$  and  $b$  named "weight" and "bias" respectively are independent of each neuron. A diagram of a neuron is shown in Figure 6. Diverse neurons can be linked forming a network (neural network) increasing the complexity in their non-linear response. These neurons can be grouped together into layers, where the neurons on each layer takes as input the output from the previous layer neurons and the generated output is used as well as input for the next layer neurons. Obviously the neurons of the first layer take as inputs the input supplied to the emulator (input layer), whereas the outputs of the last layer neurons are the outputs generated by the emulator (output layer). Between the input and output layers, we have the so-called hidden layers, whose number determines whether we are applying machine learning (a single hidden layer) or deep learning techniques (more than 1 hidden layer). In addition, when the neuron of each hidden layer is connected with all the neurons of the previous and next hidden layers; this is known as dense layer. As a result, the output of each neuron  $j$  is a bit more complex for a dense layer  $z_j = \sigma(\sum_i w_{ij} \cdot x_i + b_j)$  where  $x_i$  are the values obtained from the  $i$  neurons on the previous layer,  $w_{ij}$  the weights of the neuron  $j$  connecting with the neuron  $i$  in the previous layer and  $b_j$  the bias of the neuron  $j$ . Here we have worked with the Figure 7 configuration: 2 dense hidden layers (deep learning) where each layer is composed of 512 neurons and the activation function of the 2 hidden layers is the sigmoid function of the form  $\sigma(p) = \frac{1}{1+e^{-p}}$  ranging from 0 to 1, while the output layer poses a linear activation function so as not to have restricted values and to obtain the correct outputs.

### 3.1.2 Emulator training

With the general idea of the configuration, the final purpose is to imitate the non-linear function behaviour and this is not possible without real results from it. The emulator is fed with some inputs of the function and the corresponding output generated to find the proper weights and biases of each neuron that decode it. This is known as the "training". The emulator learns how the function exactly works (which output generates) depending on the input introduced and then it is able to interpolate these training data results to unseen datasets. The training consists of 2 steps:

1. The forward propagation where the inputs are given to the emulator and some predicted outputs are generated. Afterwards we compare the predicted outputs and the expected real ones defining the "loss function", whose results would be minimum when the predicted values are equivalent to the expected. In this work as loss function we employ the mean



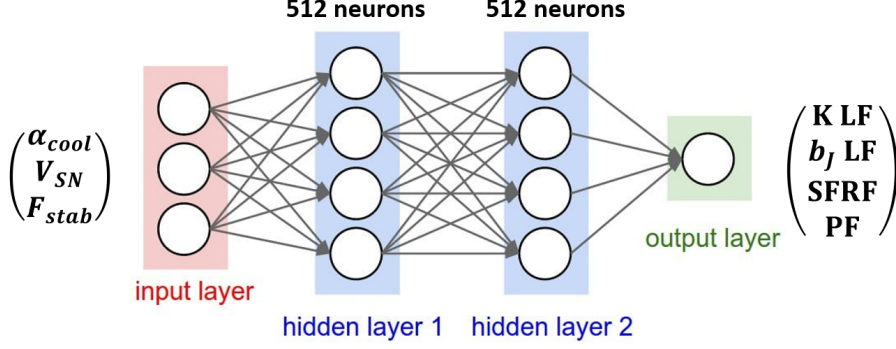


Figure 7: Neural network configuration employed in this project. The emulator mimicks the galaxy model run with a set of parameters on the UNIT DM-only simulation. Note that this simulation, the halo finder and the merger tree construction are fixed. Here we are only training the emulator against different choices of the free parameters described in § 2.5.1.

absolute error (MAE) defined:

$$MAE = \frac{1}{n} \sum_{k=1}^n |F_n(\mathbf{x}_0) - y_n| \quad (9)$$

where  $F_n(\mathbf{x}_0)$  represents the  $n$ -th predicted output from the emulator produced from an input  $\mathbf{x}_0$  and  $y_n$  the real output of the function. The MAE is dimensionless, since it is as dividing  $|F_n(\mathbf{x}_0) - y_n|$  by an error of value equal to the unity. The final objective is to minimise the loss function and the aid of an "optimiser" is fundamental for that. The optimiser supplies the way weights and biases are updated to be directed to the global minimum of the loss function.

2. The second step named as the backpropagation algorithm provides the values of the partial derivatives necessary for using the optimiser and it is based on 4 equations.  $\delta_j^l = \partial MAE / \partial z_j^l$  is the error in the  $j$ -th neuron on the  $l$ -th layer and then  $\delta_j^L$  the error in the output layer. The 4 equations are the following ones and for more detailed information about them we have used this reference (Nielsen, 2015).

$$\delta_j^L = \frac{\partial MAE}{\partial F(\mathbf{x}_0)_j} \sigma'(p_j^L) \quad (10)$$

$$\delta_j^l = \sum_i (w_{ij}^{l+1} \delta_i^{l+1}) \sigma'(p_j^l) \quad (11)$$

$$\frac{\partial MAE}{\partial b_j^l} = \delta_j^l \quad (12)$$

$$\frac{\partial C}{\partial w_{ij}^l} = z_j^{l-1} \delta_i^l \quad (13)$$

where  $l$  is the layer of the specific neuron  $j$ ,  $x_i$  its inputs and  $z_j = \sigma(p_j) = \sigma(\sum w_{ij} x_i + b_j)$  its outputs. The initial weights and biases are randomly chosen.

The Adam optimizer (Kingma and Ba, 2014) has been used, which is efficient for large models and datasets in regression problems. It updates both variables taking into account the first moment (the mean) and the second moment (the uncentered invariance) of the



gradients (partial derivatives of the loss function with respect to weights and biases). An iterative method is carried out for every weight and bias based on the following expressions where  $l$  indicates the layer of the neuron and  $(t)$  the step of iteration ( $m^{l,(0)} = 0, v^{l,(0)} = 0$ ).

$$m^{l,(t)} = \beta_1 m^{l,(t-1)} + (1 - \beta_1) \frac{\partial MAE}{\partial w_{ij}^l} \quad (14)$$

$$v^{l,(t)} = \beta_2 v^{l,(t-1)} + (1 - \beta_2) \left( \frac{\partial MAE}{\partial w_{ij}^l} \right)^2 \quad (15)$$

$$\hat{m}^{l,(t)} = \frac{m^{l,(t-1)}}{(1 - \beta_1^t)} \quad (16)$$

$$\hat{v}^{l,(t)} = \frac{v^{l,(t-1)}}{(1 - \beta_2^t)} \quad (17)$$

$$w_{ij}^{l,(t)} = w_{ij}^{l,(t-1)} - \frac{\eta}{\sqrt{v^{l,(n)}} + \epsilon} \hat{m}^{l,(n)} \quad (18)$$

when  $w_{ij}^{l,(t)}$  converges to a specific value we update the weight  $w_{ij}^l = w_{ij}^{l,(t)}$ . The constants are considered  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\eta = 0.001$  and  $\epsilon = 10^{-8}$ . We repeat the same expressions for the bias, but applying them to the  $\partial MAE / \partial b_j^l$ .

Every time these 2 steps are carried out over all the neurons when the training data is passed to the emulator is called an "epoch". In general, the emulator needs to repeat the steps over a large number of epochs to be properly trained. The strategy for training the emulator in our case has been approximately the one from Elliott et al. (2021). Real datasets results of the function are used for training, but with a difference because they are divided into 3 types of data: training data (80% of the models), validation data (10%) and test data (10%). The training data has been used following the 2 steps mentioned before. The new changes come from the validation and test data. The validation data is very relevant to generalise the emulator to data different to the training one (unseen data), which is the most difficult part. We have generated a loop of epochs where at the end of each epoch we compute the average MAE of the validation datasets, saving only the weights and biases obtained from the training data when there has been improvement in the validation MAE. Plus until the validation MAE had not improved during a specific number of epochs (in this work 500 epochs has been found to be a correct balance between computational time and improvement of the emulator), the training did not finish. Once the emulator was already trained, we have run it over the remaining test dataset to obtain another average MAE value that has allowed us to compare the performance of various emulators in case the test data used is the same. After these considerations, we were capable of properly replacing it by the non-linear function.

