Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **542,** 2808–2829 (2025) https://doi.org/10.1093/mnras/staf1389
Advance Access publication 2025 August 23

Simultaneously modelling dusty star-forming galaxies and massive quiescents: a calibration framework for galaxy formation models

Pablo Araya-Araya [1]★ Rachel K. Cochrane [2,3,4]★ Christopher C. Hayward [5,6] Laerte Sodré Jr [1] Robert M. Yates [7] Marcel P. van Daalen [8] and Marcelo C. Vicentin[1]

[1]*Departamento de Astronomia, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão 1226, Cidade Universitária, 05508-900, São Paulo, SP, Brazil*
[2]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[3]*Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA*
[4]*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA*
[5]*Eureka Scientific, Inc., 2452 Delmer Street, Suite 100, Oakland, CA 94602, USA*
[6]*Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[7]*Centre for Astrophysics Research, University of Hertfordshire, Hatfield, AL10 9AB, UK*
[8]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

Accepted 2025 August 13. Received 2025 July 25; in original form 2025 April 18

ABSTRACT

Galaxy formation models, particularly semi-analytic models (SAMs), rely on differential equations with free parameters to describe the physical mechanisms governing galaxy formation and evolution. Traditionally, most SAMs calibrate these parameters manually to match observational data. However, this approach fails to fully explore the multidimensional parameter space, resulting in limited robustness and inconsistency with some observations. In contrast, the `L-Galaxies` SAM features a unique Markov Chain Monte Carlo mode, enabling robust model calibration. Using this functionality, we address a long-standing tension in galaxy formation models: simultaneously reproducing the number densities of dusty star-forming galaxies (DSFGs) and high-redshift massive quiescent galaxies (MQs). We test nine combinations of observational constraints – including stellar mass functions, quiescent fractions, neutral hydrogen mass functions, and DSFG number densities – across different redshifts. We then analyse the resulting galaxy property predictions and discuss the underlying physical mechanisms. Our results identify a model that reasonably matches the number density of DSFGs while remaining consistent with observationally derived lower limits on the number density of high-redshift MQs, though limitations remain. This model requires high star formation efficiencies in mergers and a null dependency of supermassive black hole (SMBH) cold gas accretion on halo mass, facilitating rapid stellar mass and SMBH growth. Additionally, our findings highlight the importance of robust calibration procedures to address the significant degeneracies inherent to multidimensional galaxy formation models.

Key words: methods: numerical – galaxies: evolution – galaxies: formation – galaxies: high-redshift.

1 INTRODUCTION

Dusty star-forming galaxies (DSFGs; also known as submillimetre galaxies or SMGs) have gained significant attention since their discovery in the late 1990s (Smail, Ivison & Blain 1997; Barger et al. 1998; Hughes et al. 1998; Eales et al. 1999). Intrinsically, DSFGs are highly luminous ($L_{\rm IR} \gtrsim 10^{11}\,{\rm L}_{\odot}$), and due to the negative k-correction, they are relatively easy to detect even at high redshifts (e.g. at $z \gtrsim 4$; Cooper et al. 2022; Manning et al. 2022; Long et al. 2024). DSFGs were first identified in single-dish surveys, where large beam sizes hampered individual source localization and cross-matching to multiwavelength data (e.g. see the extensive work that led to a secure redshift for the SMG HDF850.1; Hughes et al. 1998; Downes et al. 1999; Richards 1999; Dunlop et al. 2004; Wagg et al. 2007; Cowie et al. 2009; Walter et al. 2012; Neri et al. 2014; Herard-Demanche et al. 2025).

The high angular resolution of the Atacama Large Millimeter/submillimetre Array (ALMA) has enabled the localization and detailed characterization of DSFGs across cosmic time (see Hodge & da Cunha 2020, for a review), leading to constraints on their redshift distribution, physical properties, and large-scale environments. Observational studies show that DSFGs are predominantly found at redshifts $z \sim 2$–3 (Chapman et al. 2005; Simpson et al. 2017; Dudzevičiūtė et al. 2020), approximately coinciding with the global peak of cosmic star formation activity (e.g. Madau & Dickinson 2014; Cochrane et al. 2023b). Dust-obscured star formation comprises nearly half of the total cosmic star formation rate density (CSFRD) at these epochs (Dunlop et al. 2017; Michałowski et al. 2017; Smith et al. 2017; Zavala et al. 2021). Despite this, DSFGs

★ E-mail: rcochra3@roe.ac.uk (RKC); paraya-araya@usp.br (PAA)

are relatively rare, with number densities of $N \sim 10^{-5}\,\mathrm{Mpc}^{-3}\,\mathrm{Gyr}^{-1}$ at the peak of their redshift distribution (Dudzevičiūtė et al. 2020). The brightest DSFGs are highly clustered (Blain et al. 2004; Chen et al. 2016; García-Vergara et al. 2020; Lim et al. 2020; Stach et al. 2021), and also serve as effective tracers of galaxy protoclusters (Chapman et al. 2001; Daddi et al. 2009; Dannerbauer et al. 2014; Casey 2016; Marrone et al. 2018; Miller et al. 2018; Oteo et al. 2018; Wang et al. 2021; Gouin et al. 2022; Calvi, Castignani & Dannerbauer 2023; Araya-Araya et al. 2024; Hill et al. 2025; Herard-Demanche et al. 2025). They have been proposed as potential progenitors of the massive elliptical galaxies found at the centres of present-day galaxy clusters (e.g. Toft et al. 2014). Their extreme properties – including high stellar masses ($M_\star \sim 10^{11}\,\mathrm{M}_\odot$) and intense dust-obscured star formation rates (SFRs; $\sim 10^2$–$10^3\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$) (e.g. Simpson et al. 2014; da Cunha et al. 2015; Dudzevičiūtė et al. 2020; Cochrane et al. 2021) – make DSFGs valuable laboratories for both observational and theoretical studies of galaxy evolution in extreme environments.

Historically, theoretical models have struggled to reproduce the DSFG population (see section 10 of Casey, Narayanan & Cooray 2014 for a review), particularly submillimetre (sub-mm) number counts (e.g. at 870 μm; Granato et al. 2000; Fontanot et al. 2007; Somerville et al. 2012; Cowley et al. 2019; Hayward et al. 2021). A potential solution was proposed by Baugh et al. (2005), who used the GALFORM semi-analytic model (SAM; Cole et al. 2000; Lacey et al. 2016) to show that a top-heavy.[1] Stellar initial mass function (IMF) in merger-induced starbursts could bring models into better agreement with observations. However, this was controversial, as SAMs (and galaxy formation models in general) include many free parameters, enabling alternative solutions without invoking IMF variations. For instance, Hayward et al. (2013) matched sub-mm number counts under a universal IMF, estimating fluxes from dust mass and SFR relations calibrated via radiative transfer (Hayward et al. 2011). Safarzadeh, Lu & Hayward (2017) and Lagos et al. (2019) similarly found agreement with observations using universal IMFs in the Lu et al. (2011, 2014) and SHARK SAMs, respectively.

