Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **475,** 2936–2954 (2018)
Advance Access publication 2017 December 20

doi:10.1093/mnras/stx3274

Cosmic CARNage I: on the calibration of galaxy formation models

Alexander Knebe,[1,2★] Frazer R. Pearce,[3] Violeta Gonzalez-Perez,[4,5] Peter A. Thomas,[6] Andrew Benson,[7] Rachel Asquith,[3] Jeremy Blaizot,[8] Richard Bower,[5] Jorge Carretero,[9,10,11] Francisco J. Castander,[9,10] Andrea Cattaneo,[12,13] Sofía A. Cora,[14,15] Darren J. Croton,[16] Weiguang Cui,[1] Daniel Cunnama,[17,18] Julien E. Devriendt,[19] Pascal J. Elahi,[20] Andreea Font,[21] Fabio Fontanot,[22] Ignacio D. Gargiulo,[14,15] John Helly,[5] Bruno Henriques,[23,24] Jaehyun Lee,[25] Gary A. Mamon,[26] Julian Onions,[3] Nelson D. Padilla,[27,28] Chris Power,[20] Arnau Pujol,[29,30,31] Andrés N. Ruiz,[32,33] Chaichalit Srisawat,[6] Adam R. H. Stevens,[16,20] Edouard Tollet,[34,35] Cristian A. Vega-Martínez[14] and Sukyoung K. Yi[36]

Affiliations are listed at the end of the paper

Accepted 2017 December 15. Received 2017 December 15; in original form 2017 August 29

ABSTRACT

We present a comparison of nine galaxy formation models, eight semi-analytical, and one halo occupation distribution model, run on the same underlying cold dark matter simulation (cosmological box of comoving width $125h^{-1}$ Mpc, with a dark-matter particle mass of $1.24 \times 10^9 h^{-1}\mathrm{M}_\odot$) and the same merger trees. While their free parameters have been calibrated to the same observational data sets using two approaches, they nevertheless retain some 'memory' of any previous calibration that served as the starting point (especially for the manually tuned models). For the first calibration, models reproduce the observed $z = 0$ galaxy stellar mass function (SMF) within 3σ. The second calibration extended the observational data to include the $z = 2$ SMF alongside the $z \sim 0$ star formation rate function, cold gas mass, and the black hole–bulge mass relation. Encapsulating the observed evolution of the SMF from $z = 2$ to 0 is found to be very hard within the context of the physics currently included in the models. We finally use our calibrated models to study the evolution of the stellar-to-halo mass (SHM) ratio. For all models, we find that the peak value of the SHM relation decreases with redshift. However, the trends seen for the evolution of the peak position as well as the mean scatter in the SHM relation are rather weak and strongly model dependent. Both the calibration data sets and model results are publicly available.

Key words: methods: analytical – methods: numerical – galaxies: evolution – galaxies: haloes – dark matter – cosmology: theory.

1 INTRODUCTION

Galaxy formation is one of the most complex phenomena in astrophysics as it involves physical processes that operate and interact on scales from the large-scale structure of the Universe (i.e. several Gpcs) down to the sizes of black holes (BHs, i.e. sub-parsec scales). This enormous dynamic range in scale prevents us from modelling galaxies 'ab initio' and therefore any model of galaxy formation depends upon a number of recipes that encode all the physical processes we believe are relevant at all those different scales (see e.g. Baugh 2006; Frenk & White 2012; Silk & Mamon 2012; Somerville & Davé 2015, for recent reviews). These recipes are not precisely known but are each regulated by parameters that are chosen to satisfy observational constraints. This is primarily accomplished by means of one-point functions – such as stellar mass or luminosity functions, the black hole–bulge mass relation (BHBM), the star formation rate density (SFRD), etc. (e.g. Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Croton et al. 2006;

★ E-mail: alexander.knebe@uam.es

© 2017 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

Bower et al. 2006a; Monaco, Fontanot & Taffoni 2007; Guo et al. 2013), although the first steps have been taken in the direction of extending this to two-point functions (e.g. the two-point correlation function of galaxies, van Daalen et al. 2015).

To calibrate the parameters of a galaxy formation model one picks an observational data set and adjusts the free parameters (noting that some parameters might be fixed during that process) until a sufficient match is obtained (e.g. Henriques et al. 2009a). Here, 'sufficient match' depends on several factors, including the objectives of the science project the particular galaxy formation models are being developed for. Some models are designed to reproduce certain observations better than others – at the expense of matching the latter. And as we will see later, simultaneously matching multiple observations adds additional degrees of freedom in how to weight the various calibration data sets. In that regard an idealized project comparing galaxy formation codes would use the same automated tuning method on all the models, as well as defining both a clear weighting scheme for the different observations and criteria for calibration. While it would be interesting to adopt such an approach of attaching all the models to the same automated tuning engine, we leave this for a future study. We note that several of the methods incorporated here do not presently contain a fully automated calibration procedure and hence insisting on this approach would have severely limited the scope of this project.

Any observational data input as a calibration constraint cannot be viewed as an output prediction of the model. But in that regards, it is important to note that having a model that self-consistently matches the calibration data is a non-trivial and interesting exercise; it shows that there is a plausible physical model that is consistent with the observed Universe – at least as described by the calibration data. However, properties independent of those used for model calibration can be considered genuine predictions. Of course it is usually the case that observational properties depend somewhat on one another (e.g. the 'fundamental metallicity relation' Lara-López et al. 2010; Mannucci et al. 2010; Salim et al. 2015; Lagos et al. 2016). The extent of this intrinsic overlap needs to be judged when considering the strength of the prediction. Such an approach separates the resulting galaxy properties into two broad categories: the 'prescriptions' and the 'predictions'. The extent to which a model 'prescribes' depends largely upon what it is intended to be used for. A well-calibrated highly prescribed model will be guaranteed to match a wide range of observations *by construction* and may be highly desirable for testing observational pipelines. Alternatively, a model with few prescriptions might be more suitable for examining where the physics of galaxy formation breaks down as there will be a wide range of available predictions.

In a previous study (Knebe et al. 2015), we applied 14 different galaxy formation models to the same underlying cosmological simulation and the merger trees derived from it. Those models were used with their published parameters and not recalibrated to the new simulation data. We have seen that using models *as is* introduces model-to-model variations larger than reported in previous comparisons (e.g. Fontanot et al. 2009; Lu et al. 2014). We attributed this to the missing recalibration to the new simulation and merger trees used for that particular study. In this work, we now both recalibrate the models and introduce a unified set of observations to produce a common calibration. We investigate how well multiple simultaneously applied constraints are reproduced by the models (i.e. the 'prescriptions') and also study how calibrating to different data sets affects some properties that were not used during the calibration.

We stress that the aim of this work is not to investigate the calibration process of galaxy formation models which has been subject of previous works (e.g. Benson 2014; Rodrigues, Vernon & Bower 2017). Our prime objective is to obtain models that can be used to study average, statistical properties of galaxies representative of the observable population. Further, we want to understand what is needed to achieve that goal. The common ground for all our models is the same simulation, merger tree, and the observational constraints used for calibration. We further fixed a few more ingredients that enter into each of the models such as the initial mass function (IMF) of stars, stellar yields (how chemical elements are produced in stars), and the recycled fraction (the fraction of gas available for new star formation). We are therefore left with a selection of galaxy formation models that use similar assumptions, but are different in design and how to model galaxy formation. We are additionally including a halo occupation distribution (HOD) model (MICE, Carretero et al. 2015) in the comparison: as such models relate numerical data (for a given redshift) to observations in a statistical manner, they provide – by construction – an accurate reproduction of the galactic content of haloes . While they do not provide galaxy populations that self-consistently evolve in time, they nevertheless have great value when it comes to interpreting data from on-going and future galaxy redshift surveys (especially for clustering studies, Pujol et al. 2017). In fact, the MICE model has been used to generate the flagship galaxy mock catalogue for the *Euclid* satellite mission (which is available at the COSMOHUB[1] data base).

We have made all the data for this project publicly available. A link for the observational data used for the common calibration (i.e. the so-called CARNage calibration set) can be found in Appendix A), and the resulting galaxy catalogues (ca. 40GB of data) are stored on a data server to which access will be granted upon request to the leading author.

2 THE SIMULATION DATA

The halo catalogues used for this paper are extracted from 125 snapshots[2] of a cosmological dark-matter-only simulation undertaken using the GADGET-3 *N*-body code (Springel 2005) with initial conditions drawn from the *Planck* cosmology (Planck Collaboration et al. 2014, $\Omega_{\rm m} = 0.307$, $\Omega_\Lambda = 0.693$, $\Omega_{\rm b} = 0.048$, $\sigma_8 = 0.829$, $h = 0.677$, and $n_{\rm s} = 0.96$). We use 512^3 particles in a box of comoving width $125h^{-1}$ Mpc, with a dark-matter particle mass of $1.24 \times 10^9 h^{-1}{\rm M}_\odot$. Haloes were identified with ROCKSTAR (Behroozi, Wechsler & Wu 2013a) and merger trees generated with the CONSISTENTTREES code (Behroozi et al. 2011). Even though both halo finder and tree builder have changed with respect to Knebe et al. (2015), the files passed to the modellers were in the same format. Essentially, these changes have improved the underlying simulation framework upon which this work is based: the box is bigger, there are more snapshots, the halo catalogues are more complete at early times and the tree construction includes patching when objects briefly disappear between snapshots.

3 GALAXY CATALOGUES

The nine participating galaxy formation models are listed in Table 1. While we include a reference to the paper where the model and its parameters are presented in detail, we also refer the reader to

[1] https://cosmohub.pic.es

[2] It has been shown that the number of snapshots as chosen here is sufficient to achieve convergence to within 5 per cent for galaxy masses (Benson et al. 2012).

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

Table 1. Participating galaxy formation models. We list here the reference where the model used in this work has been introduced and where the original calibration details can be found. We further provide columns summarizing the calibration approach (MCMC and PSO) and comments that are of relevance for the data sets used in this work.

Code name	Reference	Calibration approach	Comments
DLB07	De Lucia & Blaizot (2007)	Manual	SMF $z = 2$ has not been used for '-c02'
GALICS 2.0	Knebe et al. (2015)	Manual	'-c01' and '-c02' data sets are identical
GALFORM-GP14	Gonzalez-Perez et al. (2014)	Manual	—
LGALAXIES	Henriques et al. (2013)	Automated: MCMC	For '-c01' also SMF $z = 2$ has been used
MICE	Carretero et al. (2015)	Manual	neither CGMF nor BHBM for '-c02'
MORGANA	Monaco et al. (2007)	Manual	SMF $z = 2$ has not been used for '-c02'
SAG	Gargiulo et al. (2015), Cora et al. (2018)	Automated: PSO	—
SAGE	Croton et al. (2016)	Manual	'-c01' and '-c02' data sets are identical
YSAM	Lee & Yi (2013)	Manual	—

the appendix of Knebe et al. (2015) where all the models have been summarized in a concise and unified manner. However, we need to mention that the SAG model used here differs from the previous version and corresponds to the one presented in Cora et al. (2018): the model includes a revised supernova feedback scheme and a detailed modelling of environmental effects coupled with an improved orbital evolution of orphan galaxies. Further, the GALICS 2.0 model used here is exactly the one described in the appendix of Knebe et al. (2015) and *not* the one presented in Cattaneo et al. (2017).

The third column in Table 1 provides information about whether the model has been calibrated manually or using some automated technique (to be discussed in more detail below). The last column in Table 1 provides some remarks about particulars affecting the common calibration strategy.

Each model has been applied to the simulation data generating three distinct galaxy catalogues:

(i) **uncalibrated (uc)**: as has been done in Knebe et al. (2015) each model was applied without any recalibration to the new simulation, merger trees, and the common assumptions detailed in Section 3.1.

(ii) **calibration #1 (c01, SMF at $z = 0$)**: the models were calibrated to the provided SMF at $z = 0$ in the range around the SMF knee, $9.95 \leq \log_{10}(M_*/M_\odot) \leq 11.15$. We will refer to this as calibration '-c01'.

(iii) **calibration #2 (c02, SMF at $z = 0$ and 2 + extra constraints)**: in addition to the constraint used for calibration #1, the BHBM and cold gas mass at $z = 0$, the star formation rate function (SFRF) at $z = 0.15$, and the stellar mass function (SMF) at $z = 2$ have been used. We will from now on refer to this either as calibration '-c02' or simply the 'CARNage calibration'. This observational data set is motivated and described in detail in Section 3.3.

Note that the uncalibrated model will only serve as a connecting point to our previous comparison (Knebe et al. 2015) and will not enter the main part of this work. And even though results from the two calibration approaches are not mixed together in a single plot, we also chose to use different line styles for these two catalogues: results from '-c01' will always be presented as dashed lines, whereas the results from '-c02' are shown as solid lines. This facilitates comparison and allows for a quicker identification of calibrations in the plots.