### 3.2 Emulating galaxy model observables

In our study the particular function decoded by the emulator is the one that represents the SAM. For simplicity the SAM can be described as a function that takes the merger trees of a DM-only simulation and the value of the free parameters (related to the physical processes that play a role in galaxies: star formation, mergers, gas outflows and cooling or growth of black holes) and produces as output some observables (galaxy morphologies, galaxy colours, star formation, clustering, luminosity, ...) at certain bins or values. But as we have worked with the fixed UNIT subvolume, the only input are the adjustable parameters. Both the particular free parameters considered as input ( $\alpha_{cool}$ ,  $V_{SN}$  and  $F_{stab}$ ) and the observables considered as

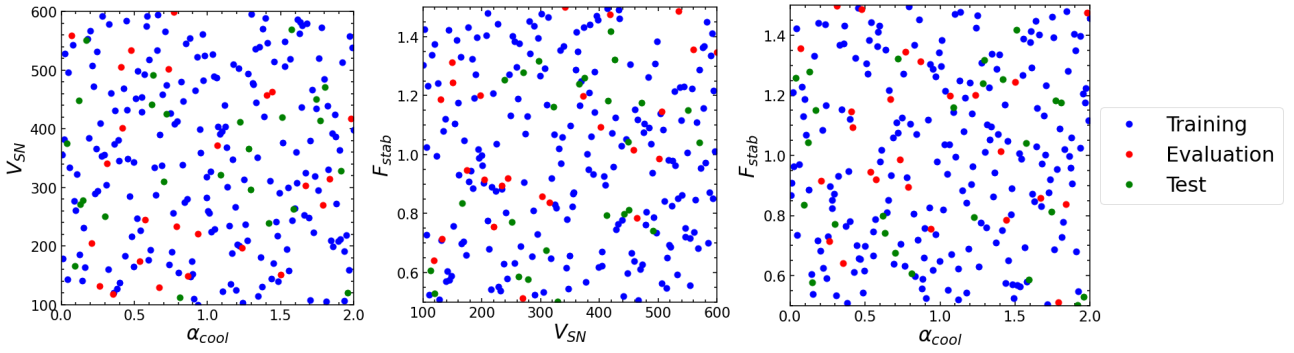


Figure 8: Latin Hypercube sampling of free parameters chosen ( $\alpha_{cool}$ ,  $V_{SN}$ ,  $F_{stab}$ ) for the 250 GALFORM models necessary for the emulator training. The blue dots correspond to the 200 used as training models, the red dots to the 25 validation models and the green dots to the 25 test models.

output observables (K LF,  $b_J$  LF, SFRF and the passive fraction) to study the parameter space has been explained before and they appeared in the Figure 7 diagram.

The emulator has only been trained over the most reliable observational data ranges of the observables when it comes to considering the resolution limits from the DM-only simulation and reasonable observational errorbars. For the LFs we have chosen Driver et al. (2012) considering the K LF limits and the r LF (equivalent to the  $b_J$  band) limits; for the SFRF we have restricted it to the values followed in Knebe et al. (2018) while in the latter case for the passive fraction the limits are taken from the observations of Henriques et al. (2015) of the galactic stellar mass function. In all the cases as the LFs and the SFRF are based on logarithms and the passive fraction on a number galaxy ratio, we have enclosed the boundaries even more if the GALFORM runs gave us 0 galaxies of a certain luminosity, mass or SFR. The final ranges are collected in Table 2 and they are also visible in Figure 5 (the shadowed regions are out of these ranges).

Input parameter	Range
$\alpha_{cool}$	(0, 2)
$V_{SN}$	(100, 600)
$F_{stab}$	(0.5, 1.5)
Output observables	Range
K LF	(-22.25, -16.25)
$b_J$ LF	(-21, -14)
SFRF	(-1.5, 1.5)
Passive fraction	(9.42, 11.02)

Table 2: Range of free parameters (inputs) and observational functions (outputs) explored by the emulators.

On the other hand, the ranges analysed for the free parameters are shown in Table 2; which have been chosen taking the ranges used for the above selected parameters in the respective calibrations of different articles such as Lacey et al. (2016) and Elliott et al. (2021), could be extending a little the boundaries as the emulator power enabled us to do it.

Moreover, taking into account the expected degeneracies in the parameter space previously mentioned in section § 2.5.1, we have created 3 different emulators all of them with the same input free parameters ( $\alpha_{cool}$ ,  $V_{SN}$  and  $F_{stab}$ ) from GALFORM depending on the observational data that we consider as output of the emulator. The aim of developing different emulators is to try to break the possible degeneracies among the parameters. The 3 emulators are summarized:

1. Emulator E1: the outputs are both the K and the  $b_J$  LFs.
2. Emulator E2: the outputs are the K and  $b_J$  LFs, plus the SFRF.
3. Emulator E3: the outputs are the K and  $b_J$  LFs, SFRF, plus the passive fraction.

For all the emulator's (E1,E2,E3) training we have run 250 GALFORM models using a Latin Hypercube sampling of the input parameters (Bower et al., 2010) to explore the parameter space as evenly as possible with the least number of galaxy models as it is shown in Figure 8. These 250 models are divided into: 200 as training models, 25 as validation models and 25 as test models; which are as well distinguished in Figure 8 and it is clear that each set of models approximately covers the whole parameter space volume.

### 3.2.1 Parameter space exploration

Once each of the emulators have been trained and tested, we are capable of properly replacing running GALFORM with them. Therefore, we are ready to carry out the extensive parameter space exploration. To do it, We produce efficient emulator predictions for 10000 parameter values sampling again the whole parameter space with a Latin Hypercube to have an adequate probability distribution and we find the set of free parameters that are most consistent with observations (best fit model). As we look for the most consistent results with observations, we calculate the  $\chi^2$  value with respect to inferred data from observations.

$$\chi_i^2 = \frac{1}{n} \sum_{k=1}^n \left( \frac{y_{n,obs} - F_n(\mathbf{x}_0)}{e_{n,obs}} \right)^2 \quad (19)$$

where  $y_{n,obs}$  and  $e_{n,obs}$  are each observational data point and its associated error, and  $F_n(\mathbf{x}_0)$  the predictions from the emulator (usually linearly interpolated to have the values at the exact points of the observations). The expression is normalized by n, the number of observational data points produce by the emulator. The expression (19) is applied for each of the observational functions that are reproduced by the emulator (2 observational functions for E1, 3 for E2 and 4 for E3), and at the end the probability distribution is computed based on the average  $\chi^2$  over all the observational functions:  $\chi^2 = (1/k) \sum_i^k \chi_i^2$ . The  $y_{n,obs}$  and  $e_{n,obs}$  observational values are taken from (Driver et al., 2012) for the K LF, from (Norberg et al., 2002) for the  $b_J$  LF, from Gruppioni et al. (2015) for the SFRF and from Bauer et al. (2013) together with Gilbank et al. (2010) for the passive fraction. All the observations appear in Figure 5 as well.