Large-box cosmological hydrodynamical simulations show mixed success. Using radiative transfer on EAGLE galaxies (McAlpine et al. 2016; Trayford et al. 2017), McAlpine et al. (2019) broadly reproduced SMG redshift distributions, but underpredicted bright-end number counts by over an order of magnitude (Cowley et al. 2019). Using scaling relations, Hayward et al. (2021) found better number counts in Illustris (Genel et al. 2014; Vogelsberger et al. 2014), though IllustrisTNG (Marinacci et al. 2018; Springel et al. 2018) underperforms, likely due to lower dust masses and SFRs in high-mass galaxies (also yielding too-high redshift peaks; Kumar et al. 2025). The SIMBA simulation (Davé et al. 2019), post-processed with POWDERDAY (Narayanan et al. 2021), produces consistent number counts (Lovell et al. 2021), though its brightest sources are skewed to higher redshifts. Recently, Kumar et al. (2025) used the FLAMINGO simulation (Schaye et al. 2023) with updated scaling relations to match number counts and redshift distributions under the Hayward et al. (2013) calibration (but not the newer Lovell et al. 2021 one).

Reproducing the number density of massive quiescent galaxies (MQs; $M_\star \gtrsim 10^{10.5}\,\mathrm{M}_\odot$, sSFR $\lesssim 10^{-11}\,\mathrm{yr}^{-1}$) at high redshift is also a major challenge – especially in light of *James Webb Space Telescope* (*JWST*). Lagos et al. (2025) compared several models – GAEA, GALFORM, SHARK, SIMBA, IllustrisTNG, and EAGLE – and found all underpredict MQs by 0.3–> 1 dex compared to recent *JWST* results (Carnall et al. 2023; Valentino et al. 2023; Alberts et al. 2024; Nanayakkara et al. 2024). Similar tensions appear in SIMBA-C (Szpila et al. 2025) and L-Galaxies (Vani et al. 2025), though MAGNETICUM (Kimmig et al. 2025) matches MQs at high-z – but overpredicts them by an order of magnitude at low redshift (Lagos et al. 2025).

Overall, while some simulations reproduce sub-mm number counts, most perform worse at modelling MQs – particularly their number densities (though these remain observationally uncertain; e.g. Valentino et al. 2020, 2023). For instance, Lagos et al. (2018, 2019) found that SHARK v1.0 matched sub-mm counts but underpredicted MQs by ~ 1 dex, a shortcoming improved in v2.0 (Lagos et al. 2024). Illustris reproduces number counts (Hayward et al. 2021) but underpredicts quiescent galaxies (Merlin et al. 2019), as does SIMBA (Merlin et al. 2019; Lovell et al. 2021). Conversely, IllustrisTNG and EAGLE better reproduce MQs but underpredict sub-mm counts (Cowley et al. 2019; Hayward et al. 2021). These inconsistencies complicate our understanding of the formation and evolution of both population, particularly since DSFGs and MQs may be connected through evolutionary pathways (Daddi et al. 2010; Tacconi et al. 2010; Casey et al. 2014; Valentino et al. 2020; Hayward et al. 2021). Resolving this tension is therefore a key challenge for theoretical astrophysics.

As mentioned above, one of the main reasons why modifying the IMF to solve the sub-mm number counts tension remains controversial is the high-dimensional parameter space of galaxy formation models. For example, SAMs typically have more than 10 free parameters that are often manually tuned. This 'calibration' process does not fully explore the range of possible scenarios and their physical implications, potentially obscuring alternative solutions. Consequently, robust calibration techniques are essential to rigorously test how well galaxy formation models reproduce observations. However, performing a comprehensive calibration is computationally expensive, as it requires extensive parameter-space exploration. In practice, this is infeasible for large-volume hydrodynamical simulations.

Unlike most models, the L-Galaxies SAM has incorporated a systematic calibration framework since Henriques et al. (2013), using a Markov Chain Monte Carlo (MCMC) approach, known as the 'MCMC mode'. This feature makes L-Galaxies uniquely flexible[2] by enabling calibration against multiple observables, including the stellar mass function (SMF), luminosity function, and quiescent fraction across different redshifts. In this work, we use the MCMC mode of the Henriques et al. (2020) version of L-Galaxies to systematically explore solutions to the SMG–MQ tension. Specifically, we calibrate the model using different sets of observational constraints, incorporating, for the first time, the number density of SMGs as a direct constraint. We then compare the galaxy properties predicted by the best-fitting models, run on the Millennium simulation (Springel 2005), across different calibration data sets. Finally, we analyse the dominant physical mechanisms driving these differences and assess the level of degeneracy in our most promising model.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of the SAM used in this study. In Section 3, we describe our calibration framework, including the MCMC mode and the observational constraints used. We then present our model

[1] A $\mathrm{d}n/\mathrm{dlog}\,m$ constant is assumed in Baugh et al. (2005) while $\mathrm{d}n/\mathrm{dlog}\,m \propto m^{-1}$ is implemented by Lacey et al. (2016).

[2] The Lagos et al. (2024) version of SHARK also implemented a calibration method, but the final parameter choices were still refined via visual inspection.

predictions, physical interpretations, and an analysis of degeneracies in Section 4. In Section 5, we discuss our results. We conclude with a summary of our findings in Section 6.

Throughout this work, we adopt the Planck Collaboration XVI (2014) cosmology: $\sigma_8 = 0.829$, $H_0 = 67.3\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.685$, $\Omega_m = 0.315$, $\Omega_b = 0.0487$, $f_b = 0.155$, and $n = 0.96$, consistent with the cosmologically rescaled version of the Millennium simulation (Angulo & Hilbert 2015).

2 GALAXY FORMATION MODEL

In this work, we use the Henriques et al. (2020) version of the L-Galaxies SAM of galaxy formation. In this section, we briefly describe the principal aspects of this model.

The L-Galaxies SAM is optimized to run on the Millennium and Millennium-II *N*-body dark matter-only simulations (Springel 2005; Boylan-Kolchin et al. 2009). In practice, L-Galaxies runs on the merger trees created with the SUBFIND algorithm (Springel et al. 2001). Additionally, L-Galaxies performs a cosmology scaling (Angulo & White 2010), updating halo properties according to new cosmological parameters – in this case, the Planck Collaboration XVI (2014) cosmology. After the cosmology scaling, the Millennium simulation volume is $(713.6\,\mathrm{cMpc})^3$ with a dark matter particle mass resolution of $m_p = 1.43 \times 10^9\,\mathrm{M_\odot}$. Here, we only run L-Galaxies on the Millennium simulation.

The evolution of baryonic components is modelled by a set of differential equations that describe astrophysical processes. Initially, primordial gas begins to accrete on to sufficiently massive dark matter haloes. The infalling gas is first added to the hot gas reservoir and subsequently transitions to the cold gas reservoir through radiative cooling. This version of L-Galaxies follows the evolution of cold gas in concentric rings within galaxies. This gas is further separated into H I and H_2, with only the latter forming stars, either through a secular process (based on H_2 surface density) or merger-induced starbursts. Besides triggering star formation, mergers are the main mechanism in the model for growing supermassive black holes (SMBHs), where SMBH mass growth is linked to the energy released from active galactic nuclei (AGNs). AGN feedback is a crucial process regulating star formation in massive galaxies. Moreover, mergers also affect galaxy morphology, destroying discs and contributing to the growth or formation of the galaxy bulge.

Star formation and the evolution of the stellar component are related with various astrophysical processes in the galaxy evolution context. As stars reach their final stages, supernovae (SNe) and stellar winds release metals and energy into the interstellar medium (ISM) and circumgalactic medium (CGM). In this version of L-Galaxies, metal enrichment from asymptotic giant branch (AGB) stars, SNe-Ia, and SNe-II is considered. Coupled with these events, the release of energy plays a crucial role in regulating subsequent star formation, i.e. SN feedback. The SN feedback in L-Galaxies operates in two ways: (re)heating and ejecting gas. The former (a) heats the cold gas within galaxies, transferring some to the surrounding CGM, and/or (b) reheats the CGM, thereby delaying cooling. On the other hand, when the energy release is significant, a fraction of the hot gas is ejected and later reincorporated after some time.