3.1 Common modelling

In this work, we have populated one simulation with eight different semi-analytical models (SAMs) of galaxy formation and evolution and one halo occupation model. In order to further align the various galaxy formation models they all assumed a Chabrier IMF, a metallicity yield of 0.02 and a recycled fraction of 0.43.[3] The first uncalibrated ('-uc') stage has populated the simulation described in Section 2 using the models with parameters as previously published (see Table 1 for the list of references), but with a *Planck* cosmology and the assumptions just mentioned with regards to IMF, yield and recycle fraction. The results from this uncalibrated comparison are presented in Appendix B which demonstrates that changing the simulation and merger trees has not changed our findings in Knebe et al. (2015) and that the same results as those obtained previously are recovered, respectively.

3.2 Calibration

Models of galaxy formation and evolution provide a research tool that can be used to explore a vast range of dynamical scales, from stellar nurseries to the large-scale structures seen in the observed Universe. Covering such an enormous dynamical range in scales implies that approximations are needed in order to reduce the computational cost to a feasible level. Moreover, not all the relevant physical processes are well understood and constrained observationally. Thus, any model of galaxy formation has free parameters or parameters that can vary within a reasonable range constrained by direct observations. These free parameters summarize our ignorance with respect to the physical processes pertinent for the formation of galaxies. They are chosen by comparing certain model properties to observations. This choice depends on the physical processes and cosmic times one is interested in. We refer to calibration as the process of choosing these free parameters.

For this work, two sets of observational data were given, one for '-c01' and one for '-c02', but the specific calibration approach was left to each group of modellers, so it could be close to those used for their published models.

3.2.1 Calibration approach

The calibration approach can be split into two distinct categories: manual and automated. In Table 1, we provide this information for each of the models and describe these two classes here in more detail:

[3] Since the SAG code does not use an instantaneous recycling approximation but relies on a set of tables with yields and ages for stars in different mass ranges, the yield and recycled fraction are not fixed to these values.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

(i) *Manual calibration*: this calibration approach explores the parameter space in a manual way, basing the choices of parameters to be varied on previous knowledge (e.g. Lacey et al. 2016). When this approach is used, only a handful of the total free model parameters are modified. The others are inherited from previous work and thus, they have intrinsically been chosen with respect to a certain set of observations. For instance, the SMF at $z = 0$ only constrains physical processes related to the star formation and feedback in galaxies, but not other processes such as the growth of BHs or the growth of bulges (Rodrigues et al. 2017). However, in '-c01', the models that were calibrated manually are leaving the parameters controlling these later processes as they were in previous incarnations of the models and thus, indirectly inheriting previous knowledge.

(ii) *Automated calibration*: this calibration approach explores either the whole parameter space or a subset in a numerical manner. For this work, there are models either using Monte Carlo Markov Chains (MCMC, Henriques et al. 2009a, used by LGALAXIES) or a particle swarm optimization (PSO, Ruiz et al. 2015, used by SAG). When models are calibrated using a numerical exploration of the parameter space, there is a choice on which parameters are set free. If some parameters are set free, but no adequate observable is used during the calibration, the resulting best set of parameters would not be a robust choice as they were basically unconstrained by the choice of the observational data set. While PSO focuses on quickly finding a best-fitting model given the observational constraints and chosen priors, the MCMC approach enables to have a full understanding of the statistical significance of a given set of parameters. However, in this case, special care needs to be taken on how the observational errors are considered (e.g. Benson 2014). Beyond these two numerical approaches, SAMs have also been calibrated using emulators (Bower et al. 2010), an approach not used by any of the models presented here.

As detailed above, the calibration process often entails that the models retain some 'memory' of any previous calibration; this is especially true for the models that are tuned manually. This may particularly affect the 'prescriptions' and 'predictions' of the models. If a model usually uses, for instance, the evolution of the cosmic star formation rate density (cSFRD) or the stellar-to-halo mass relation (SHM) as one of its constraints for the parameters, this might still be reflected in the catalogues presented here. Practically what this means is that in general each model did not start each calibration process with an entirely clean slate but rather began with a set of parameters previously understood to produce a not completely unreasonable result.

3.2.2 Calibration steps and parameter changes

The calibration #1 catalogue ('-c01') starts from the parameters used for '-uc' (i.e. the original parameters presented in the model's reference publication) and uses the observed SMF at redshift $z = 0$ as the only constraint. The calibration #2 catalogue ('-c02') starts from the parameter values found for '-c01', but now adds the provided SFRF, cold gas mass fraction (CGMF), and BHBM relations at redshift $z \sim 0$ as well as the SMF at $z = 2$ to the constraints (see Section 3.3 for further details). The intention here was to add the minimum number of additional calibration data sets which simultaneously constrained the key physical processes required for a model of galaxy formation. In order to limit the number of different calibrations for each model, we decided to simultaneously apply all four of the constraints in '-c02'. At this stage, each modeller was given the freedom to assign weight to each of the additional constraints as they saw fit. The idea here was not to expect a perfect fit to all constraints but rather to provide a 'best case' solution to the entirety of the constraints.

Here, there is a brief discussion of the changes seen in the model parameters when going from '-uc' to '-c01' and eventually '-c02'. Most of the models applied changes to the same parameters when calibrating to '-c01' and '-c02'. The only exception to this was GALFORM-GP14 where active galactic nuclei (AGN) feedback was changed for '-c01', but not for '-c02' and the level of stellar feedback only changed for '-c02'. In general, the parameter changes required primarily focused on changing the feedback (either stellar or AGN) and/or the treatment of mergers. The latter impacts upon bulge, disc, and BH growth which are amongst the parameters modified by most (but not all) modellers. However, the only changes for GALICS 2.0 were the limiting mass above which the accretion of gas on to galaxies is shut down (the $M_{\rm shock\,max}$ parameter in appendix 2 of Knebe et al. 2015) and a lowering of all outflow rates by a factor $\sim$1.3. The HOD model MICE starts from the galaxy luminosity function and then converts it to stellar masses: for '-c01' this conversion has been updated to obtain a better fit to the provided SMF at $z = 0$. For '-c02', this has been implemented even more self-consistently for all redshifts while also changing the calibration of the SFR. We close by mentioning that some of the models only changed a few parameters for the calibration sequence (e.g. YSAM only adjusted four parameters), whereas models applying an automated calibration process feature changes in substantially more parameters (e.g. for LGALAXIES seven parameters were varied).

3.3 The 'CARNage calibration' data set

The observational data sets used for the calibration are detailed in Appendix A where we also provide a link for download. They are designed to constrain different aspects of galaxy formation and evolution, yet observationally are well established. Galaxy formation models aim at encapsulating the main physical processes governing galaxy formation and evolution in a set of coupled parametrized differential equations. These parameters are not arbitrary but set the efficiency of the various physical processes being modelled – and they have to be tuned using observational data. All models are hence calibrated against a key set of observables by which different physics in the models are fixed – depending on the actual observations used. The observational data sets used for the 'CARNage calibration' where chosen in a way to be as complementary as possible (references are given in Table 2):

Stellar mass function: the SMF at $z = 0$ is a fundamental property that can constraint alone much of the stellar formation and feedback processes that shape the formation and evolution of galaxies. In addition to the observed SMF at $z = 0$ (that formed the basis of calibration '-c01'), we added the observed SMF at $z = 2$. While this is still somewhat observationally uncertain, it provides a constraint on the evolution of the stellar mass assembly. For the process of calibration, we further agreed that when tuning to the SMF, the model stellar masses should be convolved with a $0.08(1 + z)$ dex scatter to account for at least part of the observational errors in measuring stellar masses. Note, while each code wrote to the resulting output files the internally calculated stellar masses for each galaxy, these masses were subjected to this scatter only during the calibration process. And in order to mimic this effect when again comparing the model stellar masses to observations, we modify them in the same way while reading them in from the galaxy catalogues. i.e. our analysis pipeline convolved the model data with aforementioned scatter before comparing to observations.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

Table 2. Observational data sets used for the common calibration approach: SMF, SFRF, CGMF, and BHBM. The actual data used for this paper can be downloaded using the hyperlink provided in Appendix A. The last column indicates for which calibration set data was used.

Observation	Redshift	Reference	Calibration
SMF	$z = 0$	Baldry et al. (2012); Li & White (2009); Baldry, Glazebrook & Driver (2008)	c01 + c02
SMF	$z = 2$	Tomczak et al. (2014); Muzzin et al. (2013); Ilbert et al. (2013); Domínguez Sánchez et al. (2011)	c02
SFRF	$z = 0.15$	Gruppioni et al. (2015)	c02
CGMF	$z = 0$	Boselli et al. (2014)	c02
BHBM	$z = 0$	McConnell & Ma (2013); Kormendy & Ho (2013)	c02

Star formation rate function: while the SMF constrains the amount of mass in stars, the SFRF quantifies the change in the SMF as a function of time. It provides information on the efficiency of transforming (cold) gas into stars (and the fraction of stars whose mass is lost back as gas) in a given galaxy and hence drives any shape changes or evolution of the SMF. In the models, the cold gas fraction (CGF) at low redshift is thus determined both by the star formation law and by the effectiveness of quenching processes, such as AGN feedback and stripping, that remove cold gas from galaxies.
Cold gas mass fraction: stars can only form when sufficient cold gas is present in a galaxy. Therefore, an important tracer for star formation is the fraction of cold gas to stellar mass.
Black hole–bulge mass relation: the observed relation between the mass of a galaxy's central BH and the mass (or velocity dispersion) of a galaxy's bulge suggests that the central BH plays a key role in galaxy evolution. The BHBM relation is used to constrain BH masses as there is otherwise a large degeneracy between the BH mass and the AGN feedback efficiency in the models (Henriques et al. 2009b; Croton et al. 2006; Bower et al. 2006b; Mutch, Poole & Croton 2013; Croton et al. 2016): BH growth plays a critical role in galaxy evolution (e.g. Cheung et al. 2016, for recent observational results from the Mapping Nearby Galaxies at Apache Point Observatory (MaNGA) survey).

We refrained from using galaxy luminosities, even though they are directly observable, because their calculation requires an additional layer of modelling which adds extra complexity to the comparison – something to be avoided for this paper and left for a future study.

3.4 Common analysis

All provided galaxy catalogues have been post-processed with a common analysis pipeline that is also made available alongside the numerical and observational data sets. While catalogues contain galaxies with much smaller masses, we limit all of the comparisons presented here to galaxies with stellar masses $M_* > 10^9 h^{-1}$ M$_\odot$ – a mass threshold appropriate for simulations with a resolution in dark-matter mass comparable to the Millennium simulation (see Guo et al. 2011).

We further remark that the points for the models have been obtained by binning the data using logarithmically spaced bins on the x-axis and then calculating the median in that bin for both the x- and y-values.

4 CALIBRATION WITH THE LOCAL SMF

Calibration #1 presented in this section only uses the SMF at redshift $z = 0$ as a constraint and we will explore its *prescription* in Section 4.1. We continue in Section 4.2 to examine these '-c01' models compared to the observations used later on for calibration '-c02'.

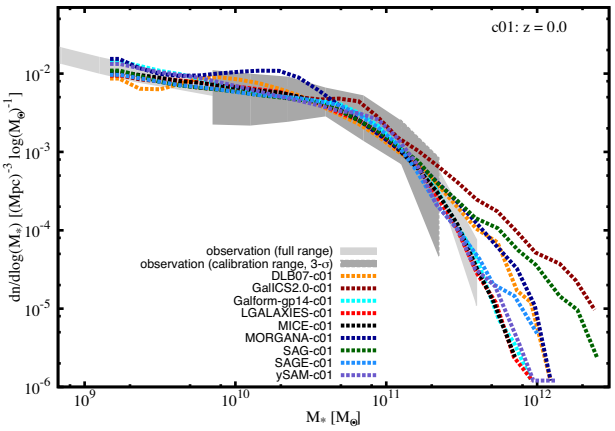

Figure 1. SMFat redshift $z = 0$ for calibration #1 (c01) compared to the observational data as described in detail in Appendix A1. The light-shaded region shows the compilation and its 1σ errors, whereas the dark-shaded region shows the mass range used to calibrate the models and its 3σ errors.

Table 3. χ^2-values for models with respects to observational data. The first column after the model name is for the SMF at $z = 0$ in the calibration '-c01' data set; the following two columns are for the SMF at $z = 0$ and 2, respectively, in the calibration '-c02' data set. Note, only the range used during model calibration has entered into the χ^2-calculation.