## 4 Results and discussion

### 4.1 Emulator training using luminosity functions

Both the  $b_J$  and K luminosity functions are output used to train the emulator. We work in particular with this 2 visible and nIR bands since the strongest observations are carried out in them. In this emulator we hoped to find huge degeneracies. The emulator has been trained with the 250 models and in Figure 9 we have found that the MAE of both the training and validation sets decreases approaching to the minimum value of 0. The meaning of this is that the emulator

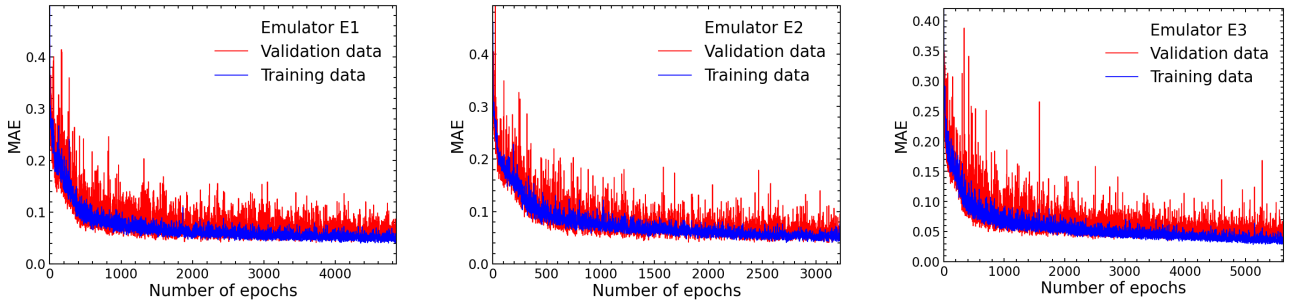


Figure 9: MAE of the training data (blue curve) and the validation data (red curve) as a function of the number of epochs (in each epoch the weights and biases are recalculated and updated as long as the performance over the validation data has improved) for the 3 emulators developed. The training stops when the MAE of the validation data has not improved for 500 consecutive epochs.

performance is improving. Moreover, we have no signs of overfitting or underfitting since both datasets decreases during training or at the end of the final epochs the values are kept almost constant. However, we can realise the validation set results are much more noisy due to the fact these datasets are unseen data for the emulator; but despite the huge dispersion, the mean values of both training and validation are not so far away.

After the training and evaluation, the emulator still needs to be tested. The results from the test part are in Figure 10 where the blue curves in both LFs show the ratio between emulator predictions and GALFORM model results with a median value close to 1, but not stable around it at all. The dispersion for its part is approximately of 5%. Besides, the  $b_J$  LF seems to be better generated by the emulator than the K-band luminosity.

## 4.2 Emulator training using luminosity functions and star formation rate function

This second emulator had as main objective to break some of the degeneracies found in the first emulator, since we were introducing a new observable as the SFRF. Again, the emulator training and validation results appear in Figure 9 and they are quite similar to the ones in the previous case even with an equivalent minimum MAE value reached. Perhaps the main difference is that it has needed around 1000 less epochs to finish the training, but it is not significant at all. The testing results of Figure 10 (orange curves) have a median value closer to 1 than for the E1 emulator, and the dispersion seems to be slightly low (roughly 3%) including the percentiles results from the SFRF. For these reasons, this new emulator seemed to produce best results.

## 4.3 Emulator training using luminosity functions, star formation rate function and passive fraction

If the second emulator was planned to break some of the degeneracies, the final emulator was going to be the one that produced even better results considering that we were including the passive fraction directly in the calibration. First, the emulator curves of training and validation data as a function of the epochs in Figure 9 are quite similar to the previous cases with the maximum number of epochs needed for the training, with a total of more than 5000.

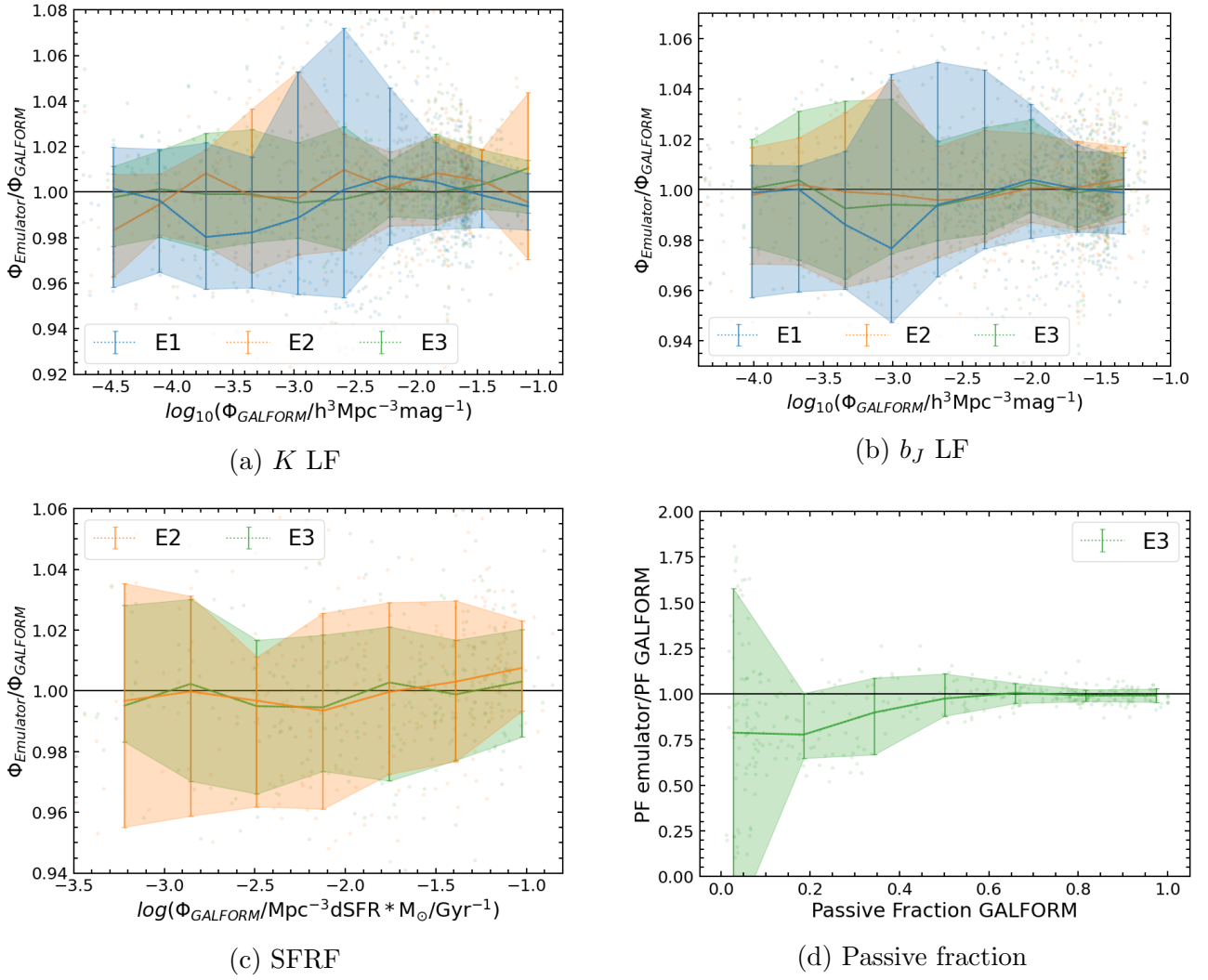


Figure 10: Percentiles of the ratio between the GALFORM model results of the observational functions and the emulator predictions from the 25 testing models. The solid central lines represent the median values, while the upper and lower limit of the errorbars are the 84% and 16% respectively. There are results from the 4 observables: a) the K LF and b) the  $b_J$  LF, both reproduced by the 3 emulators (E1,E2,E3); c) the SFRF reproduced by 2 emulators (E2,E3) and d) the passive fraction only reproduced by the final E3 emulator. In the background in a more transparent colour, the points that generate these percentiles appear.

The Figure 10 shows the predictions for the test models. In this emulator we have to focus on the green curves where visually the median value of the ratio is quite close to 1 with an stability around it (not oscillating too much). On the other hand, the dispersion of the ratio is the smallest with around 2% and is also roughly symmetric. Both characteristics are valid for the LFs, the SFRF and the extreme nearer to 1 in the passive fraction. The exception are the values close to 0 in the passive fraction where the median and the dispersion even diverges. This is due to the fact that the emulator was not properly trained and produced various negative values of the passive fraction with no physical meaning (also values greater than 1 with no meaning either). This can be seen in Figure 11, where we have the test predictions and models not in the form of ratio percentiles, but as the 4 observational functions calibrated. Here once more it is visible that the function that is imitated in a less reasonable way is the passive fraction, since even in the SFRF, the prediction for the peak found at more or less  $\text{SFR}=0$  is adequate. In the background in this Figure we can see the models that have been used as training, and it is noticeable that the test covers the distinct behaviours, which means they

are a proper set of data for testing. A possible solution for the future is to implement a new emulator whose output values satisfy the condition to be in the  $[0,1]$  range.