Environmental effects, such as tidal stripping, disruption, and ram pressure stripping, are also included in L-Galaxies. These processes occur when a halo is accreted by a more massive one. Among their effects on galaxies, these processes can remove hot gas atmospheres, modify galaxy components, and disrupt small systems.

The Henriques et al. (2020) version of L-Galaxies has 19 free parameters in total (table 1 in Henriques et al. 2020), of which 15 were constrained using the MCMC mode of the model. Here, we follow the Henriques et al. (2020) configuration to calibrate the model, constraining the same 15 free parameters.

To date, five more recent modifications of L-Galaxies have been published (Ayromlou et al. 2021; Yates et al. 2021; Izquierdo-Villalba et al. 2022; Murphy, Yates & Mohamed 2022; Spinoso et al. 2023; Yates et al. 2024) since Henriques et al. (2020). However, these versions introduce new treatments for specific astrophysical processes while still using the Henriques et al. (2020) version as a base. Therefore, in this work, we choose to use the default Henriques et al. (2020) version of L-Galaxies.

3 CALIBRATION METHOD

3.1 The MCMC mode

First introduced by Henriques et al. (2009) and Henriques & Thomas (2010), the MCMC mode of L-Galaxies enables exploration of the model's free parameter space and its calibration against a set of observational data. Since the Henriques et al. (2013) version, the MCMC mode has operated on a representative subset of merger trees designed to approximate, as closely as possible, the predictions of observables – such as the SMF, red/passive fractions, and number densities – when compared to those derived from the full cosmological volume. This approach significantly accelerates the MCMC process, making it feasible for calibrating galaxy formation models. However, not all haloes in a given set of merger trees necessarily represent the overall predictions across all redshifts. Thus, a critical input for the MCMC mode is a carefully selected subsample of haloes within the representative subset of merger trees at a given redshift, which, when combined, reproduce the results of the full model. Each halo in this sample is assigned a weight reflecting the number of similar haloes in the entire simulation volume. In Section 3.2, we introduce a new method for selecting both a sample of merger trees and the subsample of haloes within those trees that effectively represent the predictions of the entire simulation.

The MCMC mode also requires the set of observables to constrain the model. The default L-Galaxies version already includes a large set of observables at different redshifts, both one-dimensional (such as SMFs, luminosity functions in different bands, and cold gas mass functions) and two-dimensional relations (such as black hole-bulge mass, stellar metallicity-stellar mass, and size-stellar mass relations). For instance, in Henriques et al. (2020), the model was calibrated using the SMFs and fraction of quiescent galaxies both at $z = 0$ and $z = 2.0$, and the neutral hydrogen mass function (HIMF) at $z = 0$ as observational constraints. These observables are compared with the predictions generated at each MCMC step (from the sample of haloes at a given redshift), and the likelihood is estimated. In Section 3.3, we describe the sets of observational constraints that we use in this work.

In practice, multiple chains are run in parallel, each beginning from an initial point randomly displaced by a value $\sigma_{\mathrm{initial}}$. The Metropolis–Hastings algorithm is then applied using a lognormal Gaussian proposal distribution with width σ.

3.2 Sample of merger trees

As discussed in Henriques et al. (2013), implementing an MCMC approach to calibrate galaxy formation models within cosmological

simulations remains a computational challenge. In fact, the implementation of a robust calibration is still not feasible for hydrodynamical simulations. This is primarily due to the need to track the evolution of millions of galaxies while repeatedly varying the model's free parameters. A key alternative, introduced in Henriques et al. (2009) and Henriques & Thomas (2010), involves using subvolumes of the simulation. While this approach improves efficiency, the subvolumes may not fully represent the predictions of the entire simulation. Specifically, probes of rare populations, such as the massive end of the SMF and the fraction of quiescent galaxies, can be under- or overpredicted in certain regions.

To address this, Henriques et al. (2013) developed a method for selecting samples of merger trees that accurately represent the predicted luminosity function of a fiducial model. This approach not only improves the representativeness of the sample but also significantly reduces computational time, as far fewer haloes and merger trees are needed to reproduce the predictions of the full volume. However, initial tests performed as part of this project showed that their method fails to adequately represent the number density of SMGs – a key observable for this study – primarily due to the rarity of these galaxies. This limitation might also apply to other non-conventional observables that probe rare populations. Therefore, the selection of merger tree samples should be based on the specific observational constraints required for the study.

Here, we introduce a new method for selecting samples of merger trees that can also be applied to generate representative samples for other observables used in the calibration of galaxy formation models. This method involves two main steps. The first step involves selecting a reasonable subsample of merger trees consistent with the observable predictions of the entire simulation at a given redshift. We choose $z = 2.8$ in this work, as our motivation is to test whether we can match the SMG density and the quiescent galaxy fraction at this epoch. In particular, our observables, which we describe in Section 3.3, are the SMF, the quiescent galaxy fraction as a function of stellar mass ($f_{\rm Q}$), the SMG number density ($n_{\rm SMG}$), and the HIMF. Since the default 2020 version of `L-Galaxies` critically underpredicts the submillimetre number counts even more than the Henriques et al. (2015) version, we use here a fiducial model that was found from preliminary recalibration results. This was obtained similarly to configuration 'all' (see Section 3.4), but with a lower number of MCMC steps ($\sim 2,000$). In general, the fiducial model predicts similar sub-mm number counts and fractions of quiescent galaxies as the Henriques et al. (2015) model and the Henriques et al. (2020) model, respectively. Compared to Henriques et al. (2020), the fiducial model implements a higher AGN feedback efficiency and star formation from merger-induced starbursts, but a lower efficiency in forming stars by secular process. Then, we construct a 2D grid (20×20 bins) of the virial mass–stellar mass relation for all central galaxies (galaxies containing the most-bound dark matter particle in each FOF halo group) at $z = 2.8$, as shown in Fig. 1.

We could also include other key galaxy properties in this selection process, such as SFR, local overdensity, or metallicity. However, increasing the number of dimensions also increases the number of cells required to select a representative sample of haloes. For this reason, we chose to consider only the halo and stellar masses of the central galaxy, which is sufficient to obtain a representative sample, as shown below. Notice that we set a $M_\star = 10^6\,{\rm M}_\odot/h$ lower limit in Fig. 1, which is below the stellar mass of well-modelled galaxies due to mass resolution, when run on `Millennium`. This ensures the sample of haloes we select represents the predictions of the entire simulation volume. The upper limits are set as the maximum stellar and halo masses.