Model	$\chi^2_{z=0,c01}$	$\chi^2_{z=0,c02}$	$\chi^2_{z=2,c02}$
DLB07	3.0	30. 0	41.0
GALICS 2.0	0.8	0.8	19.0
GALFORM-GP14	0.9	3.4	14.0
LGALAXIES	0.7	1.7	2.0
MICE	0.5	0.9	0.91
MORGANA	11.7	9.2	83.0
SAG	0.5	2.1	0.19
SAGE	0.2	0.2	15.0
YSAM	1.4	5.7	44.0

4.1 The calibration

For calibration #1, the individual model parameters have been tuned to the same (compiled) SMF (in the mass range $9.95 \leq \log_{10}(M_*/M_\odot) \leq 11.15$).[4] We compare the '-c01' models against the used observational SMF with the latter shown in Fig. 1 as a light-shaded region (see Appendix A1 for more details about this compilation of observations); we additionally indicate the mass range used for the model calibration as a dark-shaded region (3σ error bars). Most of our galaxy formation models lie well within

[4] Remember, this mass range – bracketing the knee of the SMF – was agreed upon by all modellers during the Cosmic CARNage workshop.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

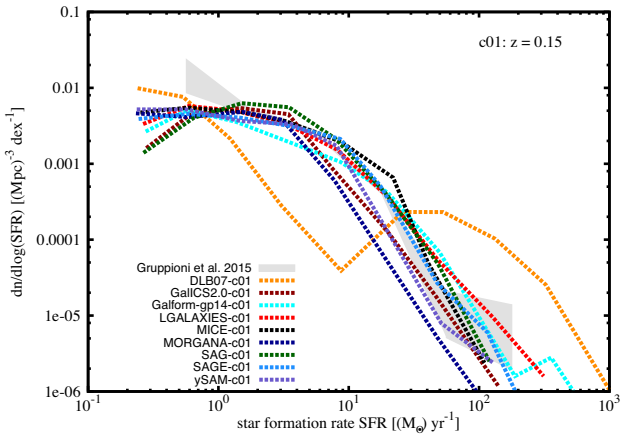

Figure 2. SFRF at redshift $z = 0.15$ for calibration #1 (c01). The shaded regions shows the observations and its 1σ errors.

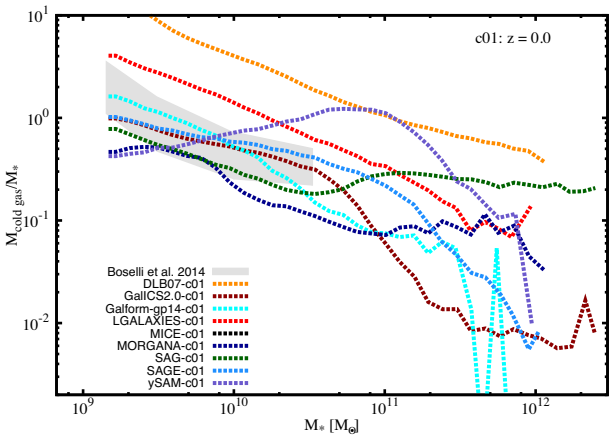

Figure 3. CGF as a function of stellar mass at redshift $z = 0$ for calibration #1 (c01). The shaded regions shows the observations and its 1σ errors.

that '3σ area' as indicated by the low values given in Table 3 where we list the respective χ^2-values (alongside the corresponding values for the other calibration and redshifts to be discussed later). Comparing this to the equivalent fig. 2 of Knebe et al. (2015, as well as the uncalibrated analogue in Appendix B), we see a clear tightening when using the SMF at $z = 0$ as a common constraint. The most apparent deviations from the observations and amongst the models are at the high-mass end where there are only a few objects with little constraining power: the data in that range is dominated by systematic errors which is why this range has been excluded from the calibration set.

4.2 Beyond the calibration

Even though the models have not been (actively) calibrated against anything else but the SMF at redshift $z = 0$, we will now study the model data for the reminder of the observational 'CARNage calibration' set. This allows us to investigate how the residual scatter seen for the SMF in Fig. 1 propagates into other galaxy properties and their respective correlations. For most of the subsequent plots in this subsection (except the BHBM relation), the 'CARNage set' will be represented as a shaded region indicating 1σ error bars.

4.2.1 Star formation rate function

In Fig. 2, we present the SFR distribution function, i.e. the number density of galaxies in a certain SFR interval compared to the observations of Gruppioni et al. (2015). Within the range of models considered here, most of them already lie close to the observations and they follow the same general trend even *before* this is used as a constraint, i.e. the form of the SFRF is largely already imposed by the requirement of matching the SMF at $z = 0$. However, while SFR and stellar mass are certainly connected, this relation has to be viewed with care because of the recycle fraction of exploding stars: the integral over the SFR is not necessarily the total stellar mass. And we have found (though not explicitly shown here) that requesting the DLB07 model (and to a lesser extent MORGANA, too) to fit the provided SMF at $z = 0$ degrades the quality of the matching to the SFRF: when using the uncalibrated '-uc' data for DLB07 (and MORGANA), we find that the SFRF for the two models lies within the observed 1σ range (not explicitly shown here).

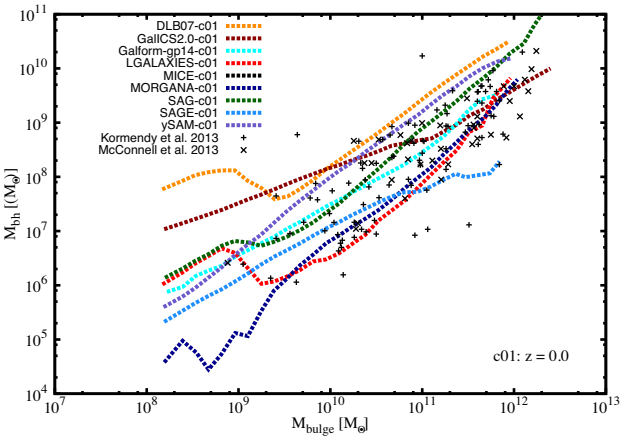

Figure 4. BH versus bulge mass at redshift $z = 0$ for calibration #1 (c01) alongside the observational data.

4.2.2 Cold gas mass fraction

The CGF in galaxies as a function of stellar mass is shown in Fig. 3. In this figure, model galaxies are compared to the observational data from Boselli et al. (2014). Almost all the models (bar YSAM) reproduce the declining trend seen in the observations. However, leaving this property unconstrained leads to a substantial model-to-model variation amplitude-wise. We can already deduce some interesting conclusions from this figure: for instance, if unconstrained by the CGMF both DLB07 and LGALAXIES would prefer higher CGMFs than observed.

4.2.3 Black hole–bulge mass relation

In Fig. 4, we examine how the BHBM relation is reproduced by the models without constraining to it – in comparison to the observational data of McConnell & Ma (2013) and Kormendy & Ho (2013). While one might conclude that such an agreement is related to the requirement of fitting the SMF at $z = 0$, we confirm (though not explicitly shown here) that using the uncalibrated '-uc' data set gives a plot very similar to Fig. 4. This could be interpreted as the BHBM being largely insensitive to the parameters governing the SMF.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

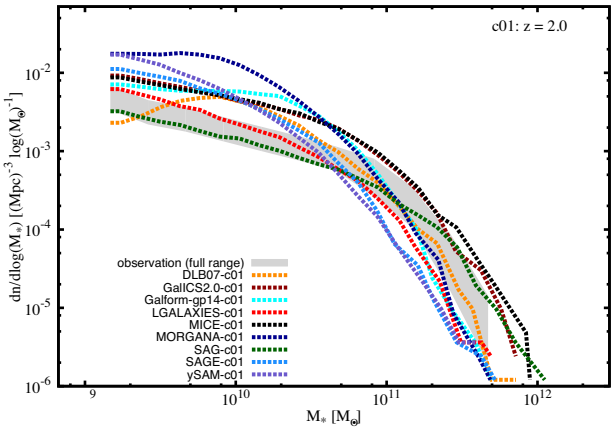

Figure 5. SMF at redshift $z = 2$ for calibration #1 (c01). The shaded regions shows the observations and its 1σ errors.

4.2.4 Stellar mass function at z = 2

All the previous observational data sets were restricted to (or close to) redshift $z = 0$. We now extend our investigations to higher redshifts by considering the SMF at $z = 2$. We can observe in Fig. 5 that this poses a challenge for the majority (if not all) of the models. The scatter is considerably larger than for redshift $z = 0$. We reconfirm that reproducing high-redshift observations is a challenge for most models: only SAG and LGALAXIES lie within the 1σ error range and in both cases the physics in the model was tuned to reproduce the SMF evolution (Henriques et al. 2014).

We see that all the other models lie above the observations at small-to-intermediate masses, indicating that when unconstrained by this observation they predict the presence of a large number of small objects that are not observed.

4.3 Discussion

In this work, we required all models to use the same observational data during parameter calibration and, for this first approach, only use the SMF at redshift $z = 0$. We find that the scatter seen in fig. 2 of Knebe et al. (2015, and also Fig. B1 here) substantially tightens and now lies within the 3σ error bars of the observations – at least for the mass range considered during the calibration. As another example, the model SFRF and the CGF follow the observational trends reasonably well, albeit still showing pronounced model-to-model variations. We remark that while for some of the models, the change from the uncalibrated data set to calibration '-c01' clearly improved the match to the SMF at $z = 0$, this was accompanied by a degrading of the match to other observational data. This is particularly prominent for the CGF where the uncalibrated data set (not shown here) shows far less model-to-model variation than seen in Fig. 3.

We have further found that the model SMF at redshift $z = 2$ exhibits scatter to the same degree as found for models when not recalibrated (cf. fig. 2 in Knebe et al. 2015, again). We reconfirm the well-known problem that galaxy formation models readily overproduce low-mass galaxies at high redshift (e.g. Fontanot et al. 2009; Weinmann et al. 2012; Somerville & Davé 2015; Hirschmann, De Lucia & Fontanot 2016).

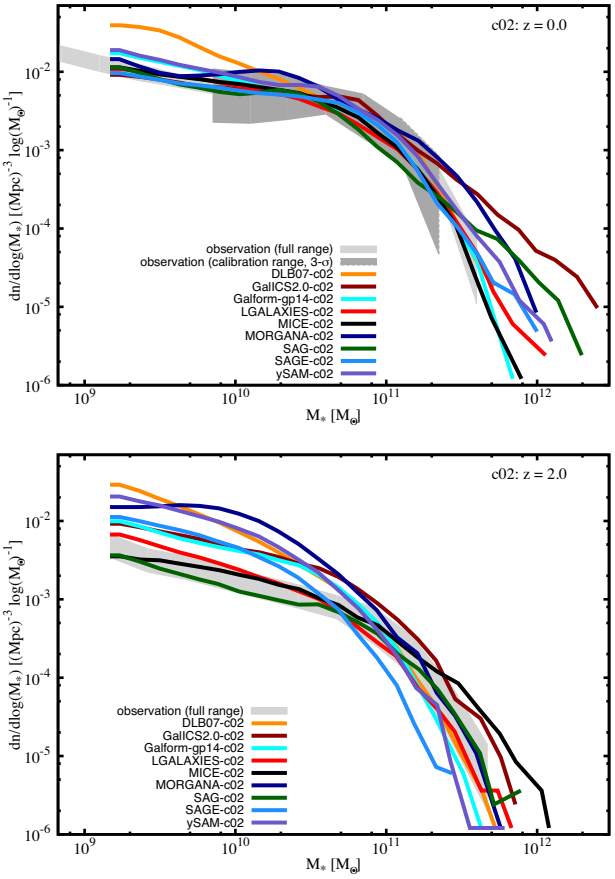

Figure 6. SMF at redshift $z = 0$ (top panel) and $z = 2$ (bottom panel) for calibration #2 (c02).

5 CALIBRATION WITH THE CARNage DATA SET

The 'CARNage calibration' data set has been introduced and motivated already in Section 3.3 and its details (including a link to a public data base) are given in Appendix A. All models have now either manually or automatically tuned their parameters with that particular set simultaneously. However, the modellers were given the freedom to assign weights to each observation individually: different models might be designed to perform better for some predictions/prescriptions and hence put more emphasis on reproducing, for instance, the BHBM as opposed to the CGF. In passing, we note that there is no difference between the '-c01' and '-c02' galaxy catalogues for GALICS 2.0 and SAGE: their respective catalogues are based upon the same set of calibration parameters.[5]

5.1 The calibration

5.1.1 Stellar mass function at z = 0 *and* 2

In Fig. 6, we show the SMF as given by the models for both redshift $z = 0$ (upper panel) and $z = 2$ (lower panel) in comparison to the observational data. These plots are again accompanied by the respective χ^2-values listed in Table 3. We find that adding the new

[5] While all models went through the '-c02'-calibration process, both GALICS 2.0 and SAGE eventually realized that the best parameters actually agree with the ones already obtained during the '-c01' calibration.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

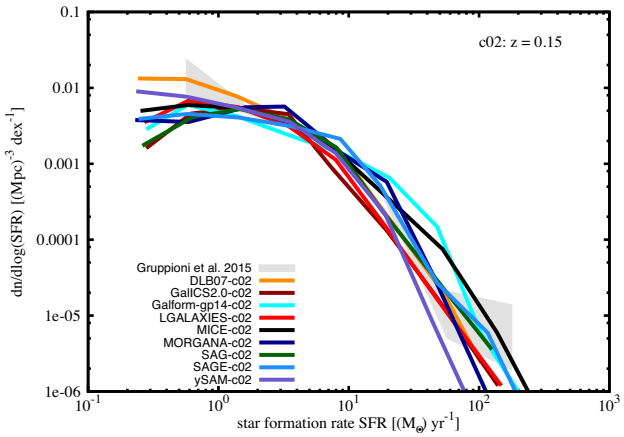

Figure 7. SFRF at redshift $z = 0.15$ for calibration #2 (c02).