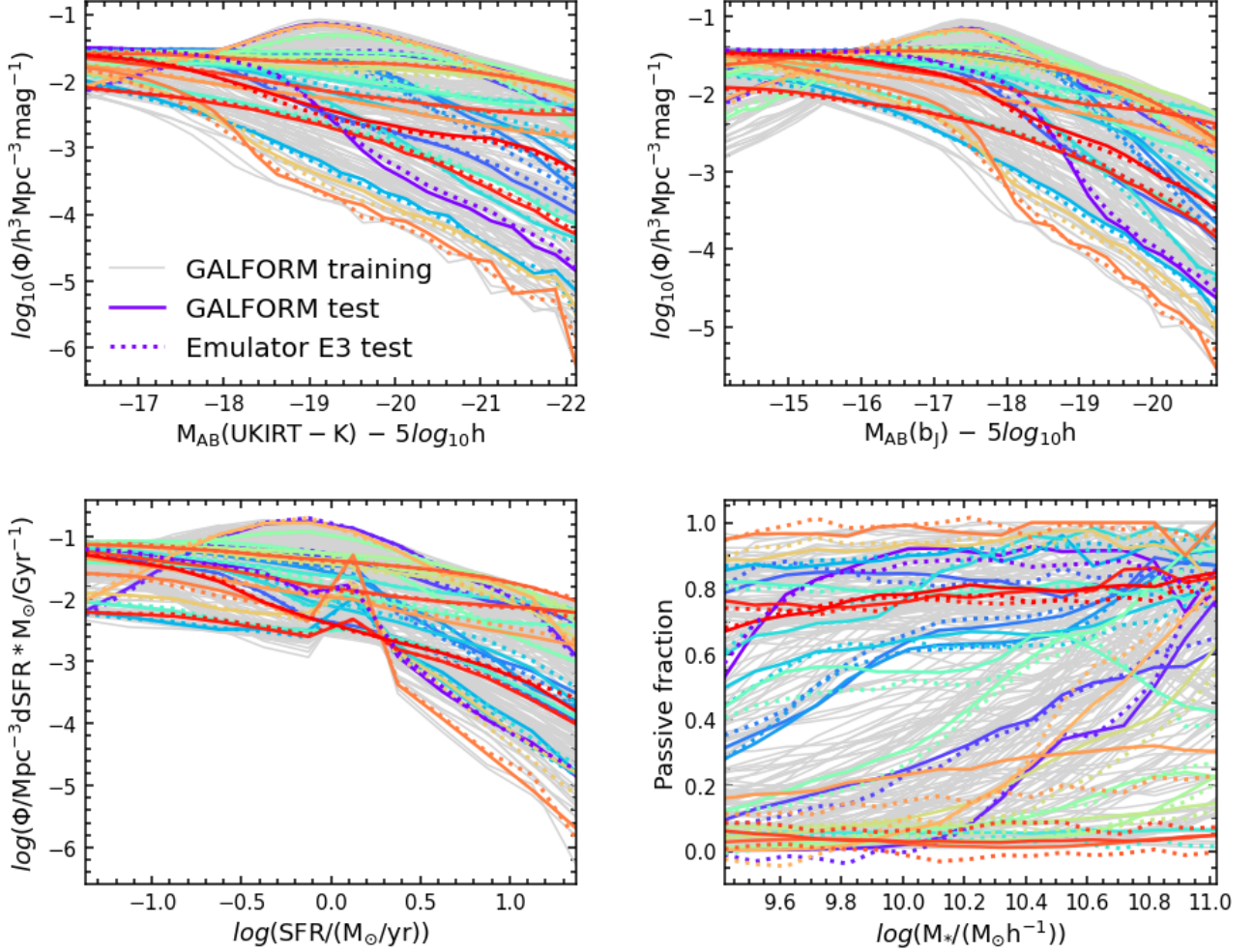


Figure 11: The 4 observables (K LF in the left top panel,  $b_J$  LF in the right top panel, SFRF in the left bottom panel and PF in the right bottom panel) for the E3 emulator where the GALFORM model results are visually compared to the emulator predictions for the E3 emulator. Plus in the background the GALFORM models used for training the emulator are shown to have a perspective of the range covered by the parameter space analysed. It is a qualitative comparison of the Figure 10 with percentiles.

#### 4.4 Comparison of emulators performance

The Figure 10 already analysed is a comparison method, although qualitative, among the emulators. The best results seem to come from the emulator E3. If the dispersion or a measurement of the mean value of the percentiles had been calculated, quantitative results would have been obtained. However, the most reliable quantitative way to compare the performance of the emulators is to consider the average MAE over the test datasets for each of them. The results appear in Table 3 and the values are all of them of the same magnitude order with no significant differences. The value obtained is lower as we introduce more observables for calibration: the performance of the second one is better than the first emulator, while the performance of the third one is even better; what is obtained also in Figure 10. Therefore, again the best performance is produced by the emulator E3, despite not reproducing properly the passive fraction;



but improving the results of the other emulators in the LFs and the SFRF. On average the different emulators are able to reproduce the opposed types of behaviour, albeit at the edges of the parameter bounds the training models are sparsely sampled and it is common that the emulator performance is less reliable there.

Emulator	Test dataset MAE	$\alpha_{cool}(\chi^2_{min})$	$V_{SN}(\chi^2_{min})$	$F_{stab}(\chi^2_{min})$
E1 - LFs	0.049	0.65	401	0.68
E2 - LFs+SFRF	0.043	0.57	401	0.89
E3 - LFs+SFRF+PF	0.036	0.56	386	0.88

Table 3: Comparison of emulator performances taking into account the average MAE value over the 25 GALFORM test models, and comparison of the best fit (where the minimum  $\chi^2$  is obtained) parameters for each emulator.

From Figures 9 and 10 we can see that it is possible to get reasonable predictions of the LFs and the SFRF making use of the emulators. They are not perfectly accurate predictions (the median values in Figure 10 are not exactly equal to 1 with zero dispersion), but they allow us to reproduce the GALFORM results to have an overall view of the complex parameter space involved in the model. However, the passive fraction behaviour is less simple to mimic. The less reliability of the passive fraction results from both Figures could be due to the fact that the range of parameters employed in Table 2 for the exploration have a wide variety of behaviours as it is particularly visible on the right bottom panel of Figure 10. The whole range in mass and in y-axis between 0 and 1 is covered by the 200 grey curves, unlike what happens with the LFs and the SFRF, where the functions have a pretty standard tendency with certain variations introduced when exploring the regions of the parameter space. Other possibility is that the complex nature of selecting passive galaxies with implicit dependences on the SFR (since it is the ratio of quenched galaxies over the total, condition that is imposed on the sSFR) or on the mass are impossible to be decoded by our emulator.

## 4.5 Fitting the observations with the emulators

Once the emulator performances have been analysed, it is the turn of going into detail in the 3-dimensional parameter space of the 3 cases above. Results from the later systematic exploration of the parameter space using 10000 points distributed along a Latin Hypercube are only shown for the E3 emulator since the results are quite similar in all of them. That means the E1 emulator shows the expected degeneracies, but it also means that the expected degeneracy break either introducing the SFRF or the PF have not been successful. Despite the results from training where emulators performed better when more observational constraints were included, which seemed to indicate that some change in the degeneracies may have happened; either E2 emulator including the SFRF or E3 emulator including both the SFRF and the passive galaxies are still constrained in the same way. The Figure 12 gives us the  $\chi^2$  distribution (probability distribution) for the emulator E3, as the third emulator is supposed to be the one with the best performance.