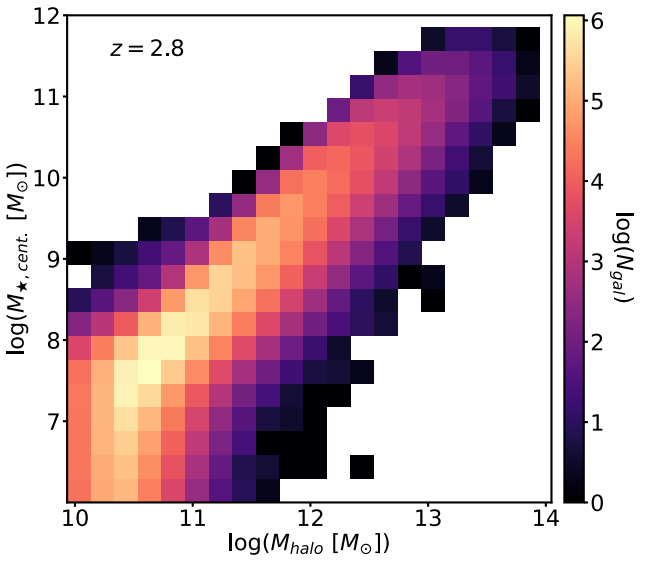

Figure 1. The stellar mass – halo mass relation for central galaxies in a 20×20 grid. We use this distribution to obtain a preliminary sample of dark matter haloes and identify an optimal sample merger trees.

We start our procedure by randomly selecting two haloes in each non-zero cell of the 2D grid. This grid has 215 non-zero cells. Hence, we select 430 FOF haloes/central galaxies. Then, we assign the weight of each halo as the number of selected haloes divided by the total number in its respective cell. Our algorithm *accepts* a sample of haloes if the sum of the average relative errors between the sampled observables and the full predictions is lower than the set of haloes previously accepted. After testing 20 000 sets of haloes, the algorithm retrieves the set with the lowest sum of average relative errors. This first step results in a representative sample of merger trees.

As mentioned earlier, it is not necessarily the case that the observables from all haloes in the set of merger trees are consistent with the entire volume predictions at all redshifts. Therefore, the second step in this process involves selecting a subsample of haloes (from the selected merger trees) at all redshifts separately. Note that `L-Galaxies` runs over the sample of merger trees, and the runtime is almost independent of the halo subsample size. In this step, haloes are selected based only on their virial mass. To find the optimal number of mass bins, we choose the highest number that produces histograms without empty bins (from the set of merger trees) within the mass limits. The procedure starts by testing only one halo per mass bin. If no better sample is found after 50 trials, the number of selected haloes per bin increases.[3] This algorithm ends after 20 000 tests. Note that the number density of SMGs at $z \lesssim 1.5$ is very low, so we do not include this observable when estimating representativeness for $0 < z < 1.5$.

In general, the average relative errors are ~ 10 per cent, ~ 25 per cent, ~ 5 per cent, and ~ 20 per cent for SMFs, $f_{\rm Q}$ s, $n_{\rm SMG}$, and HIMF, respectively. We compare the sampled and full-volume predictions in Appendix A. This procedure yielded a sample of merger trees/haloes representing all the observables we will use to constrain our model, without significantly increasing runtime.

[3]If the mass bin contains fewer haloes than requested, we use all of them without repetition.

3.3 Observational constraints

A crucial input to the L-Galaxies MCMC mode is the set of observational constraints, the observables, as the algorithm compares the proposed model at each MCMC step to this data set. Here, we describe the updates to the observational data used in this study.

3.3.1 Stellar mass function

Henriques et al. (2015, 2020) calibrated the free parameters of L-Galaxies using combined SMFs from SDSS at $z = 0$ (Baldry, Glazebrook & Driver 2008; Li & White 2009; Baldry et al. 2012) and ULTRAVISTA at $z = 2$ (Ilbert et al. 2013; Muzzin et al. 2013). In contrast, we use the SMFs derived by Leja et al. (2020) and the quiescent fractions presented by Leja et al. (2022). The primary motivation for this change is that the SMFs of Leja et al. (2020) provided a resolution to the tension between the observed SFR density and the stellar mass density (Madau & Dickinson 2014; Leja et al. 2015; Tomczak et al. 2016), which is evident in many data sets commonly used for calibration (e.g. Baldry et al. 2012; Ilbert et al. 2013; Muzzin et al. 2013; Tomczak et al. 2014). We provide a summary of these data sets here.

Leja et al. (2020) employed the Prospector SED fitting code (Leja et al. 2017, 2019b), which uses non-parametric star formation histories (SFHs) to construct modelled SEDs. As shown by Leja et al. (2019a), non-parametric SFHs recover input SFHs with significantly less bias compared to parametrized SFHs. Leja et al. (2019b) demonstrated that this method yields stellar masses approximately 0.1–0.3 dex larger and total SFRs approximately $0.1 - 1$ dex lower than previous studies, suggesting a reconciliation between these two observables. Leja et al. (2020) developed a model to describe the evolution of the SMF (referred to as the 'continuity model') by fitting SMFs derived from the 3D-*Hubble Space Telescope* (Skelton et al. 2014) and COSMOS2015 (Laigle et al. 2016) surveys, in the range $0.2 < z < 3.0$. Given the redshift limits of the model, we use the continuity model to derive the SMF at $z = 0.4$ and $z = 2.8$. To achieve this, we followed the procedure outlined in appendix B of Leja et al. (2020) to generate the posterior distribution of the median SMF at each redshift and its associated $1\,\sigma$ uncertainty.

Fig. 2 illustrates the differences between the SMFs used in this work and those used by Henriques et al. (2015, 2020). Across almost the entire stellar mass range above the completeness limit, the SMFs derived from the Leja et al. (2020) continuity model (L20, in Fig. 2) exceed the data set used by Henriques et al. (2015) and Henriques et al. (2020, H20) by 0.1–0.2 dex. This difference generally increases with stellar mass. However, at the highest stellar masses, number densities are low and can be significantly affected by cosmic variance; in the most massive bin error bars are larger and number density estimates are comparable between H20 and L20 at $z \sim 0.4$, $z \sim 1.0$, and $z \sim 2.8$.

3.3.2 Quiescent fraction

Following Henriques et al. (2015, 2020), we used the fraction of 'red/quiescent' galaxies as a function of stellar mass at different redshifts to calibrate the model. In both of these earlier studies, data from Muzzin et al. (2013), Ilbert et al. (2013), and Tomczak et al. (2014) were combined with UVJ colour–colour criteria applied to define 'red' galaxies. At $z = 0$, the Baldry et al. (2004) $u - r$ colour cut was used. However, as discussed in Rodríguez-Puebla et al. (2017), systematic effects – such as the assumed IMF, stellar population synthesis (SPS) model, photometric calibrations, and dust extinction – can bias the models when comparing different data sets. For simplicity and to maintain a homogeneous sample, we used the Leja et al. (2022) data to construct this observational constraint. This data set is the same as that used in Leja et al. (2020) to derive the continuity model of the SMF.

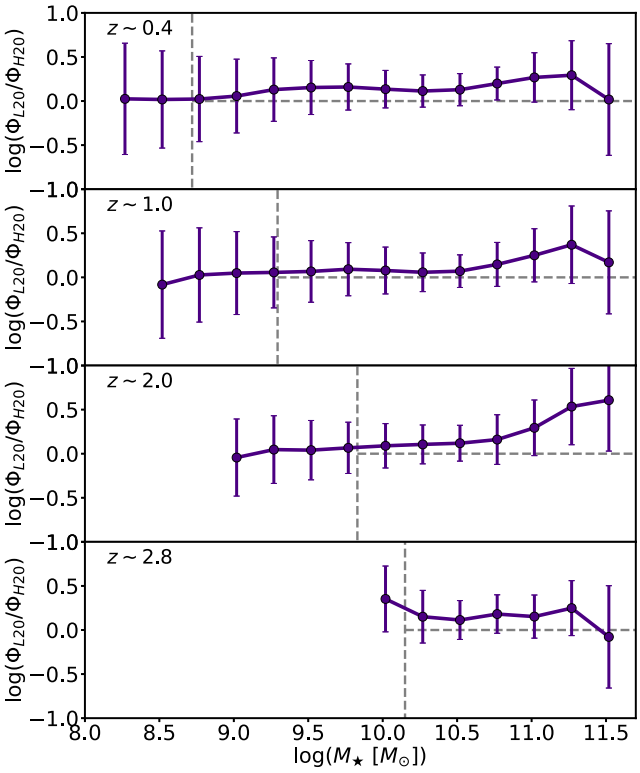

Figure 2. Comparison between the SMFs of Leja et al. (2020) (continuity model used in this work; L20) and those used in Henriques et al. (2020) (compilation from literature; H20) at $z \sim 0.4$, 1.0, 2.0, and 2.8. The error bars represent the propagated uncertainties from both observed SMFs. Dashed grey horizontal and vertical lines indicate unity and the mass completeness of the Leja et al. (2020) data set, respectively. The L20 SMFs indicate higher galaxy number densities, particularly at the massive end.