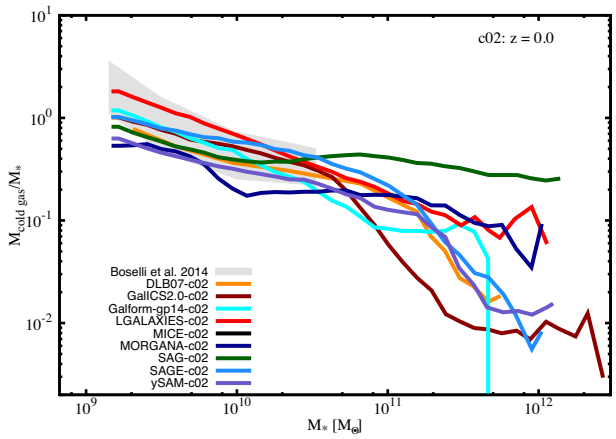

Figure 8. CGF as a function of stellar mass at redshift $z = 0$ for calibration #2 (c02).

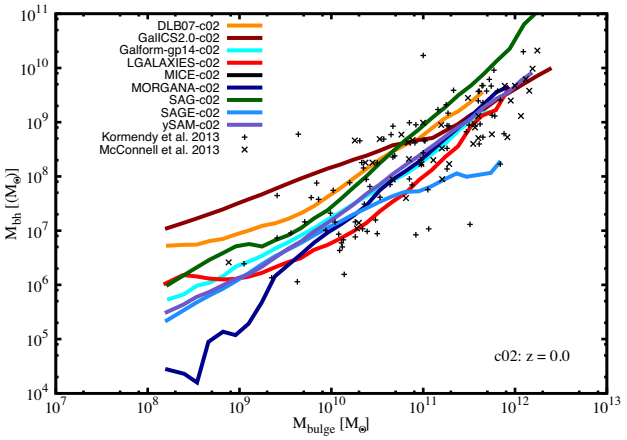

Figure 9. BH versus bulge mass at redshift $z = 0$ for calibration #2 (c02).

constraints (including one at higher redshift) reduces the agreement at redshift $z = 0$ for most models with the scatter between the models clearly increased. This scatter now spans the 3σ band at redshift $z = 0$ (for the considered mass range, see Section 4.1) and they still show substantial variation at higher redshift: at $z = 2$ only the LGALAXIES, SAG, and MICE models lie close to the observational band across all stellar masses.

Even when constraining the SMF at $z = 2$, most models clearly overproduce galaxies at the low-mass end at $z = 2$, as already noted before. Suffice it to say that the $z = 2$ SMF provides significant tension and it is already well known that it is difficult to concurrently obtain good fits to the SMF at both redshifts (e.g. Fontanot et al. 2009; Weinmann et al. 2012; Somerville & Davé 2015; Hirschmann et al. 2016). A more in-depth study of the physics of simultaneously matching the redshift $z = 0$ and 2 SMF will be presented in a companion paper (paper II, Asquith et al., in preparation). That work reasserts that tension exists to some extent in all the SAMs of galaxy formation studied here. In that paper, we investigate the evolution of the SMF with redshift for all galaxies (passive and star forming) up to $z = 3$ and find that all the models, despite the wide range of physical processes implemented, produce too many small galaxies at high redshift. These excess galaxies appear to be mainly star forming and are not present in the latest observations (Mortlock et al. 2015; Muzzin et al. 2013).

But we also noted before (see Section 4.2.1) that there is an interplay between matching the SMF and SFRF, especially for the DLB07 and MORGANA models. And the increased model-to-model variation for the SMF seen here might also be attributed to an improved matching of the SFRF as presented in the following subsection.

5.1.2 Star formation rate function

In Fig. 7, we can appreciate that when adding the SFRF as a constraint (along with the four additional constraints used in stage '-c02') the scatter seen before in Fig. 2 noticeably tightens.

5.1.3 Cold gas mass fraction

We have seen before that leaving the CGF unconstrained leads to a substantial model-to-model variation in amplitude. This is somewhat alleviated by using it as a calibration constraint as can be verified in Fig. 8: all models lie within the 2σ range of observations. The most prominent change happens for YSAM, which had a generally rising trend when the observed CGMF was not used as a constraint.

5.1.4 Black hole–bulge mass relation

For the BHBM relation – as presented in Fig. 9 – we find that all models lie within the observations with a slight tightening of the range when this relation is added as a constraint to the models. Note that in practice several models already included the BHBM relation as one of their usual constraints (and hence keeping its 'memory'); we therefore did not expect a considerable change when adding it as a constraint for data set '-c02'.

The rather large BH masses at the low-$M_{\rm bulge}$ end for GALICS 2.0 – in comparison to the other models and even when adding this relation as a constraint – are related to the treatment of major mergers, which instantaneously converts 1 per cent of the gas into a central BH, while most of the remaining gas is ejected to the high-mass loading factors of low-mass galaxies and hence not available for star formation. In mergers with $M_{\rm gas} \gg M_*$, this assumption can lead to remnants with $M_{\rm bh}/M_{\rm bulge} \gg 0.01$.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

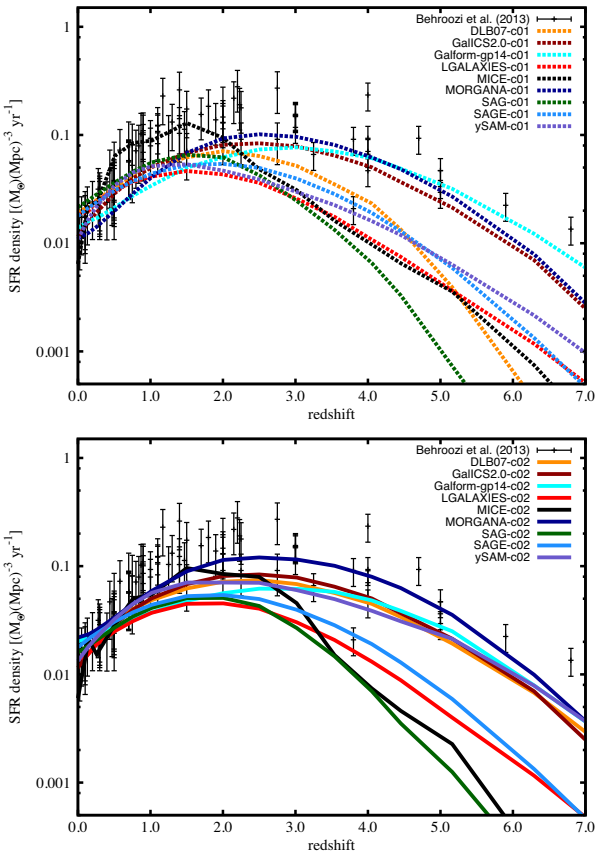

Figure 10. Evolution of the SFRD.

5.2 Beyond calibration

While all distribution functions and correlations (apart from the SMF at $z = 0$) studied for data set '-c01' in Section 4.2 could be considered predictions, the '-c02' data set has been constrained by the local SFRF and SMF as well as the redshift $z = 2$ SMF. But will this be sufficient to 'predict', for instance, the so-called Madau–Lilly plot (Madau & Dickinson 2014; Madau et al. 1996; Lilly et al. 1996), i.e. the evolution of the cSFRD. We note that this plot is closely related to the SMF as well as the SFRF: for instance, the integral over all masses of the SMF at a fixed redshift corresponds to the integral of the cSFRD up to that redshift; further, the integral over all SFR values in the SFRF gives the point in the cSFRD at the corresponding redshift. We have previously seen that matching higher redshift observations is far from trivial. This discrepancy is well known (e.g. Fontanot et al. 2009; Weinmann et al. 2012; Somerville & Davé 2015; Hirschmann et al. 2016) and is somewhat driven by the fact that the integral under the observational curve is inconsistent with the observed stellar density today, a requirement that is enforced in the models (Nagamine et al. 2004; Davé 2009; Wilkins, Trentham & Hopkins 2008), but influenced and modified by the recycled fraction as mentioned before.

Fig. 10 shows the results for the evolution of the cSFRD for each model, with observational data taken from Behroozi, Wechsler & Conroy (2013b). We find that all the models reproduce the form of the SFRD evolution with a pronounced peak and a significant decrease at late times of approximately the observed amplitude. While the model-to-model variations appear to be the same for both calibrations '-c01' and '-c02', we note that individual models substantially change their behaviour from one to the other. For instance, the DLB07 model has a higher SFR at early times in '-c02'. SAG and LGALAXIES are towards the bottom end of the SFRs at early times as is the HOD model, MICE. For LGALAXIES, this is related to a lack of resolution in the provided N-body simulation used here: in Henriques et al. (2014), it has been shown that with the addition of the Millenium-II simulation with an increased mass resolution, the model matches the observations at high-z (but still falls well below at $z = 1$-2). For MICE, this is due to not having calibrated at those high redshifts; it only applied evolutionary correction up to $z \sim 3$. However, these are the three models that provide the best match to the SMF at $z = 2$. We close by mentioning that both the SAGE and YSAM models usually use the cSFRD as a constraint during their calibration; however, they utilize observational data presented in Somerville, Primack & Faber (2001, SAGE) and Panter et al. (2007, YSAM) respectively.

We end this subsection with a word of caution: we applied a general lower limit for galaxies entering our plots, i.e. $M_* > 10^9 h^{-1}$ $\mathrm{M}_\odot$. But this will bias the results presented in Fig. 10 as it leaves out star formation taking place in smaller objects which is even more relevant at early times and high redshifts, respectively. In order to investigate the size of this effect, we have performed two different tests. First, we have lowered the mass threshold in several steps from $M_{*,\mathrm{cut}} = 10^9 h^{-1}$ $\mathrm{M}_\odot$ down to $M_{*,\mathrm{cut}} = 10^6 h^{-1}$ $\mathrm{M}_\odot$ always using the galaxies as provided in the respective catalogue. We confirm that this does not alter the behaviour of the models for redshifts $z < 2$, but increase the SFRD at higher redshifts bringing them into closer agreement with the observations. Second, we performed a more elaborate test to investigate resolution effects entering this plot: instead of simply adding the galaxies below the resolution limit, at all redshifts we fit the SMF to a Schechter function of the form $\mathrm{d}n/\mathrm{dlog}\, M_* = n(M_*/M_0)^p \exp(-M_*/M_0)$ (with free parameters n, p, andM_0) and the relation between SFR and stellar M_* to a power-law $\mathrm{SFR} = A M_*^q$ (with free parameters A, q). These best-fitting functions are then used to add the contribution from galaxies with $M_* < 10^9 h^{-1}$ $\mathrm{M}_\odot$ to the cSFRD given as $\int_{M_*=10^6 h^{-1}\mathrm{M}_\odot}^{M_*=10^9 h^{-1}\mathrm{M}_\odot} \mathrm{SFR}(M_*)\mathrm{d}n/\mathrm{d}M_*\, \mathrm{d}M_*$. We confirm again that the conclusions drawn from Fig. 10 for the comparison between models remain unchanged when post-correcting in this way, but the curves are shifted upwards bringing them into better agreement with the observations. However, as both these methods have significant uncertainties in the correction to be applied, we decided to simply describe their effects rather than incorporating them in the figure.

5.3 Discussion

We have seen in the previous section that constraining the SMF at redshift $z = 0$ alone will already be sufficient to give the right trends in correlations and shapes of distribution functions, but with noticeable model-to-model amplitude variation as well as large offsets to the observations for some models – especially when higher redshifts are considered. In particular, models have difficulty producing the correct number density of low-mass galaxies at higher redshift (Fontanot et al. 2009; Weinmann et al. 2012; Somerville & Davé 2015; Hirschmann et al. 2016) – a problem that has been addressed by adding some form of preventive/ejective stellar feedback (White, Somerville & Ferguson 2015; Hirschmann, De Lucia & Fontanot 2016) and/or modulation of reaccreation (Henriques et al. 2013) to the models in order to recover the evolution of the SMF.