On the one hand, the  $(\alpha_{cool}, V_{SN})$  plane shows a elongated structure in the region more consistent with observations (the dark blue colour marks the lower  $\chi^2$  values as seen in Figure 12 colorbar), which is an indicative of the strong degeneracy between both parameters found in Bower et al. (2010) and it is visible in the 2D distribution. Focusing on the  $(V_{SN}, F_{stab})$  surface, albeit it is clear that we look for parameter values larger than  $V_{SN} = 300$  consistent with several articles as

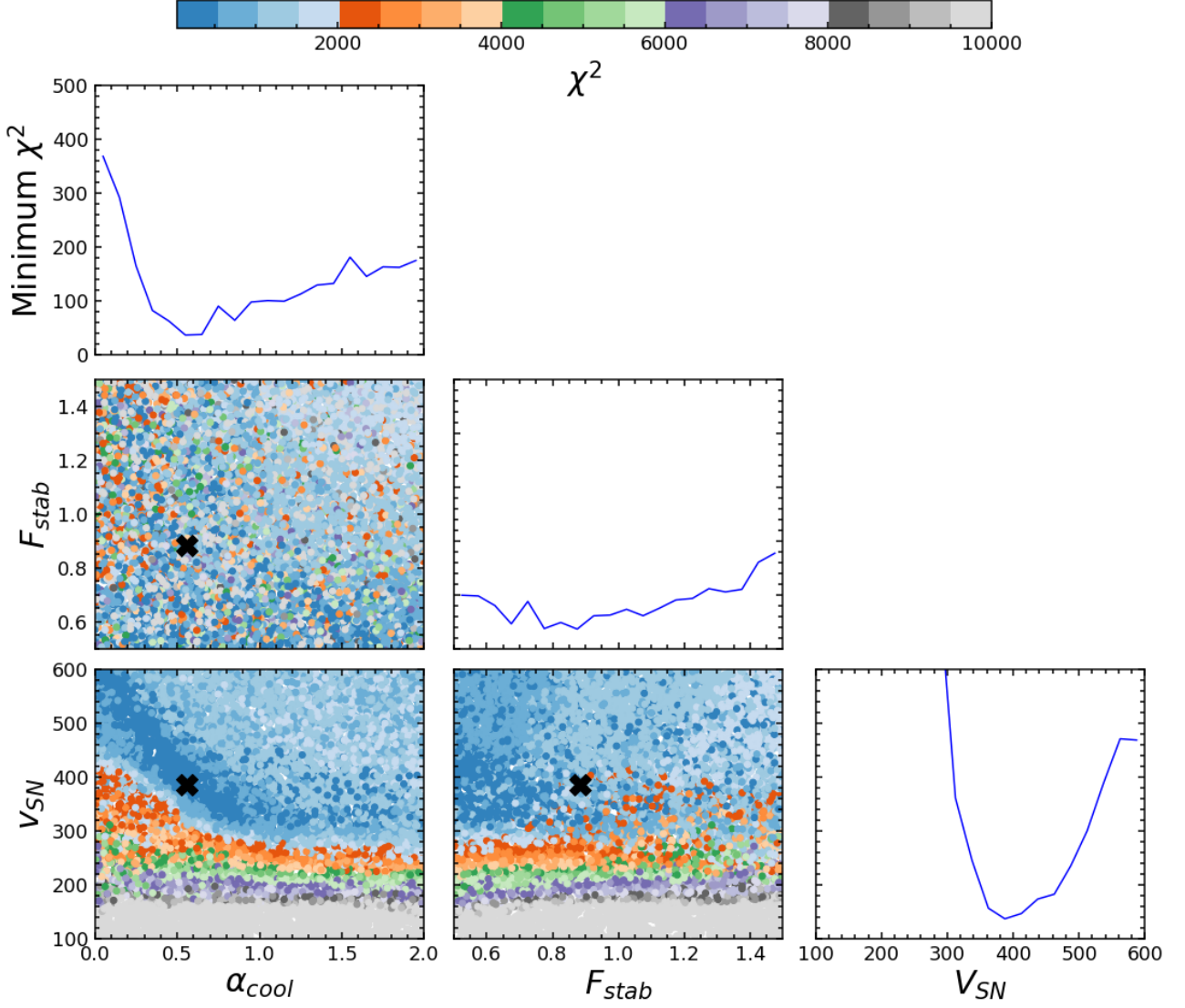


Figure 12:  $\chi^2$  probability distribution of the 3D parameter space in the case of the emulator E3. The 2D panels show the possible parameter pairs and are collapsed in the remaining dimensions. The colorbar indicates the  $\chi^2$  with the dark blue as the lower values, and the black cross locates the position in the parameter space of the minimum  $\chi^2$ . The values of the parameters that produce the best fit to observations (minimum  $\chi^2$ ) are collected in Table 3. The 3 remaining curves represent the minimum  $\chi^2$  in different bins of the particular parameter values collapsed in the remaining 2 directions. The ranges on the vertical axis of the 1D collapsed graphics are the same, as well the horizontal axis in these graphics which correspond to the ones used for the parameters in Table 2.

Bower et al. (2010) or others (in the previous  $(\alpha_{cool}, V_{SN})$  plane this conclusion could have also been drawn). It is very straightforward to notice the non-validity when  $V_{SN} < 300$  so that the faint end of the  $K$  and  $b_J$  LFs at  $z=0$  agree with observations. In the area of more consistent  $V_{SN}$  parameters, the lower  $\chi^2$  values are not concentrated on a specific region, but the dark blue colour is more scattered. Even in the  $(\alpha_{cool}, F_{stab})$  plane (the other known degeneracy due to the fact only AGNs are present in bulges in our model) this effect is more noticeable: the whole 2D area does not have a defined distribution, where the same  $\chi^2$  values do not have any dominated preference. This is another effect produce by a constraint between both parameters. A way to improve the clarity of the exploration would be to include contour lines on the probability distribution of Figure 12.



Looking at the 1D projections of the minimum  $\chi^2$  in Figure 12 as well it is visible that the effect of  $F_{stab}$  is less relevant because the minimum is more or less constant along the range (0.5-1.5) of the parameter, unlike what happens with the other 2 projections where there is a clear minimum in each case. This is another characteristic of degeneracies. Therefore, the best fit  $F_{stab}$  parameter here obtained is expected to vary for a different training set, as the results are particularly noisy. In addition, we have identified as it is said in Elliott et al. (2021) that  $F_{stab}$  is not as constrained to a tight range as may happen with other parameters. As  $F_{stab}$  is a limit in the expression (8) it seems logical that it is more difficult to constrain it, and to do that we should have chosen observables related to bulges and disks and not the ones used.

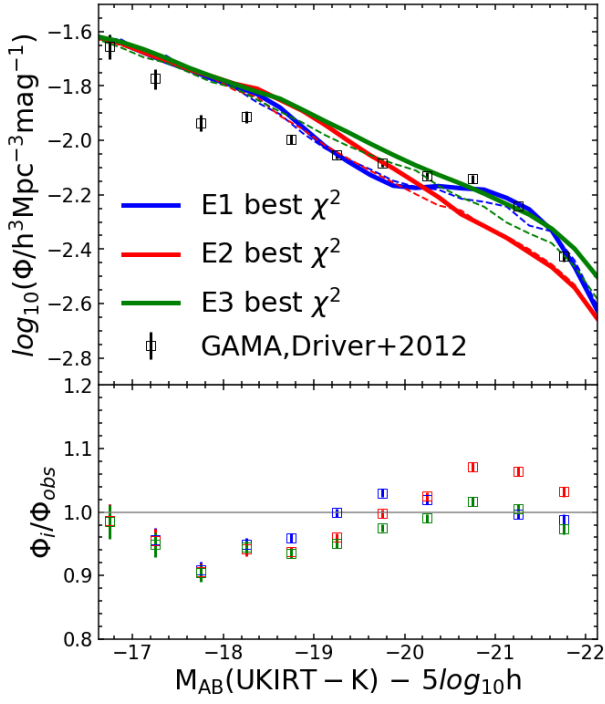
On the other hand, the fact that we obtain similar results training over different observables (3 emulators) means that the additional functions, in this case the SFRF and the passive fraction, do not include any new constraint. The explanation to this could be that the luminosity bands used are implicitly related to the star formation activity, specially in the case of the  $b_J$  luminosity since it traces young objects where star formation plays a major role. Plus the passive fraction has been explained that is computed using the sSFR, which in turn is proportional to star formation. In the end, the 2 new added observables to improve the emulator have not presented new physics. Another option might be that the galaxy model treatment of some physical processes related to star formation or another process is not described properly and it would need to be revised, what eventually would produce the absence of constraints as well. In Elliott et al. (2021) for example it is said that the SN feedback parameters largely constrain not only the LFs, but the rest of the observational functions.