Additionally, we replaced the fraction of 'red' (using UVJ colour–colour criteria) galaxies with the fraction of 'quiescent' galaxies, defining them using a specific star formation rate (sSFR) threshold of $\log(\mathrm{sSFR/yr^{-1}}) \leq -11$. This choice was motivated by the need for a consistent selection at different redshifts, as well as the need to avoid systematic biases introduced by the different SPS models used to compute galaxy magnitudes. For instance, Leja et al. (2020) and Leja et al. (2022) used the Flexible Stellar Population Synthesis (FSPS) code (Conroy, Gunn & White 2009), whereas the available SPS models used in L-Galaxies include Bruzual & Charlot (2003), Maraston (2005), and Charlot & Bruzual (2007). Instead, we quantified the fraction of galaxies with $\log(\mathrm{sSFR/yr^{-1}}) \leq -11$ in five stellar mass bins ($\log(M_\star/\mathrm{M_\odot}) =$ [9.0–9.5, 9.5–10.0, 10.0–10.5, 10.5–11.0, 11.0–12.0]) at $z = 0.4$, 1.0, 2.0, and 2.8. Similarly to our use of the SMFs as observational constraints, we calibrate the model using only the quiescent fraction (f_Q) at two redshifts, following the calibration framework of Henriques et al. (2020), but adopting $z = 0.4$ and $z = 2.8$ instead of $z = 0.0$ and $z = 3.0$. The primary reason for calibrating the parameters using only two redshifts is to preserve the ability to

compare the model predictions with independent observational data at other redshifts.

In order to obtain robust estimates, for each galaxy in the Leja et al. (2022) data set, we sampled 10 000 values based on the uncertainties in (1) redshift, (2) stellar mass, and (3) sSFR, assuming a Gaussian distribution centred on the most likely value and with σ equal to the uncertainties in these properties. We then estimated the median of the 10 000 sampled $f_{\rm Q}$ values and the associated 1 σ error. $f_{\rm Q}$ is also influenced by sample size, so we computed the total error in $f_{\rm Q}$ as the quadratic sum of the uncertainty-induced error and the Poissonian error.

3.3.3 *Number densities of sub-mm galaxies*

As shown in Araya-Araya et al. (2024), the Henriques et al. (2015) version of `L-Galaxies` underpredicts the sub-mm number counts when using the Cochrane et al. (2023a) scaling relations to model the observed-frame 870 μm flux densities (S_{870}; since `L-Galaxies` does not make predictions for sub-mm fluxes, S_{870} was modelled in that work as a function of SFR, $M_\star$, $M_{\rm dust}$, and redshift, based on detailed radiative transfer post-processing on highly resolved zoom-in galaxies; see Cochrane et al. 2023a). This underprediction of sub-mm number counts persists in the Henriques et al. (2020) version of the SAM. Motivated by obtaining a better match to observationally derived sub-mm number counts, we included SMG number density measurements as an additional observational constraint.

Ideally, the full sub-mm number counts would be used as observational constraints. However, the MCMC mode estimates likelihoods by comparing the proposed model to the observables at a few specific snapshots (redshifts). For simplicity, we instead use the SMG number density as an observational constraint. We estimated the SMG number density using the Dudzevičiūtė et al. (2020) catalogue, which provides photometric redshifts for SMGs observed in the AS2UDS survey (Simpson et al. 2017; Stach et al. 2018, 2019; Dudzevičiūtė et al. 2020). This catalogue consists of 870 μm continuum ALMA follow-up observations of SCUBA-2 detections in the UDS field (S2UDS; Geach et al. 2017), covering an area of 0.96 deg^2. The ALMA survey targeted S2UDS sources with 4 σ detections (i.e. $S_{850} \geq 3.6$ mJy), but the sample is incomplete at these flux densities (see Geach et al. 2017). Therefore, we adopted $S_{870} = 5.2$ mJy (where completeness exceeds 90 per cent) as the flux density threshold for our calibration.

To estimate the number density, we selected all SMGs with photometric redshifts within $z_c \pm \Delta z$, where $z_c = 2.8$ (our highest redshift for SMF and $f_{\rm Q}$), and $\Delta z = 0.35$. The Δz value was chosen based on the average width ($16^{\rm th} - 84^{\rm th}$ percentiles) of the photometric redshift probability density functions of SMGs presented by Dudzevičiūtė et al. (2020). However, note that the number densities in our model do not change significantly – within the error bars – even if smaller values of Δz are adopted. Similar to our approach for $f_{\rm Q}$, we sampled 10 000 values of redshift and S_{870} flux density for each galaxy, accounting for the uncertainties in these estimates. The SMG number density was calculated as the median of the 10 000 sampled values, with the observational uncertainty taken as the standard deviation. Again, we included the Poissonian error contribution due to the sample size. Finally, the number density of galaxies with $S_{870} \geq 5.2$ mJy at $z \sim 2.8$ is $n_{\rm SMG} = (2.48 \pm 0.3) \times 10^{-5}\, h^3\, {\rm Mpc}^{-3}$.

To implement this new observable as a constraint, we modified the `L-Galaxies` code to estimate S_{870} using the Cochrane et al. (2023a) scaling relations and, then, estimate the SMG number density. Since the Cochrane et al. (2023a) scaling relations are parametrized by the average SFR over the last 10 Myr, we used the instantaneous SFR (`SfrInst`) output from `L-Galaxies`. However, this SFR does not account for the contribution from merger-induced starbursts (see Section 4.2.1 for a description of how star formation is modelled in `L-Galaxies`), which occur instantaneously in the model. We therefore modified `L-Galaxies` to incorporate this component into the `SfrInst` parameter.

3.3.4 *Neutral hydrogen mass function*

Like Henriques et al. (2020), we included the H I mass function at $z = 0$ as an additional constraint, alongside the SMF and $f_{\rm Q}$ at two higher redshifts. The observational data combine results from Zwaan et al. (2005), Haynes et al. (2011), and Jones et al. (2018). Recall that in this study, we use SMFs and $f_{\rm Q}$s at $z = 0.4$ and $z = 2.8$, rather than at $z = 0$ and $z = 2.0$ as in Henriques et al. (2020), as explained earlier in this section.

3.4 MCMC configurations

In order to assess how sensitive `L-Galaxies` is to the observational constraints used for calibration, we ran the `L-Galaxies` MCMC mode for nine different sets of constraints. These sets are referred to as configurations throughout this work. In principle, we expect to obtain a different model for each configuration. Understanding which physical models favour specific observables is crucial to identifying the key discrepancies between galaxy formation models and observations.