When adding additional (orthogonal) constraints, the scatter generally tightens and models move towards closer agreement with each

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

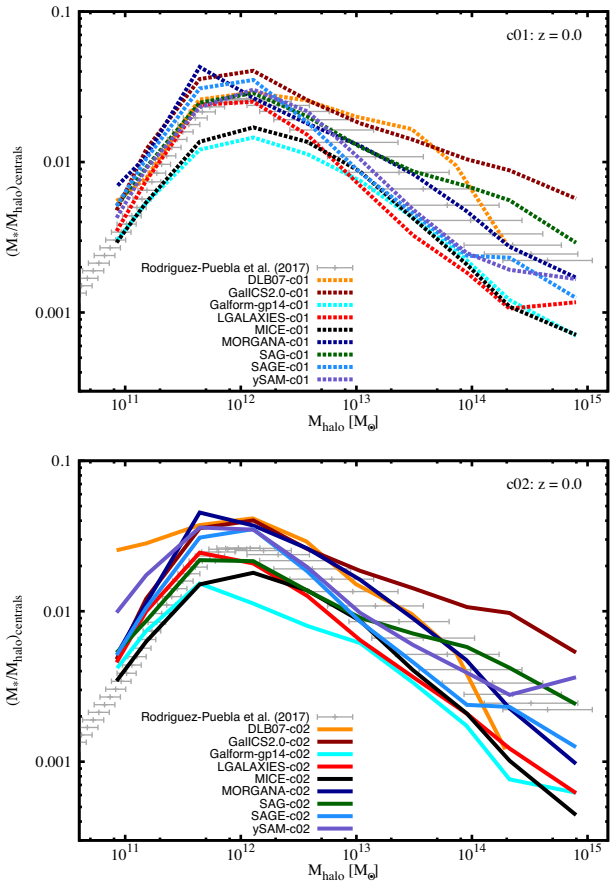

Figure 11. SHM as a function of halo mass for central galaxies only. The values shown are medians in the respective bin.

other. This is most prominent for the CGMF, i.e. the model-to-model variation substantially reduces when comparing Fig. 3 ('-c01' calibration that only uses the SMF at redshift $z = 0$) to Fig. 8 ('-c02' calibration that now includes additional observational data at different redshifts as described in Section 3). Further, the SFRF scatter also decreases going from '-c01' to '-c02'. We can see that there is one model in particular, DLB07, which is a clear outlier in both the CGF and SFRF for '-c01', and shows dramatic improvement going to '-c02'. The other models that have a notable change in CGF have an accompanying change in SFRF – typically gas fractions drop and SFR goes up. However, for the SMF at redshift $z = 0$, we note a marginal increase in the scatter as it is no longer the sole constraint.

When moving to a non-calibrated (yet related) property like the evolution of the SFRD, we find that switching from '-c01' to '-c02' will not tighten the scatter across models. Rather it impacts upon certain models more than others, e.g. DLB07 see an increase in amplitude at higher redshift; as the DLB07 model did not use the SMF at $z = 2$ when calibrating, this improvement is mainly due to the use of SFRF at $z = 2$ as an additional constraint.

We will return to the evolution of the SMF in a companion paper (paper II, Asquith et al., in preparation) where we more closely investigate the evolution of the SMF with redshift for the same models presented here, and we separate the galaxies into 'passive' and 'star-forming' classes.

6 STELLAR-TO-HALO MASS RATIO

All the galaxy formation models presented and studied here populate given dark-matter haloes with galaxies whose properties depend on the particulars of the formation history of the halo they are placed in. Subsequent galaxy evolution then shapes the galaxy SMF leading to the well-known shape that can be roughly described as two power laws: at the low-mass end supernova feedback suppresses star formation, whereas various feedback mechanisms due to the accretion of gas on to a BH are responsible for a suppression of star formation at the high-mass end (see Silk & Mamon 2012, for a succinct review).

None of the galaxy properties in the previous sections have been related to the halo the galaxy resides in. Here, we provide a link between the two by investigating the ratio between SHM as a function of halo mass. This ratio – normalized by the cosmic baryon fraction – can also be interpreted as an 'efficiency of star formation' that depends on halo mass, i.e. how many of the maximally available baryons have been converted into stars. Its correlation with halo mass shows a distinct peak whose position coincides with the knee of the SMF. The temporal evolution of the SHM relation has caught a lot of attention recently: there appears to be no consensus yet whether it is evolving with redshift or not. Some authors claim that the peak position evolves (rises) with increasing redshift (Moster, Naab & White 2013; Behroozi, Wechsler & Conroy 2013b; Leauthaud et al. 2012; Matthee et al. 2017) as opposed to works indicating no such evolution (Hudson et al. 2015; McCracken et al. 2015). Likewise, the same works indicate that the peak value of the SHM relation either evolves (Moster et al. 2013; Hudson et al. 2015; Matthee et al. 2017) or remains constant (Behroozi et al. 2013b; Leauthaud et al. 2012) with redshift. Results stemming from SAMs and hydrodynamical simulations are likely sensitive to the particulars of the modelling (Mitchell et al. 2016). As mentioned before, both stellar and AGN feedback leave their imprint in the SHM relation, but the same also holds for disc instabilities and mergers: a halo with its galaxy falling into another larger halo will see the halo and stellar mass added to the host halo in case of merging – irrespective of the star formation efficiency of the host.

Here, we are addressing this point with the catalogues from our galaxy formation models, but limiting the analysis to central galaxies only. Orphan galaxies – by definition – do not have a dark-matter halo (for a detailed discussion and definition of 'orphan' galaxies and halo mass, respectively, please refer to Knebe et al. 2015); and as subhaloes lose mass while orbiting about their host, their (satellite) galaxy will show a more complex SHM relation and are therefore also excluded from the analysis presented here. Further, while other work has shown that the SHM relation is different for passive and star-forming galaxies (e.g. Mitchell et al. 2014), we leave such a classification for a future investigation.

We further include the scatter in the SHM relation in our study here: while halo and SAMs are based upon the assumption that galaxy evolution is directly related to halo growth, halo mass alone is not sufficient to explain the stellar mass of galaxies. This then naturally leads to a scatter in the SHM relation and, for instance, Matthee et al. (2017) found in the Evolution and Assembly of GaLaxies and their Environments (EAGLE) simulation that this scatter increases with redshift, but also decreases with halo mass. Wang, De Lucia & Weinmann (2013) even claim that how galaxies populate the scatter in the SHM relation plays an important role in determining the correlation functions of galaxies.

For the comparison presented here it is worth remembering again that all galaxy formation models used the same halo catalogues and

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

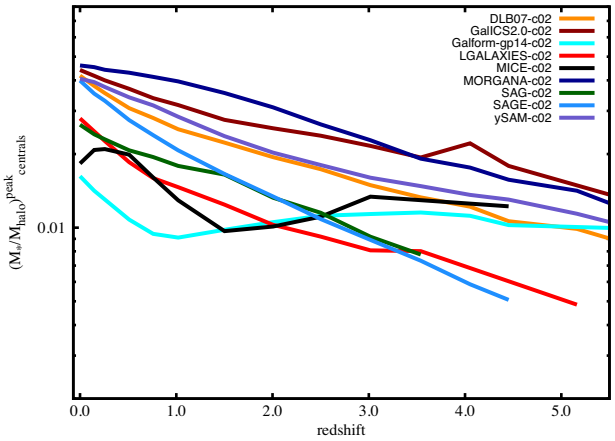

Figure 12. Redshift evolution of the peak value of the SHM relation for '-c02' models.

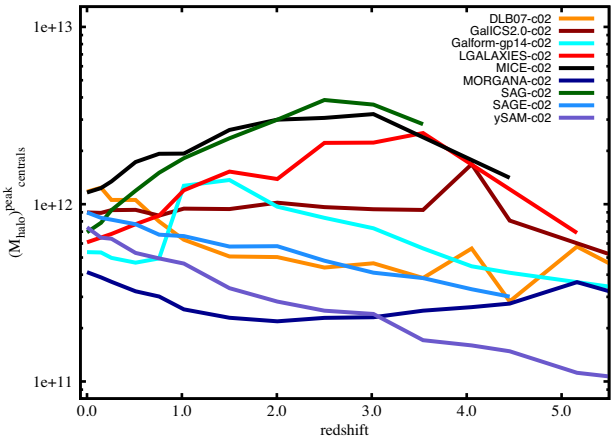

Figure 13. Redshift evolution of the peak position of the SHM relation for central galaxies in the '-c02' models.

hence the same halo masses. Therefore, all differences seen here in the SHM can purely be ascribed to the variations in the modelling of the stellar component of galaxies.

6.1 SHM relation

We start with inspecting the SHM relation at redshift $z = 0$ for both the '-c01' and '-c02' data sets. The results (for central galaxies) can be viewed in Fig. 11 alongside the best-fitting model of Rodríguez-Puebla et al. (2017).[6] Even though that model encapsulates data for both central and satellite galaxies, we only compare against centrals – as suggested by Rodriguez-Puebla (private communication, but see also Rodríguez-Puebla, Drory & Avila-Reese 2012). To generate this plot, the data have been binned logarithmically in halo mass for the x-axis and the y-axis shows the median $M_*/M_{\rm halo}$ of all central galaxies in that bin. For most of the models, the additional constraints of the '-c02' calibration lead to no appreciable difference. The SHM is essentially determined purely by the SMF, with the difference that here it only applies to central galaxies. We note that the SHM from the DLB07 model agrees significantly better with the observational results when the additional '-c02' constraints are used, while this relation remains more or less unaffected for the other models.

We have also performed the test where we added the stellar mass of all satellites to the stellar mass of the central galaxy as the halo mass of the central is 'inclusive' (i.e. contains all the subhalo masses). This gives rise to a clear effect at the high-mass end of the SHM: the model-to-model variation is marginally reduced for halo masses $M_{\rm halo} > 10^{13}$ M$_\odot$ and above the observational data by about a factor of two for $M_{\rm halo} > 10^{14}$ M$_\odot$. It indicates the importance of how to count stellar and halo mass in theoretical models when comparing to observations. Rather than showing the respective plots here we present them in Fig. C3 in Appendix C.

6.2 SHM peak value $(M_*/M_{\rm halo})^{\rm peak}$

In Fig. 12, we show the redshift evolution of the peak value $(M_*/M_{\rm halo})^{\rm peak}$ of the SHM relation. The value is obtained by spline interpolation using four times as many bins as shown in Fig. 11, but smoothing the curve to reduce the noise; further, only bins with at least 50 galaxies are considered.[7] We observe a general trend for all models in the sense that the star formation efficiency declines with increasing redshift – as reported before by, for instance, Moster et al. (2013) and Hudson et al. (2015). However, there appears to be only little (if any) evolution for GALFORM and MICE. This agrees with the findings of Mitchell et al. (2016) who reported a very strong dependence of $(M_*/M_{\rm halo})^{\rm peak}$ on model parameters, especially for the GALFORM model also used in their work.

6.3 SHM peak position $(M_{\rm halo})^{\rm peak}$

The value of $M_{\rm halo}$ where the SHM relation peaks – referred to as $(M_{\rm halo})^{\rm peak}$ here – coincides with the knee of the SMF: it is the halo mass for which star formation is most effective and least influenced by either stellar or AGN feedback (Moster et al. 2010; Yang et al. 2012). Assuming a simple relation with the typical mass of collapsed objects $M_\star$[8] and its evolution within a hierarchical structure formation scenario we naïvely expect $(M_{\rm halo})^{\rm peak}$ to drop with redshift. While a range of models (SAGE, GALFORM, YSAM, and DLB07) show such a trend, at least marginally, SAG, LGALAXIES, and MICE actually have a rising $(M_{\rm halo})^{\rm peak}$ value until redshift $z \sim 3$ – noting that these two SAM models are the ones applying an automated calibration procedure and MICE is the HOD model. The remaining models MORGANA and GALICS 2.0 favour no evolution. As outlined before, there is no clear consensus yet in the literature as to whether this value is evolving (Behroozi et al. 2013b; Moster et al. 2013; Leauthaud et al. 2012) or not (Hudson et al. 2015; McCracken et al. 2015). It is clear from our analysis that this quantity is model-dependent.

6.4 SHM relation scatter

The relation between stellar and halo mass is related to the star formation efficiency, and the evolution of galaxies is thought to be dominated by the growth of halo mass (White & Rees 1978). But the correlation between M_* and $M_{\rm halo}$ is not tight – as suggested by observations and models (both SAMs and hydrodynamical simulations, as shown in Guo et al. 2015). That means that the

[6] The data (including the error estimates for the halo masses) of Rodríguez-Puebla et al. (2017) are for redshift $z = 0.0$ and has been kindly provided by Aldo Rodriguez-Puebla.

[7] The actual curves can be viewed in Fig. C1.

[8] Usually defined as the mass of a 1σ peak in the density field at a given redshift, and not to be confused with M_*.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

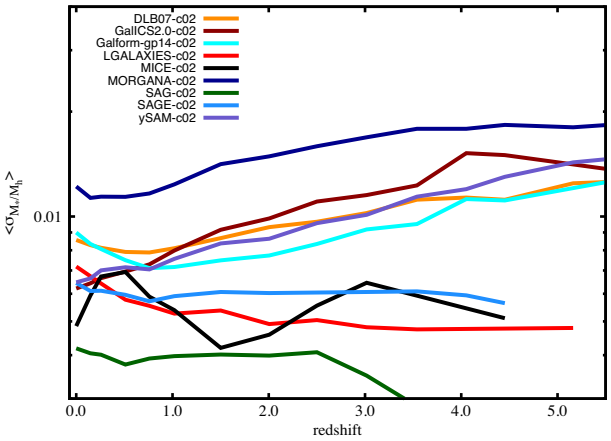

Figure 14. Redshift evolution of the mean of the scatter of the SHM relation for central galaxies in the '-c02' models.

halo mass alone is not sufficient to predict the stellar mass of the galaxy residing in it. The situation is actually far more complex for satellite galaxies whose halo has lost dark matter due to tidal stripping or are already completely disrupted leaving us with an orphan galaxy. Because of this, we will again restrict our analysis to central galaxies.