## 4.6 The best fit models

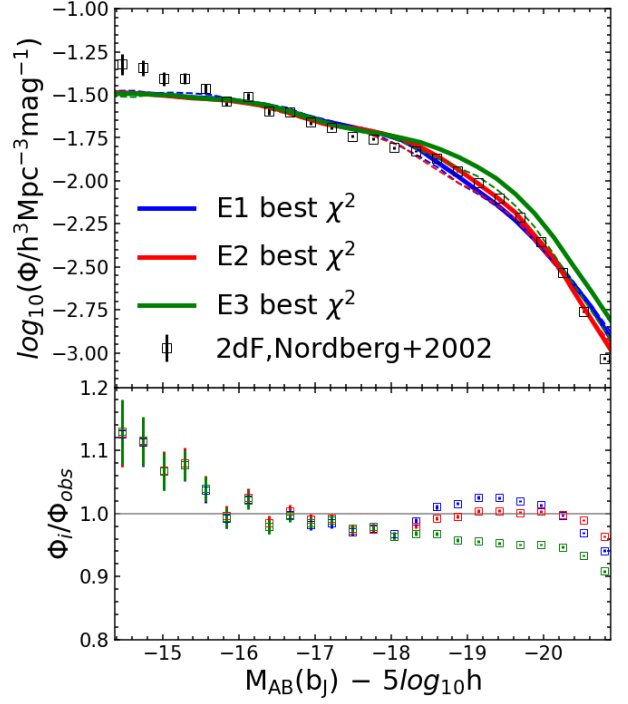
We also are interested in the minimum  $\chi^2$  value reached in each emulator, marked by a black cross in the parameter space of Figure 12. Even though, the Figure only shows the E3 emulator results (which is supposed to be the best performed emulator) the particular values obtained for the free parameters which produce the best fit to the inferred observational data are quite similar in all the emulators as the probability distribution was practically the same visually in all cases. The values are close to  $\alpha_{cool} \approx 0.6$ ,  $V_{SN} \approx 400$  and  $F_{stab} \approx 0.8$  and they appeared explicitly in Table 3.

The reference values of these parameters from the flavour used as basis (VGP20) are  $\alpha_{cool} = 0.8$ ,  $V_{SN} = 370$  and  $F_{stab} = 0.8$  obtained calibrating the same K and  $b_J$  LFs and the SMBH mass-bulge stellar mass relation at  $z=0$ . The results are quite closer particularly for the SN feedback and disk instabilities, while for the AGN feedback it has varied because we have not used observations directly connected to AGNs as the the SMBH mass-bulge stellar mass. Moreover, we have reached also best fit parameter values close to the results in Elliott et al. (2021), although in that case the number of input parameter and output datasets used to train the emulator was much larger. Our best fit parameters are also consistent with the ones in Lacey et al. (2016) and Bower et al. (2010).

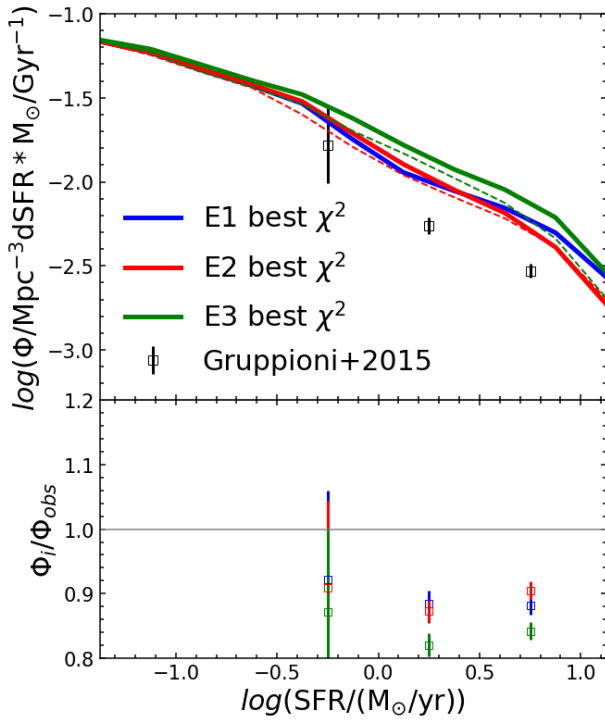
Finally, knowing the best fit parameters to observations obtained from the emulator, we have run GALFORM itself with the 3 best fit sets for each emulator. After this, we reproduce the 4 observational functions (K LF,  $b_J$  LF, SFRF and PF) in such a way that we have been able to compare the results from GALFORM and from the emulator of the different observables, as well as a comparison with the inferred data coming from observations in Figure 13. In general the results are reasonable compared to observations, although they are not extremely consistent.



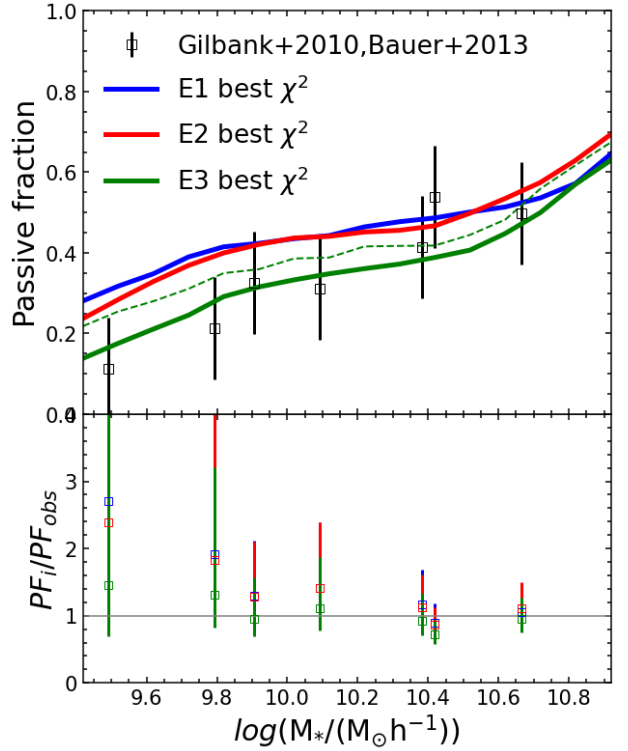
(a)  $K$  LF



(b)  $b_J$  LF



(c) SFRF



(d) Passive fraction

Figure 13:  $K$  LF (a),  $b_J$  LF (b), SFRF (c) and PF (d) for the best fit parameters (minimum  $\chi^2$  found in the parameter space): the prediction from the different emulators and the corresponding result running the GALFORM model with these parameters. Observations are also included for comparison to the best fit, , with the ratio between them and the GALFORM best fit results taking into account into account the errorbars.

Figure 13 shows that to first order the K LF is reproduced with only one slope, but instead visually we can observe that the observational results are more complex with at least 2 slope changes. Maybe the E1 emulator gives a result with a more complicated behaviour, but in the end it is not the same as the observational data. These slope changes are not mimicked correctly neither by the best fit emulator predictions (dashed lines) nor by GALFORM. The  $b_J$  results seem to be more consistent with observations, except at the tail of the faintest galaxies and all the emulator predictions and GALFORM results for the best fit coincide more or less with the exception of the GALFORM model for the E3 emulator, which distances itself from the rest of curves, regardless of being the best performed (since it distances with respect to the dashed curve coming from the emulator predictions).