The configurations tested in this work are listed in Table 1. The observables (SMFs, $f_{\rm Q}$, HIMF, and $n_{\rm SMG}$) at a given redshift ($z =$ 0.4 and 2.8) used as constraints for the different configurations are denoted by the check marks. Note that for configuration base$_{\rm H20}$, we used the same constraints as in Henriques et al. (2020).

3.5 Running the MCMC mode

We ran the `L-Galaxies` MCMC mode with the sample of merger trees obtained as described in Section 3.2, using the observational data detailed in Section 3.3, and for the nine configurations listed in Table 1. The free parameters of the model were initially randomly sampled with a standardized space displacement from the starting point of $\sigma_{\rm initial} = 0.1$, and thereafter randomly sampled from a lognormal distribution with $\sigma = 0.25$ as in Henriques et al. (2020). We used a modified version of the MCMC mode that compares the likelihood at each step in logarithmic space, improving the efficiency of convergence. This is necessary to avoid numerical underflow,

Table 1. List of MCMC configurations used throughout this paper.

Config	SMF		$f_{\rm Q}$		$n_{\rm SMG}$	HIMF
	$z = 0.4$	$z = 2.8$	$z = 0.4$	z =2.8	$z = 2.8$	$z = 0$
0: base$_{\rm H20}$	✓	✓	✓	✓		✓
1: base$_{\rm L20}$	✓	✓	✓	✓		✓
2: all(base$_{\rm L20}$+$n_{\rm SMG}$)	✓	✓	✓	✓	✓	✓
3: no hi-z SMF	✓		✓	✓	✓	✓
4: no hi-z $f_{\rm Q}$	✓	✓	✓		✓	✓
5: no hi-z SMF, $f_{\rm Q}$	✓		✓		✓	✓
6: no HI MF	✓	✓	✓	✓	✓	
7: no low-z $f_{\rm Q}$	✓	✓		✓	✓	✓
8: no low-z SMF, $f_{\rm Q}$		✓		✓	✓	✓

which can round likelihoods to zero due to the high dimensionality of the likelihood space and the presence of multiple observational constraints. In some cases, we encountered likelihood values on the order of $\gtrsim 10^{-100}$.

For each configuration, we ran the MCMC mode with 96 chains for approximately 5000 steps. Although the number of steps is lower compared to traditional MCMC fitting, the high number of chains ensures that the free parameter space is thoroughly explored. Tests confirmed convergence: we did not find any new accepted point (with a higher likelihood) within the final $\gtrsim$1000 steps. After obtaining an initial best-fitting model (i.e. the set of parameters with the highest likelihood) for each configuration, we performed an additional run of 2000 steps. For these new runs, we set the starting point to the previously obtained best-fitting parameters and sampled the proposed parameters with $\sigma_{\rm initial} = 0.05$ and $\sigma = 0.15$.

For each configuration, the MCMC process consumed approximately 110 000 CPU hours. In total, this work used around 1 million CPU hours. The 15 best-fitting parameters obtained for each configuration are presented in Appendix B.

4 RESULTS

In this section, we present the predictions of the galaxy properties obtained by the best fit of each configuration (Section 4.1), alongside key figures that aid in interpreting the main physical aspects of the models (Section 4.2).

4.1 Predictions for galaxy properties

After obtaining a best-fitting model from each configuration, we ran `L-Galaxies` for the entire `Millennium` volume following the procedure described in Section 3.5. Here, we present the main predictions, starting with the properties used to calibrate the model.

4.1.1 Predicted stellar mass functions

In Fig. 3, we show the SMFs for each calibrated configuration at $z = 0.4$, 1.0, 2.0, and 2.8. SMF data at $z = 0.4$ and 2.8 were used to calibrate some configurations. The figure includes two observed SMFs: the SMF used to calibrate the Henriques et al. (2020) model (grey diamonds) and the SMF derived by Leja et al. (2020) (black dots; see Section 3.3.1 for details). As expected, the predicted SMFs for the 'base$_{\rm H20}$' configuration are lower across all redshifts compared to other configurations, as the Henriques et al. (2020) data set was used for calibration.

Overall, our predictions match the observationally inferred SMFs well at $z = 0.4$ and 1.0 but underpredict the number densities at $z = 2.0$ and 2.8, particularly at the massive end. Another noteworthy result is the prediction from the 'no low-z SMF, $f_{\rm Q}$' configuration, which significantly overestimates the number density of massive galaxies at low redshift, due to the lack of constraints there.

4.1.2 Predicted fraction of quiescent galaxies

The second main observational constraint used in this work is the fraction of quiescent galaxies in stellar mass bins, $f_{\rm Q}$, defined by $\log({\rm sSFR}/{\rm yr}^{-1}) < -11$ for all configurations except 'base$_{\rm H20}$'.[4] Fig. 4 shows the $f_{\rm Q}$ predictions at $z = 0.4$, 1.0, 2.0, and 2.8.

[4]For this configuration, the default Henriques et al. (2020) definition was retained exclusively for the calibration process.

Despite using a different definition of quiescent galaxies during the calibration of the 'base$_{\rm H20}$' configuration, its $f_{\rm Q}$ predictions are similar to its counterpart ('base$_{\rm L20}$'), which was also calibrated with different observational data. Both configurations ('base$_{\rm H20}$' and 'base$_{\rm L20}$') provide the best match to observational results up to $z = 2.0$. As expected, when $f_{\rm Q}$ at low redshift is not used as a constraint (in configurations 'no low-z $f_{\rm Q}$' and 'no low-z SMF, $f_{\rm Q}$'), the fraction of quiescent galaxies is significantly underpredicted across all redshifts analysed. This highlights the importance of including $f_{\rm Q}$ as a calibration constraint. In general, the other configurations, which incorporated the number density of SMGs as a constraint, tend to underpredict $f_{\rm Q}$, with the discrepancy being most significant at $z = 2.0$ and 2.8. Among these, the 'no HIMF' configuration performs the best in predicting $f_{\rm Q}$.

4.1.3 Predicted H I mass function

Following Henriques et al. (2020), we use the HIMF at $z = 0$ as an observational constraint in all configurations except 'no HIMF'. Fig. 5 shows our predictions. Overall, the best-fitting models successfully reproduce the observed HIMF and exhibit similar distributions. However, notable differences arise at both the low-mass and high-mass ends of the distribution. For instance, the 'base$_{\rm H20}$' configuration significantly underpredicts the number density at the low-mass end compared to other configurations and, like 'base$_{\rm L20}$', shows an excess at the high-mass end. The strong downturn at low H I mass may be driven by the resolution limit of `Millennium`, which is $\sim 10^{9.5}\,{\rm M}_{\odot}$.

Interestingly, despite the 'no HIMF' configuration not being constrained by the HIMF, its predictions match the observational data well. This result suggests that the HIMF may not be a critical observational constraint.

4.1.4 Predicted SMG population

The primary novelty of this work is the inclusion of the number density of bright sub-mm galaxies, defined here as all galaxies with $S_{870} \geq 5.2$ mJy, as an observational constraint in the calibration process. Our main motivation for adding this constraint is to better match the observed sub-mm number counts. We first present our predictions for the SMG number densities, $n_{\rm SMG}$, at $z = 2.0$ and $z = 2.8$ compared to observational data in Fig. 6. It is important to note that only the $z = 2.8$ $n_{\rm SMG}$ was used to constrain all models, except for the 'base$_{\rm H20}$' and 'base$_{\rm L20}$' configurations, where no SMG constraints are used in the calibration. Both of these configurations critically underpredict the SMG number densities by more than two orders of magnitude, as has been seen in many previous works (see Section 1). This underscores the importance of this constraint for simultaneously reproducing the SMG and quiescent populations.