Putting aside the origin of the scatter in the SHM relation, we acknowledge that this scatter could be both halo mass and redshift dependent – as shown and investigated in great detail in Matthee et al. (2017). For that reason, it might be best to investigate the evolution of the scatter in appropriately chosen halo mass bins. However, for the work presented here, we refrain from it and simply define the mean scatter $\langle\sigma_{M_*/M_{\rm halo}}\rangle$ as the arithmetic mean of the scatter in each of the bins used for Fig. 11; and that scatter corresponds to half of the 25–75 percentiles of the distribution of $M_*/M_{\rm halo}$ values in that bin. For the relation of the scatter with halo mass – at least for redshift $z = 0$ – we refer the reader to Fig. C2 where the medians and 25–75 percentiles are shown for each model. Again, only bins with at least 50 galaxies in it will be considered.

We show the redshift evolution of $\langle\sigma_{M_*/M_{\rm halo}}\rangle$ in Fig. 14. We find that the evolution – if any – is very mild, but the normalization is highly model dependent. This is agreement with the findings of, for instance, Guo et al. (2015) where the scatter in the SHM relation and its evolution was also investigated for GALFORM and LGALAXIES: the larger dispersion seen for the former model comes from treating feedback differently in discs and bulges as opposed to only depending of the halo mass for the latter model.

7 CONCLUSIONS

This work continues the efforts presented in Knebe et al. (2015) of comparing a plethora of galaxy formation models applied to the same cosmological simulation. Here, we have compared nine galaxy formation models run on a larger simulation than previously, with a box of comoving width $125h^{-1}$ Mpc, and with a dark-matter particle mass of $1.24 \times 10^9 h^{-1}$ M$_\odot$. We also use the same halo catalogues (as identified with ROCKSTAR, Behroozi et al. 2013a) and merger trees (generated with CONSISTENTTREES, Behroozi et al. 2011). In this work, more unifying assumptions have also been made and models have been calibrated to a common set of observational data. However, some 'memory' of any previous calibration which served as the starting point for this work is retained; this is especially true for the models that are tuned manually. This may particularly affect the 'prescriptions' and 'predictions' of the models. All the data used here are publicly available.[9]

The nine models summarized in Table 1 have been calibrated in two ways: (i) just to the SMF at $z = 0$ and (ii) to the SMF, the SFRF, the cold gas mass, and the BHBM at $z \sim 0$ together with the SMF at higher redshift $z = 2$, i.e. the 'CARNage calibration' set (described in Section 3.3).

When calibrating all our models just to the SMF at $z = 0$ – as presented in Section 4 – the scatter is significantly reduced compared to models run with parameters fixed by other data sets. This indicates that the main conclusion of our earlier paper (Knebe et al. 2015), i.e. that the scatter was driven by the lack of recalibration, was correct. We reconfirm that galaxy formation models need to be recalibrated to the specific simulation, halo finder and merger tree being considered and, in general, cannot simply be rerun using parameters obtained from a different underlying simulation. At the high-mass end, where there are few objects, some models overproduce the number of massive galaxies. When calibrating simultaneously to all five constraints – as studied in Section 5 – the scatter in the SMF at $z = 0$ naturally grows. Even so, the scatter is still less than that obtained when using models with standard parameters tuned to other simulation data sets.

The observed shape of the SFRF at $z = 0$ is reproduced by models calibrated just to the SMF at $z = 0$. When this SFRF is added as a constraint, the scatter is reduced. This same trend is also seen for the cold gas mass. The observed BHBM relation is reproduced by the models even before this is used as part of the calibration, but several codes included this in previous incarnations of the models and hence the starting parameters used for the calibrations presented here introduce a 'memory' of this into the catalogues.

The most difficult constraint to match proved to be the SMF at $z = 2$. Most models struggle to obtain an acceptable fit and little improvement is seen whether this constraint is included or not. Even though this is a well-known problem (Fontanot et al. 2009; Weinmann et al. 2012; Mitchell et al. 2014; Henriques et al. 2014; Somerville & Davé 2015; Hirschmann et al. 2016), we will study this phenomenon in more detail for the models presented here in a companion paper (paper II, Asquith et al., in preparation). But note that the models that best respond to the addition of the high-redshift observational data during the calibration are the two that apply an automated tuning procedure (i.e. LGALAXIES and SAG) and the HOD model MICE.

Viewing, for instance, Fig. 3 (CGF for '-c01') or 10 (cSFRD for '-c01' and '-c02') that show a noticeable model-to-model variation for properties not used during the calibration, one might be inclined to question the predictive power of galaxy formation models. But before jumping to such conclusions, one needs to remember that these models are 'tools' that help to explore and eventually understand galaxy formation. And, as mentioned before, calibration itself is a tool too, and not a goal. Whenever a certain property is not matched satisfactorily, the model eventually allows for deeper insight to be gained regarding the driving astrophysics of said property.

In one way or another, all SAM and HOD models capture our present-day knowledge about galaxy formation, which is certainly not yet complete. These techniques are sufficiently rapid to be able to explore different physical prescriptions and change model parameters to probe the poorly understood aspects of galaxy formation. This is actually where the strength of them lies: while the other

[9] The resulting galaxy catalogues (ca. 40 GB of data) are stored on a data server to which access will be granted upon request to the leading author.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

approach of studying galaxy formation by means of hydrodynamical simulation is also based upon similar assumptions – at least regarding star formation and feedback – it is considerably slower than SAM and HOD models and hence parameter-space exploration is rather prohibitive for it. Furthermore, theoretical models might also focus on different aspects of galaxies; while one model may aim at providing reasonable gas fractions, another model may have its strength in reproducing the SFRF – with or without calibrating to it. And the model-to-model variation eventually is a reflection of model design and implementation of the actual physical phenomena into these tools.

We have also studied the SHM and its evolution with redshift in Section 6. The SHM relation represents the conversion efficiency from baryonic matter into stars within dark haloes and hence gives great insight into galaxy formation. However, there is still controversy as to whether the SHM relation evolves with redshift or remains constant. We know from observations (and models) that the SFR peaks at about redshift $z = 2$–3. But how exactly does this relate to the evolution of haloes? Focusing only on central galaxies, we find that for the majority of the models used here there is a clear trend for the position and value of the peak in the SHM relation to decline with redshift, albeit with prominent variations in the normalization of this evolutionary trend. The situation is less clear for the scatter in the SHM relation: some models predict it to be marginally declining, whereas other favour an increase.

In summary, we have presented a range of galaxy formation models calibrated to the same set of observations. These observations were chosen to be as complementary as possible and designed to test different aspects of the galaxy formation process. This choice places the various processes that govern galaxy formation into tension and the general success of the models demonstrates their robustness. The data from these galaxy formation models are available to the community. This paper provides a solid ground upon which build future explorations of the physical processes that govern the evolution of galaxies.

ACKNOWLEDGEMENTS

The authors would like to express special thanks to the Carnegie Observatories for their support and hospitality during the workshop ‘Cosmic CARNage’ where all the calibration issues were discussed and the roadmap laid out for the work presented here. This work further builds up upon the results obtained during the three week workshop ‘nIFTy Cosmology’ hosted by the Instituto de Fisica Teorica (IFT-UAM/CSIC in Madrid) and we also like to thank it for its hospitality and support, via the Centro de Excelencia Severo Ochoa Program under grant number SEV-2012-0249. We further acknowledge the financial support of the 2014 University of Western Australia Research Collaboration Award for ‘Fast Approximate Synthetic Universes for the SKA’, the Australian Research Council (ARC) Centre of Excellence for All Sky Astrophysics grant number CE110001020, and the two ARC Discovery Projects DP130100117 and DP140100198. We also recognize support from the Universidad Autonoma de Madrid for the infrastructure of the ‘nIFTy Cosmology’ workshop.

We like to express our special thanks to Rachel Somerville, Gabriella De Lucia, and Pierluigi Monaco for stimulating discussions; they all provided highly valuable comments on draft versions of this paper that helped to substantially improve it and bring it to its present shape. We further thank the anonymous referee whose comments and suggestions helped to improve the paper.

AK is supported by the *Ministerio de Economía y Competitividad* and the *Fondo Europeo de Desarrollo Regional* (MINECO/FEDER, UE) in Spain through grant AYA2015-63810-P. He also acknowledges support from the ARC grant DP140100198. FJC acknowledges support from the Spanish Ministerio de Economía y Competitividad project ESP2015-66861. SAC acknowledges grants from Consejo Nacional de Investigaciones Científicas y Técnicas (PIP-112-201301-00387), Agencia Nacional de Promoción Científica y Tecnológica (PICT-2013-0317), and Universidad Nacional de La Plata (UNLP 11-G124), Argentina. DJC acknowledges receipt of a QEII Fellowship from the Australian Government. PJE is supported by the SSimPL programme and the Sydney Institute for Astronomy, DP130100117. FF acknowledges financial contribution from the grants PRIN INAF 2010 ‘From the dawn of galaxy formation’ and PRIN INAF 2014 ‘Glittering kaleidoscopes in the sky: the multifaceted nature and role of Galaxy Clusters’ (1.05.01.94.02). VGP acknowledges support from a European Research Council Starting Grant (DEGAS-259586) and from the University of Portsmouth’s Dennis Sciama Fellowship. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the Science and Technology Funding Council (STFC) DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. The work of BH was supported by Advanced Grant 246797 GALFORMOD from the European Research Council. MH acknowledges financial support from the European Research Council via an Advanced Grant under grant agreement number 321323 NEOGAL. PM has been supported by a FRA2012 grant of the University of Trieste, PRIN2010-2011 (J91J12000450001) from MIUR, and Consorzio per la Fisica di Trieste. NDP was supported by BASAL PFB-06 CATA, and Fondecyt 1150300. Part of the calculations presented here were run using the Geryon cluster at the Center for Astro-Engineering at U. Catolica, which received funding from QUIMAL 130008 and Fondequip AIC-57. CP acknowledges support of the ARC through Future Fellowship FT130100041 and Discovery Project DP140100198. WC and CP acknowledge support of ARC DP130100117. AP acknowledges support from beca FI and 2009-SGR-1398 from Generalitat de Catalunya, project AYA2012-39620 and AYA2015-71825 from MICINN, and from a European Research Council Starting Grant (LENA-678282). RA is funded by the STFC through a studentship. RSS thanks the Downsbrough family for their generous support. PAT acknowledges support from the Science and Technology Facilities Council (grant number ST/L000652/1). SKY acknowledges support from the Korean National Research Foundation (NRF-2017R1A2A1A05001116). This study was performed under the umbrella of the joint collaboration between Yonsei University Observatory and the Korean Astronomy and Space Science Institute. The supercomputing time for the numerical simulations was kindly provided by KISTI (KSC-2014-G2-003).

The authors contributed to this paper in the following ways: AK, FRP, VGP, and PAT formed the core team with AK, FRP, and VGP writing the paper. AK and FRP organized week 2 of the ‘nIFTy Cosmology’ workshop where this work was initiated. AB organized the follow-up workshop ‘Cosmic CARNage’ where all the discussions about the common calibration took place and out of which this paper emerged. JO supplied the simulation and halo catalogue for the work presented here. The following authors performed the SAM or HOD modelling using their codes, in particular FJC, AC, SAC, DC,

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

ARHS, FF, VGP, BH, JL, PM, CVM, and SKY actively ran their models with the assistance of JH and CS. WC, DC, PJE, CP, and JO assisted with the analysis and data format issues. All authors proof read and commented on the paper.

This research has made use of NASA's Astrophysics Data System and the arXiv preprint server.