In the SFRF only the E2 and E3 emulators have been calibrated using this specific observational function. That is why we only have red and green dashed lines. However, the results of E1 compared to the ones from the other 2 emulators are quite similar, since the best fit parameters found were closer in all cases. Again the E3 GALFORM model curve separates with respect to the emulator prediction (the emulator underpredicts the result both in the SFRF as in the LFs in spite of having an almost symmetric dispersion in Figure 10 percentiles). The last function to examine is the passive fraction whose results can be compared with emulator predictions only in the case of the E3 emulator (the only one calibrated with this function). At least this E3 emulator is the most consistent with observations, meaning the training with this function has worked, whereas the other GALFORM results are more distant due to the difficulty mentioned above when representing the fraction of quenched galaxies. The errorbars in this passive fraction panel are different from the ones in Figure 5, because as the  $\chi^2$  expression (19) involved the observational error and here we had 2 sets of observations, we have taken an equal error for all the points as the average distance between both sets of observations.

## 5 Summary and conclusions

Galaxies are complex systems in which physical processes spanning many orders of magnitude interplay in non-linear ways. Moreover, in our best cosmological model, galaxies are formed in gravitational potential wells generated by dark matter. Due to this complexity, simulations are needed to interpret observational data and to extract information from it at different scales. In this work we use a dark matter (DM) only simulation (Chuang et al., 2019), together with a semi-analytical model of galaxy formation, GALFORM (Gonzalez-Perez et al., 2020), to explore some of the limitations of such modelling approach. To run GALFORM we needed to also install locally and run the Jiang et al. (2014) code, after matching properties between this code and the halo finder (Behroozi et al., 2013a) and merger tree code (Behroozi et al., 2013b) that had already been run on the UNIT simulation.

Semi-analytic models of galaxy formation encapsulate those processes relevant to galaxy evolution into analytical equations with a set of free parameters. The non-linearity among the parameters give rise to degeneracies and to a multi-dimensional complex parameter space. In this work, we use deep learning techniques to carry out a systematic exploration of the probability distribution of 3 of these free parameters to understand them better. In particular, we use an emulator with a set-up following that in Elliott et al. (2021) and we focus on parameters that control the stellar and nuclear feedback:  $\alpha_{cool}, V_{SN}$  and  $F_{stab}$  described in § 2.5.1. To calibrate the model with emulating tools that allow for the parameter exploration and the understanding of degeneracies, we have developed 3 emulators. The input of the emulators is the set of free parameters chosen  $(\alpha_{cool}, V_{SN}, F_{stab})$  whereas the output are several observables:

some luminosity functions necessary to have the correct galaxy statistics, as well as the star formation rate function and the passive fraction. We have trained the emulators with 250 SAM models spanning the ranges shown in Table 2 for each parameter and sampled in a Latin Hypercube. The E3 emulator is the one that produce the best performance results, although the other 2 emulator performances are quite similar.

Once we have the 3 trained emulators, we find the set of parameters that best match the corresponding observational data, by measuring the  $\chi^2$  of 10000 emulators run with parameters spread within the ranges tabulated in Table 2 and determined using a Latin Hypercube with respect to observational data. The results we have reached are similar in the 3 emulators, with an also equivalent systematic parameter space exploration. The  $\chi^2$  of E1 compared to observations shows a clear degeneracy between  $\alpha_{cool}$  and  $V_{SN}$ , as well as between  $\alpha_{cool}$  and  $F_{stab}$ . Therefore, we have introduced more observables to try to break these degeneracies, developing the E2 emulator involving the SFRF. However, the SFRF introduction does not improve the results. In a similar way, the passive fraction considered in the E3 emulator does not vary the outcome.

The main conclusion to stand out is the fact that the parameter degeneracies have not been broken for the  $(\alpha_{cool}, V_{SN})$  and  $(\alpha_{cool}, F_{stab})$  pairs and the strong constraints are kept regardless of the introduction of new observables in the SAM calibration, as the SFRF and the passive fraction. The reason for that could be due to the fact that the  $b_J$  luminosity function already traces implicitly the star formation. Thus, neither the SFRF (rate of star formation) nor the passive fraction (quenched galaxies obtained imposing a condition to the specific star formation rate) seem to introduce independent physics since although they model different physical processes, they are primary related to the stellar mass and star formation as well. The other possible reasoning is that some physical processes are being missed in the galaxy model or that modifications on some of them need to be developed, though it seems an explanation based on weaker evidences. Taking all of this into consideration, the main conclusion from these results is that both the SFRF and the PF do not add a constraint.

An interesting extension for this work will be to fix the best fit parameters  $(\alpha_{cool}, V_{SN}, F_{stab})$  of the emulators in GALFORM and then to examine the gas stripping in satellite galaxies performing a second consecutive deep learning calibration varying only a new parameter  $\epsilon_{strip}$ , which takes into account the fraction of hot gas that is stripped when a galaxy becomes a satellite based on the model developed by Font et al. (2008). To do it, the idea would be to follow the same methodology of this work just calibrating the passive fraction. Another possibility would be to study in more detail the degeneracies found in this project using promising independent observables for the calibration (e.g. related to galaxy sizes) that allow for breaking finally the constraints. It is also feasible to extend the emulating techniques to other physical problems or to other SAMs as SHARK (Lagos et al., 2018).

## 6 Bibliography

- M. Abadi. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. <https://www.tensorflow.org/>.
- A. E. Bauer, A. M. Hopkins, M. Gunawardhana, E. N. Taylor, I. Baldry, S. P. Bamford, J. Bland-Hawthorn, S. Brough, M. J. I. Brown, M. E. Cluver, M. Colless, C. J. Conselice, S. Croom, S. Driver, C. Foster, D. H. Jones, M. A. Lara-Lopez, J. Liske, Á. R. López-Sánchez, J. Loveday, P. Norberg, M. S. Owers, K. Pimbblet, A. Robotham, A. E. Sansom, and R. Sharp. Galaxy And Mass Assembly (GAMA): linking star formation histories and stellar mass growth. , 434(1):209–221, Sept. 2013. doi: 10.1093/mnras/stt1011.
- C. M. Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports on Progress in Physics*, 69(12):3101–3156, Dec. 2006. doi: 10.1088/0034-4885/69/12/R02.
- P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. , 762(2):109, Jan. 2013a. doi: 10.1088/0004-637X/762/2/109.
- P. S. Behroozi, R. H. Wechsler, H.-Y. Wu, M. T. Busha, A. A. Klypin, and J. R. Primack. Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. , 763(1):18, Jan. 2013b. doi: 10.1088/0004-637X/763/1/18.
- R. G. Bower, I. Vernon, M. Goldstein, A. J. Benson, C. G. Lacey, C. M. Baugh, S. Cole, and C. S. Frenk. The parameter space of galaxy formation. , 407(4):2017–2045, Oct. 2010. doi: 10.1111/j.1365-2966.2010.16991.x.
- T. J. Broadhurst, R. S. Ellis, and K. Glazebrook. Faint galaxies: evolution and cosmological curvature. , 355(6355):55–58, Jan. 1992. doi: 10.1038/355055a0.
- G. L. Bryan and M. L. Norman. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. , 495(1):80–99, Mar. 1998. doi: 10.1086/305262.
- C.-H. Chuang, G. Yepes, F.-S. Kitauro, M. Pellejero-Ibanez, S. Rodríguez-Torres, Y. Feng, R. B. Metcalf, R. H. Wechsler, C. Zhao, C.-H. To, S. Alam, A. Banerjee, J. DeRose, C. Giocoli, A. Knebe, and G. Reyes. UNIT project: Universe N-body simulations for the Investigation of Theoretical models from galaxy surveys. , 487(1):48–59, July 2019. doi: 10.1093/mnras/stz1233.
- S. P. Driver, A. S. G. Robotham, L. Kelvin, M. Alpaslan, I. K. Baldry, S. P. Bamford, S. Brough, M. Brown, A. M. Hopkins, J. Liske, J. Loveday, P. Norberg, J. A. Peacock, E. Andrae, J. Bland-Hawthorn, N. Bourne, E. Cameron, M. Colless, C. J. Conselice, S. M. Croom, L. Dunne, C. S. Frenk, A. W. Graham, M. Gunawardhana, D. T. Hill, D. H. Jones, K. Kuijken, B. Madore, R. C. Nichol, H. R. Parkinson, K. A. Pimbblet, S. Phillipps, C. C. Popescu, M. Prescott, M. Seibert, R. G. Sharp, W. J. Sutherland, E. N. Taylor, D. Thomas, R. J. Tuffs, E. van Kampen, D. Wijesinghe, and S. Wilkins. Galaxy And Mass Assembly (GAMA): the 0.013  $\leq z \leq 0.1$  cosmic spectral energy distribution from 0.1  $\mu\text{m}$  to 1 mm. , 427(4):3244–3264, Dec. 2012. doi: 10.1111/j.1365-2966.2012.22036.x.
- E. J. Elliott, C. M. Baugh, and C. G. Lacey. Efficient exploration and calibration of a semi-analytical model of galaxy formation with deep learning. , 506(3):4011–4030, Sept. 2021. doi: 10.1093/mnras/stab1837.