Despite the inclusion of this constraint, all configurations except 'no low-z SMF $f_{\rm Q}$' underpredict the $n_{\rm SMG}$ at $z = 2.8$ by at least a factor of 2.5 (e.g. 'no hi-z SMF'). This discrepancy could be explained by the underprediction of the massive end of the SMF at $z = 2.8$. In contrast, our predictions for $z = 2.0$ align more closely with observational data. The worst-performing configuration, 'no HIMF', underpredicts the SMG number density at $z = 2.0$ by a factor of ~ 2.1.

The configurations that best match the observed SMG number densities are 'no hi-z SMF', 'no hi-z SMF, $f_{\rm Q}$', and 'all'. The 'no hi-z SMF' configuration slightly overpredicts the number density at $z = 2.0$ and is the closest match at $z = 2.8$, while the 'no hi-z SMF, $f_{\rm Q}$' configuration underpredicts the $z = 2.0$ number density

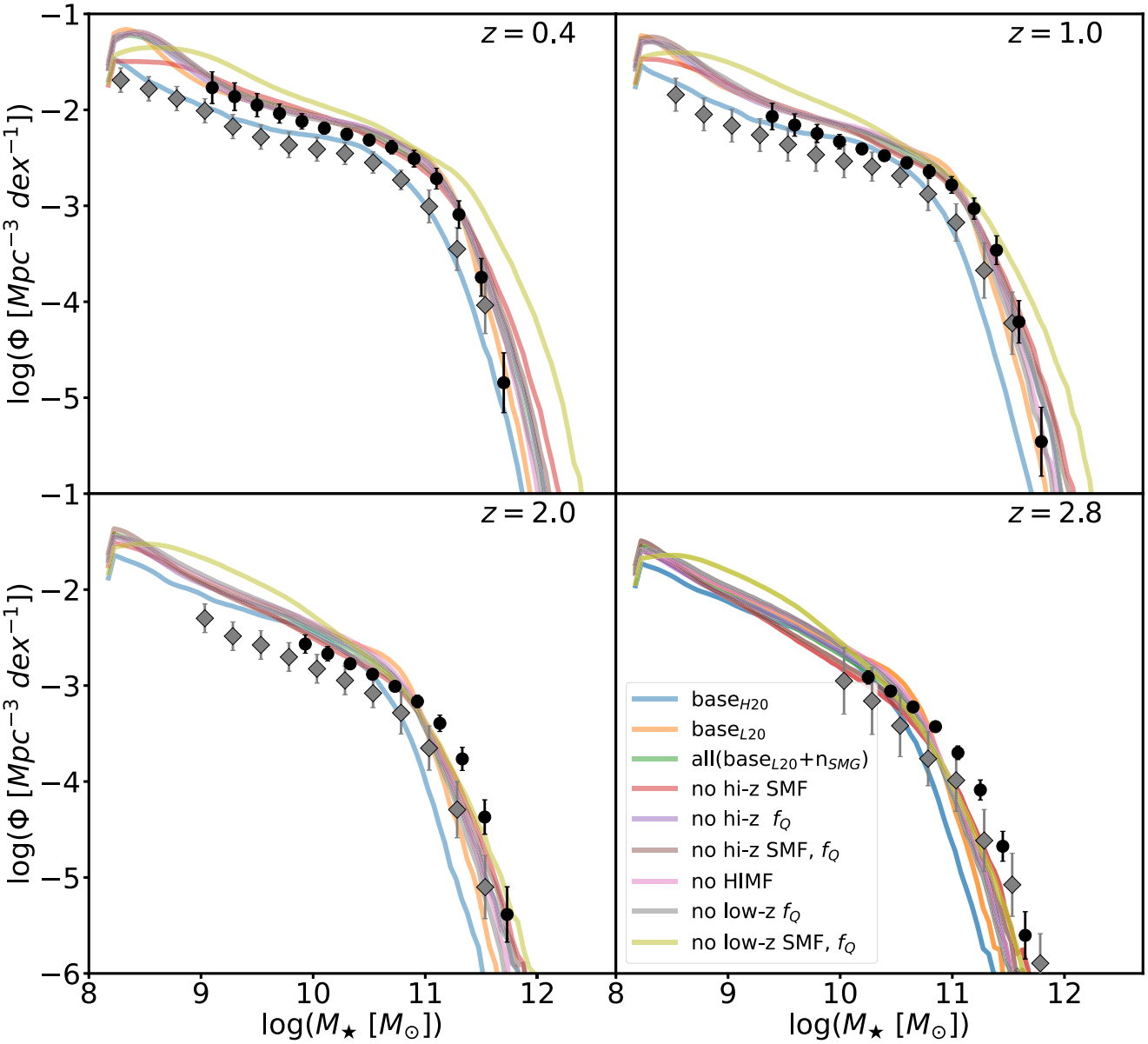

Figure 3. The predicted SMFs at $z = 0.4$, 1.0, 2.0, and 2.8, for every configuration listed in Table 1, compared to the Leja et al. (2020) continuity model (black dots) and the data set used in Henriques et al. (2020) (grey diamonds). Overall, the configurations predict consistent SMFs with the adopted observational constraints at $z = 0.4$ and 1.0. However, they slightly underpredict the observed massive end at $z = 2.0$ and 2.8.

by only a factor of 1.12. In general, the configuration 'all' presents similar number densities to configuration 'no hi-z SMF, $f_{\rm Q}$' but is slightly lower at both redshifts. Interestingly, the configuration 'no low-z SMF, $f_{\rm Q}$' is the only one that matches the observed $n_{\rm SMG}$ at $z = 2.8$, but it overpredicts the $z = 2.0$ $n_{\rm SMG}$ by a factor of 6.6. Notably, this configuration does not exhibit significant differences in the SMFs or $f_{\rm Q}$ at high redshift ($z = 2.0$ and $z = 2.8$) compared to other configurations, nor does it differ in the HIMF predictions at $z = 0$.

The most accurate way to compare observed differential number counts (number of galaxies per flux-density bin and per unit area) with our model predictions is by constructing mock galaxy catalogues. Following the prescriptions in Araya-Araya et al. (2021) (for sky galaxy positions) and Araya-Araya et al. (2024) (for sub-mm flux densities), we create a 36 deg^2 mock for the best-fitting model of each configuration. Fig. 7 shows the predicted S_{870} number counts for all configurations, compared to the observational results from Geach et al. (2017) and Stach et al. (2019). Additionally, we include the S_{870} number counts presented in Araya-Araya et al. (2024), derived from a mock catalogue (also with a 36 deg^2 area) constructed using the Henriques et al. (2015) version of `L-Galaxies`.