REFERENCES

Avila S. et al., 2014, MNRAS, 441, 3488
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Baldry I. K. et al., 2012, MNRAS, 421, 621
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2011, ApJ, 763, 18
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Conroy C., 2013b, ApJ, 770, 57
Benson A. J., 2014, MNRAS, 444, 2599
Benson A. J., Borgani S., De Lucia G., Boylan-Kolchin M., Monaco P., 2012, MNRAS, 419, 3590
Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014, A&A, 564, A66
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006a, MNRAS, 370, 645
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006b, MNRAS, 370, 645
Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, MNRAS, 407, 2017
Carretero J., Castander F. J., Gaztañaga E., Crocce M., Fosalba P., 2015, MNRAS, 447, 646
Cattaneo A. et al., 2017, MNRAS, 471, 1401
Cheung E. et al., 2016, Nature, 533, 504
Cora S. A. et al., 2018, preprint (arXiv:1801.03883)
Croton D. J. et al., 2006, MNRAS, 365, 11
Croton D. J. et al., 2016, ApJS, 222, 22
Davé R., 2009, in Jogee S., Marinova I., Hao L., Blanc G. A., eds, ASP Conf. Ser. Vol. 419, Galaxy Evolution: Emerging Insights and Future Challenges. Astron. Soc. Pac., San Francisco, p. 347
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Domínguez Sánchez H. et al., 2011, MNRAS, 417, 900
Fontanot F., De Lucia G., Monaco P., Somerville R. S., Santini P., 2009, MNRAS, 397, 1776
Frenk C. S., White S. D. M., 2012, Ann. Phys., 524, 507
Gargiulo I. D. et al., 2015, MNRAS, 446, 3820
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264
Gruppioni C. et al., 2015, MNRAS, 451, 3419
Guo Q. et al., 2011, MNRAS, 413, 101
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351
Guo Q. et al., 2015, MNRAS, 461, 3457
Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009a, MNRAS, 396, 535
Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009b, MNRAS, 396, 535
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
Henriques B., White S., Thomas P., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2014, MNRAS, 451, 2663
Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760
Hudson M. J. et al., 2015, MNRAS, 447, 298
Ilbert O. et al., 2013, A&A, 556, A55
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Knebe A. et al., 2015, MNRAS, 451, 4029
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Lagos C. d. P. et al., 2016, MNRAS, 459, 2632
Lara-López M. A. et al., 2010, A&A, 521, L53
Leauthaud A. et al., 2012, ApJ, 744, 159
Lee J., Yi S. K., 2013, ApJ, 766, 38
Lee J. et al., 2014, MNRAS, 445, 4197
Li C., White S. D. M., 2009, MNRAS, 398, 2177
Lilly S. J., Le Fevre O., Hammer F., Crampton D., 1996, ApJ, 460, L1
Lu Y. et al., 2014, ApJ, 795, 123
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388
Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115
Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, MNRAS, 465, 2381
McConnell N. J., Ma C.-P., 2013, ApJ, 764, 184
McCracken H. J. et al., 2015, MNRAS, 449, 901
Mitchell P. D., Lacey C. G., Cole S., Baugh C. M., 2014, MNRAS, 444, 2637
Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2016, MNRAS, 456, 1459
Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189
Mortlock A. et al., 2015, MNRAS, 447, 2
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Mutch S. J., Poole G. B., Croton D. J., 2013, MNRAS, 428, 2001
Muzzin A. et al., 2013, ApJ, 777, 18
Nagamine K., Cen R., Hernquist L., Ostriker J. P., Springel V., 2004, ApJ, 610, 45
Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, MNRAS, 378, 1550
Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, ApJ, 786, 54
Planck Collaboration et al., 2014, A&A, 571, A16
Pujol A. et al., 2017, MNRAS, 469, 749
Rodrigues L. F. S., Vernon I., Bower R. G., 2017, MNRAS, 466, 2418
Rodríguez-Puebla A., Drory N., Avila-Reese V., 2012, ApJ, 756, 2
Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651
Ruiz A. N. et al., 2015, ApJ, 801, 139
Salim S., Lee J. C., Davé R., Dickinson M., 2015, ApJ, 808, 25
Silk J., Mamon G. A., 2012, Res. Astron. Astrophys., 12, 917
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504
Springel V., 2005, MNRAS, 364, 1105
Tomczak A. R. et al., 2014, ApJ, 783, 85
van Daalen M. P., Henriques B. M. B., Angulo R. E., White S. D. M., 2015, MNRAS, 458, 934
Wang L., De Lucia G., Weinmann S. M., 2013, MNRAS, 431, 600
Weinmann S. M., Pasquali A., Oppenheimer B. D., Finlator K., Mendel J. T., Crain R. A., Macciò A. V., 2012, MNRAS, 426, 2797
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
White C. E., Somerville R. S., Ferguson H. C., 2015, ApJ, 799, 201
Wilkins S. M., Trentham N., Hopkins A. M., 2008, MNRAS, 385, 687
Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752, 41

APPENDIX A: THE OBSERVATIONAL DATA

Aiming at generating model galaxies with properties that can reasonably reproduce the observed statistical properties of observed galaxies, one of the objectives of the Cosmic CARNage workshop[10] – the successor of the nIFTy Cosmology workshop where the whole comparison was initiated – was to discuss appropriate observations

[10] http://users.obs.carnegiescience.edu/abenson/CARnage.html

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

to use for a common calibration of the participating galaxy formation models. We summarize the outcome of this debate here and list the chosen calibration data sets in Table 2.[11] The set of observations is designed to constrain parameters from a wide range of modelled physical processes, yet is observationally well established. The decision was to use the SMF (both at $z = 0$ and 2), the SFRF, the BHBM, and CGMF. This set of observational constraints probes several different aspects of the galaxy formation models that are all inter-related yet nevertheless sufficiently independent of each other.

However, we also like to state that calibration is a 'tool' – not a 'goal' – for galaxy formation models. What we present here as a data set is tailored to be useful for the purpose of this project, i.e. model comparison.

A1 Stellar mass function

The literature contains a great number of (local) measurements of the galaxy SMF deduced using empirically determined mass-to-light ratios. Stellar masses determined this way therefore rely on several implicit assumptions regarding the stellar IMF, star formation histories, and the integrated effects of dust attenuation. Hence, these estimates can suffer from large systematic uncertainties.

It was a matter of some debate amongst workshop participants whether it was better to compare the models to SMFs or to luminosity functions. In principle, the latter is more straightforward as the models have known star formation histories and no conversion is required for the observational data. However, the reliance on an accurate dust model outweigh this advantage. Hence, we decided to compare a quantity directly measurable in the models: the mass in stars, and treat the differences between the observational predictions as an estimate of the systematic error.

The observed SMF used here is a compilation of the data presented in Baldry et al. (2012), Li & White (2009), and Baldry et al. (2008) for redshift $z = 0$ and Tomczak et al. (2014),Muzzin et al. (2013), Ilbert et al. (2013), and Domínguez Sánchez et al. (2011) for $z = 2$. The different data sets at each redshift are formally incompatible with each other within the error bars, suggesting that there are systematic errors between them. For that reason, the maximum and minimum observational estimates (taking into account the error bars) are used as a measure of the systematic uncertainty between observations; the precise details of the procedure can be found in appendix C of Henriques et al. (2013).

A2 Star formation rate function

We use the SFRF as presented in table 1 of Gruppioni et al. (2015) for the redshift interval $z \in [0.0, 0.3]$. These data come from a flux-limited sample of galaxies observed with the *Herschel* satellite giving the total (IR+UV) instantaneous SFRs. These data were compared against model galaxies at redshift $z = 0.15$.

A3 Cold gas fractions

For the mass fraction of HI+H_2, the decision was to use the Boselli et al. (2014) data, which is based on a volume-limited sample, within the range $\log_{10}M_*/\ M_\odot \in [9.15, 10.52]$ in stellar mass. For the data used here, we combined the information for the two methods used for the X factors; see table 4 in Boselli et al. (2014). We further agreed not to use a cut for separating active from passive galaxies during the calibration – given that such a cut can be model dependant. Therefore, while the Peeples et al. (2014) data were also discussed, it eventually was not adopted as it only contains star-forming (active) galaxies.

[11] While the last column in Table 2 gives the references from which the data have been obtained, we also provide the link to a page with the actual observational data files used throughout this study: http://popia.ft.uam.es/public/CARNageSet.zip

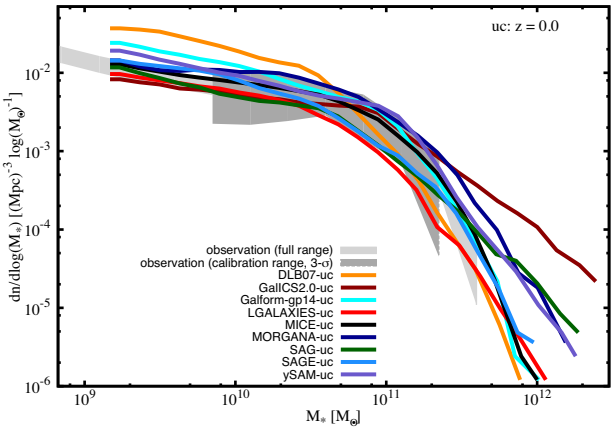

Figure B1. Same as Fig. 1, but for galaxy catalogues not recalibrated.

A4 Black hole–bulge mass relation

The BHBM used for calibration is a compilation of both the data presented in McConnell & Ma (2013) and Kormendy & Ho (2013).

APPENDIX B: UNCALIBRATED CATALOGUES

While we only studied commonly calibrated models in the main part of this paper, we show here in Fig. B1 the SMF at redshift $z = 0$ for each model when used without any recalibration and how it compares to the CARNage calibration SMF. This SMF could be compared to the upper panel of fig. 2 in Knebe et al. (2015). There is a difference though: in Knebe et al. (2015) the simulation featured a much smaller volume of $(62.5^3)h^{-1}$ Mpc^3, whereas we are using here $(125^3)h^{-1}$ Mpc^3. Further, merger trees in Knebe et al. (2015) were constructed with MERGERTREE whereas we are here using CONSISTENTREES. As has been shown in Avila et al. (2014), this will have an impact on the quality of the trees and Lee et al. (2014) further show that (and how) this impacts upon SAMs.

APPENDIX C: STELLAR-TO-HALO MASS RELATION

C1 Redshift evolution

In Fig. 11, we have presented the SHM relations for all '-c02' models at redshift $z = 0$, but later on studied the redshift evolution in Figs 12–14. Here, we now show in Fig. C1 for each model individually the SHM relations at all considered redshifts that directly entered into the calculation of the peak value and position.

C2 Individual variance

In Fig. 14, we have shown the mean scatter of the SHM relation defined as the 25–75 percentiles of the distribution in each $M_{\rm halo}$ bin. In Fig. C2, we now give an example of the scatter of each model at redshift $z = 0$.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

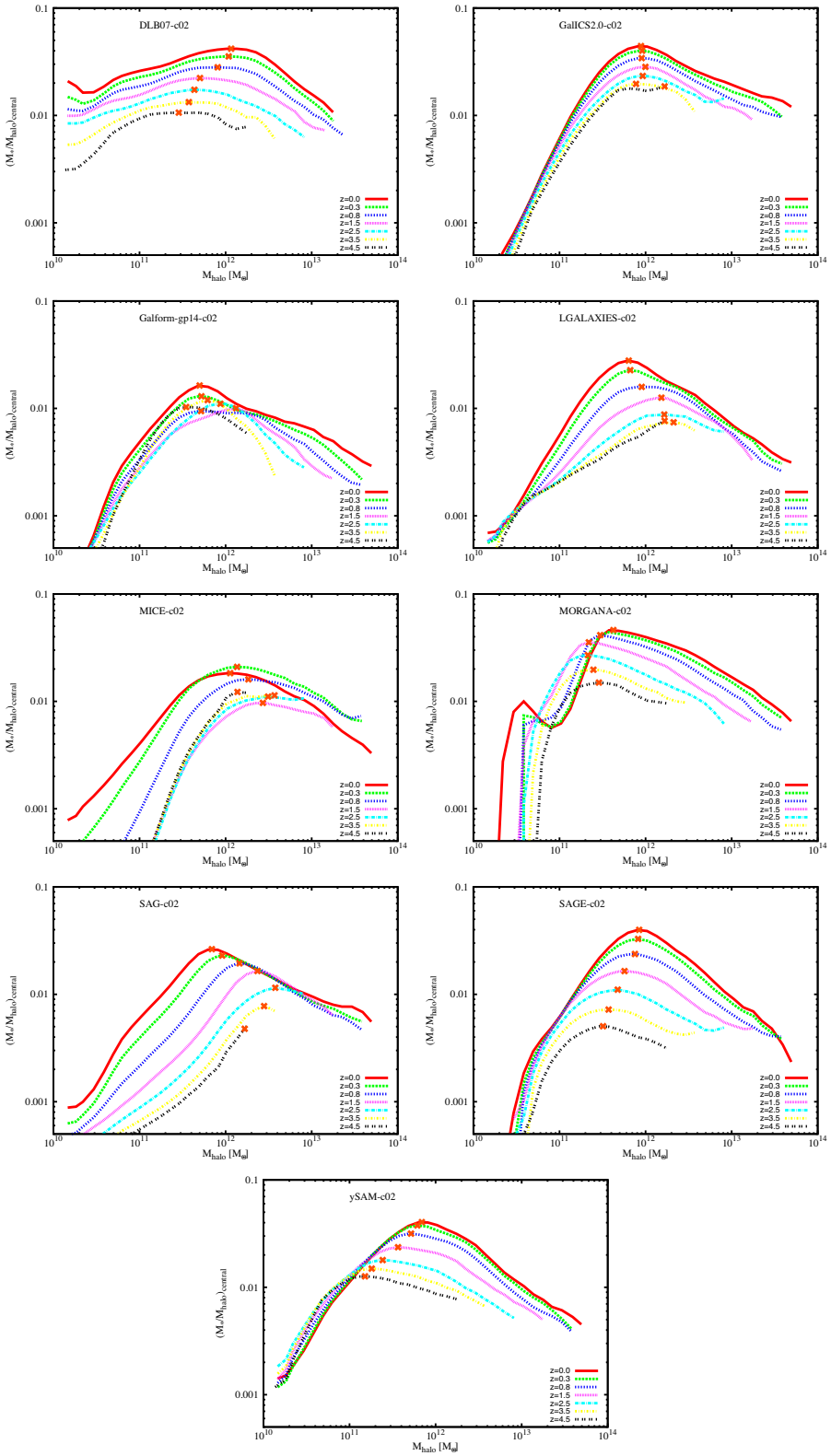

Figure C1. Evolution of the SHM relation for all '-c02' models. For clarity, only every second available redshift is shown. The grey cross marks the peak (position) used in Section 6.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