- A. S. Font, R. G. Bower, I. G. McCarthy, A. J. Benson, C. S. Frenk, J. C. Helly, C. G. Lacey, C. M. Baugh, and S. Cole. The colours of satellite galaxies in groups and clusters. , 389(4): 1619–1629, Oct. 2008. doi: 10.1111/j.1365-2966.2008.13698.x.
- M. Franx, P. G. van Dokkum, N. M. Förster Schreiber, S. Wuyts, I. Labbé, and S. Toft. Structure and Star Formation in Galaxies out to  $z = 3$ : Evidence for Surface Density Dependent Evolution and Upsizing. , 688(2):770–788, Dec. 2008. doi: 10.1086/592431.
- D. G. Gilbank, I. K. Baldry, M. L. Balogh, K. Glazebrook, and R. G. Bower. The local star formation rate density: assessing calibrations using [OII], H and UV luminosities. , 405(4): 2594–2614, July 2010. doi: 10.1111/j.1365-2966.2010.16640.x.
- V. Gonzalez-Perez, W. Cui, S. Contreras, C. M. Baugh, J. Comparat, A. J. Griffin, J. Helly, A. Knebe, C. Lacey, and P. Norberg. Do model emission line galaxies live in filaments at  $z \sim 1$ ? , 498(2):1852–1870, Oct. 2020. doi: 10.1093/mnras/staa2504.
- C. Gruppioni, F. Calura, F. Pozzi, I. Delvecchio, S. Berta, G. De Lucia, F. Fontanot, A. Franceschini, L. Marchetti, N. Menci, P. Monaco, and M. Vaccari. Star formation in Herschel’s Monsters versus semi-analytic models. , 451(4):3419–3426, Aug. 2015. doi: 10.1093/mnras/stv1204.
- B. M. B. Henriques, S. D. M. White, P. A. Thomas, R. Angulo, Q. Guo, G. Lemson, V. Springel, and R. Overzier. Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses. , 451(3):2663–2680, Aug. 2015. doi: 10.1093/mnras/stv705.
- L. Jiang, J. C. Helly, S. Cole, and C. S. Frenk. N-body dark matter haloes with simple hierarchical histories. , 440(3):2115–2135, May 2014. doi: 10.1093/mnras/stu390.
- D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, Dec. 2014.
- A. A. Klypin, S. Trujillo-Gomez, and J. Primack. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. , 740(2):102, Oct. 2011. doi: 10.1088/0004-637X/740/2/102.
- A. Knebe, F. R. Pearce, P. A. Thomas, A. Benson, J. Blaizot, R. Bower, J. Carretero, F. J. Castander, A. Cattaneo, S. A. Cora, D. J. Croton, W. Cui, D. Cunnamea, G. De Lucia, J. E. Devriendt, P. J. Elahi, A. Font, F. Fontanot, J. Garcia-Bellido, I. D. Gargiulo, V. Gonzalez-Perez, J. Helly, B. Henriques, M. Hirschmann, J. Lee, G. A. Mamon, P. Monaco, J. Onions, N. D. Padilla, C. Power, A. Pujol, R. A. Skibba, R. S. Somerville, C. Srisawat, C. A. Vega-Martínez, and S. K. Yi. nIFTy cosmology: comparison of galaxy formation models. , 451(4): 4029–4059, Aug. 2015. doi: 10.1093/mnras/stv1149.
- A. Knebe, F. R. Pearce, V. Gonzalez-Perez, P. A. Thomas, A. Benson, R. Asquith, J. Blaizot, R. Bower, J. Carretero, F. J. Castander, A. Cattaneo, S. A. Cora, D. J. Croton, W. Cui, D. Cunnamea, J. E. Devriendt, P. J. Elahi, A. Font, F. Fontanot, I. D. Gargiulo, J. Helly, B. Henriques, J. Lee, G. A. Mamon, J. Onions, N. D. Padilla, C. Power, A. Pujol, A. N. Ruiz, C. Srisawat, A. R. H. Stevens, E. Tollet, C. A. Vega-Martínez, and S. K. Yi. Cosmic CARNage I: on the calibration of galaxy formation models. , 475(3):2936–2954, Apr. 2018. doi: 10.1093/mnras/stx3274.

- E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolte, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. , 192(2):18, Feb. 2011. doi: 10.1088/0067-0049/192/2/18.
- C. G. Lacey, C. M. Baugh, C. S. Frenk, A. J. Benson, R. G. Bower, S. Cole, V. Gonzalez-Perez, J. C. Helly, C. D. P. Lagos, and P. D. Mitchell. A unified multiwavelength model of galaxy formation. , 462:3854–3911, Nov. 2016. doi: 10.1093/mnras/stw1888.
- C. d. P. Lagos, R. J. Tobar, A. S. G. Robotham, D. Obreschkow, P. D. Mitchell, C. Power, and P. J. Elahi. Shark: introducing an open source, free, and flexible semi-analytic model of galaxy formation. , 481(3):3573–3603, Dec. 2018. doi: 10.1093/mnras/sty2440.
- J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. , 490(2):493–508, Dec. 1997. doi: 10.1086/304888.
- M. A. Nielsen. Neural Networks and Deep Learning. *Determination Press*, 2015.
- P. Norberg, S. Cole, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, R. De Propris, S. P. Driver, G. Efsthathiou, R. S. Ellis, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, J. A. Peacock, B. A. Peterson, W. Sutherland, K. Taylor, and 2DFGRS Team. The 2dF Galaxy Redshift Survey: the  $b_J$ -band galaxy luminosity function and survey selection function. , 336(3):907–931, Nov. 2002. doi: 10.1046/j.1365-8711.2002.05831.x.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Bacigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, H. C. Chiang, J. Chluba, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efsthathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, M. Farhang, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, T. Ghosh, M. Giard, Y. Giraud-Héraud, E. Giusarma, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, Z. Huang, K. M. Hufenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, J. P. Leahy, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marchini, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi,

J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdureau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, B. Rouillé d'Orfeuil, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, N. Said, V. Salvatelli, L. Salvati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, L. D. Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, M. Türlér, G. Umata, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, Sept. 2016. doi: 10.1051/0004-6361/201525830.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Kesitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, Sept. 2020. doi: 10.1051/0004-6361/201833910.

R. S. Somerville and R. Davé. Physical Models of Galaxy Formation in a Cosmological Framework. , 53:51–113, Aug. 2015. doi: 10.1146/annurev-astro-082812-140951.



R. H. Wechsler and J. L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. , 56:435–487, Sept. 2018. doi: 10.1146/annurev-astro-081817-051756.