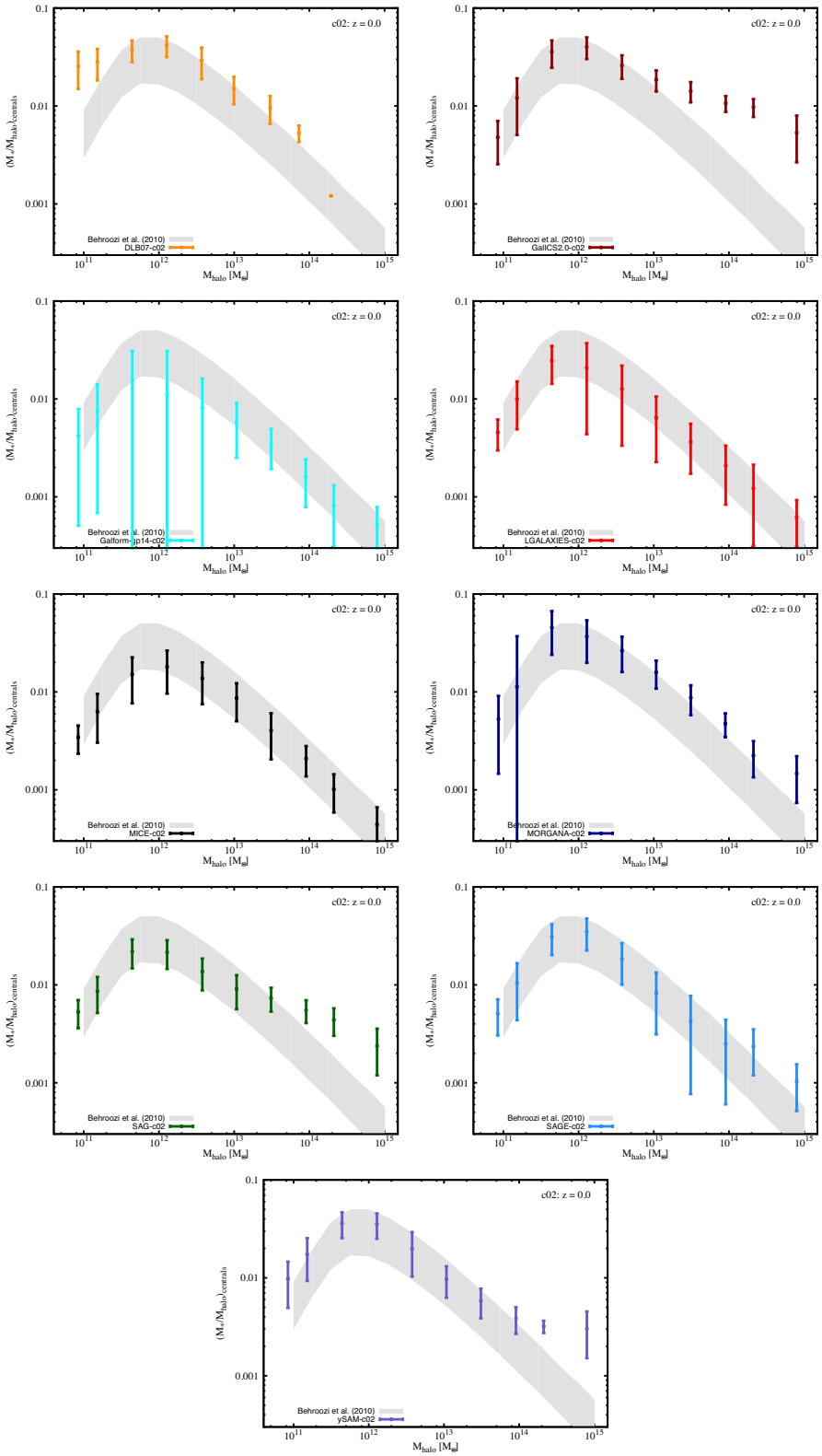

Figure C2. The SHM relation for all models at redshift $z = 0$ including error bars defined as 25–75 percentiles of the distribution in each bin.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

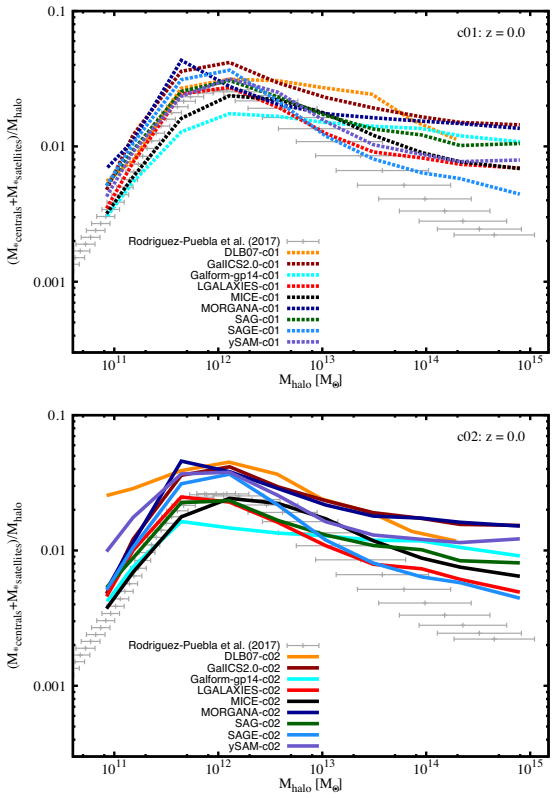

Figure C3. SHMas a function of halo mass for central galaxies only, but with the stellar mass of the satellite galaxies added to it. The values shown are medians in the respective bin.

C3 Addition of satellite stellar mass

As mentioned in Section 6.1, we have performed the test of adding the stellar mass of satellite galaxies to the stellar mass of the central when calculating the SHM ratio as the halo mass entering this ratio also contains the masses of the subhaloes. The resulting changes to the original Fig. 11 can be viewed here in Fig. C3 where the upper panel shows calibration '-c01' and the lower panel '-c02'. We notice a substantial effect, especially at the high-M_* end, i.e. for haloes that host a pronounced number of subhaloes and satellites, respectively. This test highlights the importance of how to count stellar and halo mass in theoretical models when comparing to observations, but it does not change our conclusions.

C4 Addition of cold gas mass

We have also performed the test of adding the cold gas mass to the stellar mass of the central galaxy when calculating the SHM ratio (which should then rather be called the baryon-to-halo mass relation, but we continue calling it SHM). The results can be viewed in Fig. C4 which shows the anologies to Figs 12–14. Even though some models do show rather distinct changes, the trends are nevertheless preserved and our conclusions not affected, respectively.

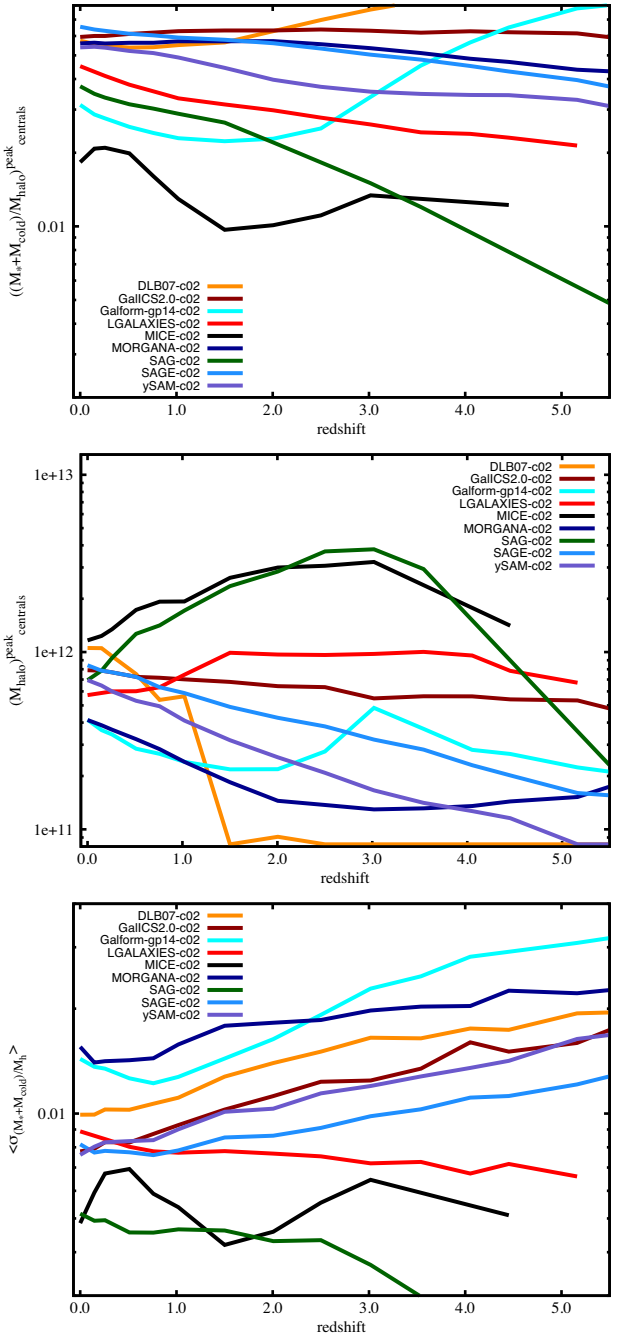

Figure C4. Redshift evolution of the peak value (upper panel), the peak position (middle panel), the mean of the scatter (lower panel) of the SHM relation for central galaxies when also adding the cold gas mass to the stellar mass of the central in the '-c02' models.

[1]*Departamento de Física Teórica, Módulo 15, Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*
[2]*Centro de Investigación Avanzada en Física Fundamental (CIAFF), Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*
[3]*School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK*
[4]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018

[5]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[6]*Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*
[7]*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[8]*Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France*
[9]*Institut de Ciències de lÉspai (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[10]*Institut dÉstudis Espacials de Catalunya (IEEC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[11]*Port d'Informació Científica (PIC) Edifici D, Universitat Autònoma de Barcelona (UAB), E-08193 Bellaterra (Barcelona), Spain*
[12]*GEPI, Observatoire de Paris, CNRS, 61, Avenue de l'Observatoire, F-75014 Paris, France*
[13]*Institut d'Astrophysique de Paris, 98bis Boulevard Arago, F-75014 Paris, France*
[14]*Instituto de Astrofísica de La Plata (CCT La Plata, CONICET, UNLP), Paseo del Bosque s/n, B1900FWA, La Plata, Argentina*
[15]*Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque s/n, B1900FWA, La Plata, Argentina*
[16]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia*
[17]*South African Astronomical Observatory, PO Box 9, Observatory, Cape Town 7935, South Africa*
[18]*Department of Physics and Astronomy, University of the Western Cape, Cape Town 7535, South Africa*
[19]*Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[20]*International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia*
[21]*Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[22]*INAF – Astronomical Observatory of Trieste, via Tiepolo 11, I-34143 Trieste, Italy*
[23]*Institute for Astronomy, ETH Zurich, CH-8093 Zurich, Switzerland*
[24]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85741 Garching b. München, Germany*
[25]*Korea Institute for Advanced Study, 85 Hoegi-ro, Dongdaemun-gu, Seoul 02455, Republic of Korea*
[26]*Institut dÁstrophysique de Paris (CNRS and Sorbonne Université), 98 bis Bd Arago, F-75014 Paris, France*
[27]*Instituto de Astrofisica, Universidad Catolica de Chile, Santiago, Chile*
[28]*Centro de Astro-Ingenieria, Universidad Catolica de Chile, Santiago, Chile*
[29]*Institut de Ciències de lÉspai, IEEC-CSIC, Campus UAB, E-08193 Bellaterra, Barcelona, Spain*
[30]*DEDIP/DAP, IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[31]*Université Paris Diderot, AIM, Sorbonne Paris Cité, CEA, CNRS, F-91191 Gif-sur-Yvette, France*
[32]*Instituto de Astronomía Teórica y Experimental (CONICET-UNC), Laprida 854, X5000BGR, Córdoba, Argentina*
[33]*Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina*
[34]*Observatoire de Paris, GEPI, CNRS and PSL Research University, 61 Avenue de l'observatoire, F-75014 Paris, France*
[35]*Université Paris Diderot, Sorbonne Paris Cité, F-75013 Paris, France*
[36]*Department of Astronomy and Yonsei University Observatory, Yonsei University, Seoul 03722, Republic of Korea*

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article-abstract/475/3/2936/4768290
by inaf user
on 09 April 2018