The first notable result is that including the number density of galaxies with $S_{870} \geq 5.2$ mJy at $z = 2.8$ as an observational constraint (a single data point) improves the consistency of the predicted S_{870} number counts with observational results. As anticipated from the comparison with direct number density measures in Fig. 6, configurations 'base$_{\rm H20}$' and 'base$_{\rm L20}$' severely underpredict the S_{870} number counts, whereas configuration 'no low-z SMF, $f_{\rm Q}$' overpredicts them. The remaining configurations show significantly better agreement with the observed number counts compared to the

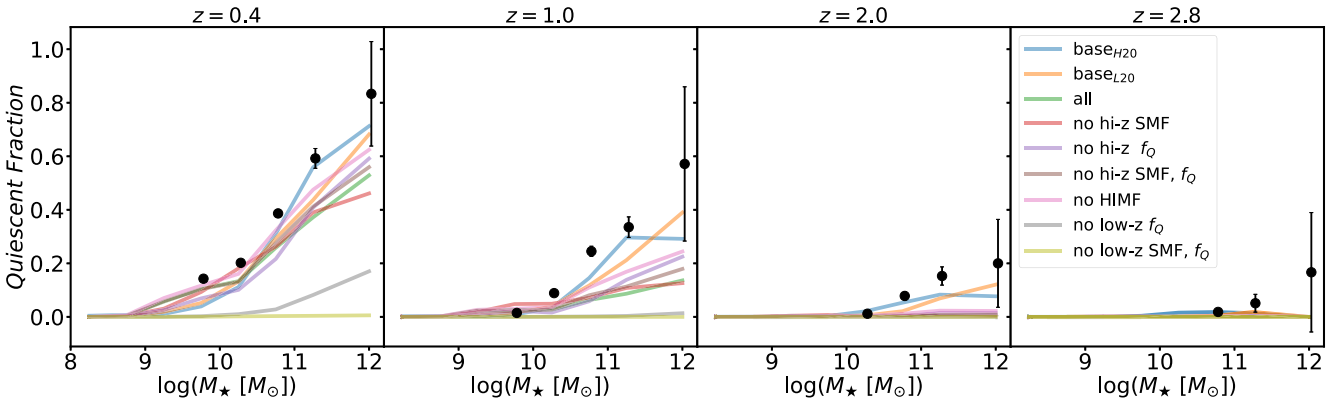

Figure 4. The predicted quiescent fraction (f_Q; $\log(\mathrm{sSFR}/\mathrm{yr}^{-1}) \leq -11$) as a function of stellar mass at $z = 0.4$, 1.0, 2.0, and 2.8, for every configuration listed in Table 1, compared to Leja et al. (2022) data (black dots). Configurations where the number density of SMGs is not input as an observational constraint ('base$_{H20}$' and 'base$_{L20}$') are in better agreement with the observational data. When f_Q at low-z is not a constraint ('no low-z f_Q' and 'no low-z SMF; f_Q'), the model critically underpredicts this population.

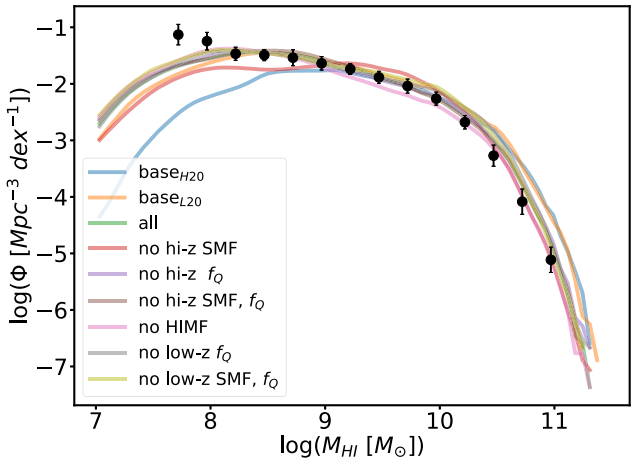

Figure 5. The predicted HIMF at $z = 0$ for every configuration listed in Table 1, compared to the observational data used as a constraint (black dots; compilation of results from Zwaan et al. 2005; Haynes et al. 2011; Jones et al. 2018). Overall, all predicted HIMFs are similar and consistent with the observational data. The main differences are at the low-mass and high-mass end of the distributions. Configurations where n_{SMG} is not used as a constraint ('base$_{H20}$' and 'base$_{L20}$') show an excess of galaxies with large H I reservoirs. Although the HIMF was not a constraint for configuration 'no HIMF', its prediction agrees with the observed HIMF.

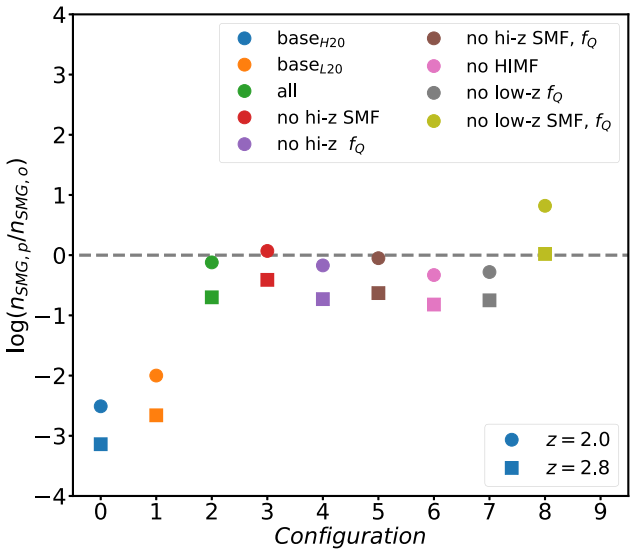

Figure 6. Deviation of the predicted number density of SMGs, $n_{\mathrm{SMG,p}}$, from the observationally inferred number density, $n_{\mathrm{SMG,o}}$ at $z = 2.0$ (dots) and $z = 2.8$ (squares), for our various models listed in Table 1. Our new models that include n_{SMG} at $z = 2.8$ as a constraint match observational estimates significantly better than those that do not (orange and blue symbols). The observed SMG number densities are best-matched by the 'all', 'no hi-z SMF', and 'no hi-z SMF, f_Q' configurations.

previous version of the model presented in Araya-Araya et al. (2024), although most still slightly underpredict observed SMG number counts. Among these, we highlight configuration 'no hi-z SMF', which nearly matches the observational data. Configurations 'all' (calibrated with all observational constraints) and 'no hi-z SMF, f_Q' also demonstrate good agreement with the data, exhibiting similar distributions. In contrast, configuration 'no HIMF', which performed better at predicting the low-z quiescent population when the n_{SMG} constraint was included, shows slightly poorer performance in reproducing sub-mm number counts. Nevertheless, even the 'no HIMF' configuration shows improved consistency compared to the previous version of the model.

Another valuable comparison with observational data is the redshift distribution of SMGs. We compare our predictions for the number of bright SMGs ($S_{870} \geq 5.2$ mJy) in redshift bins normalized by the sky area, with observational data. Fig. 8 shows the predictions of the best-fitting model for each configuration, compared to AS2UDS data (Dudzevičiūtė et al. 2020). The configuration 'no low-z SMF, f_Q', which overpredicts the sub-mm number counts (Fig. 7), exhibits a peak in the distribution at $z \sim 1.2$, significantly lower than the observed peak at $z \sim 3$. Of the two configurations that critically underpredict the SMG number density, 'base$_{H20}$' better matches the shape of the observed redshift distribution compared to 'base$_{L20}$'. All other configurations show a similar distribution shape, differing mainly in the number densities across redshift bins, which can be linked to small differences in sub-mm number counts. However, the peak redshift for these models ($z \sim 2$) is also slightly lower than the observed peak. While our best models for predicting sub-mm number counts fail to capture the sharp decrease in galaxy numbers

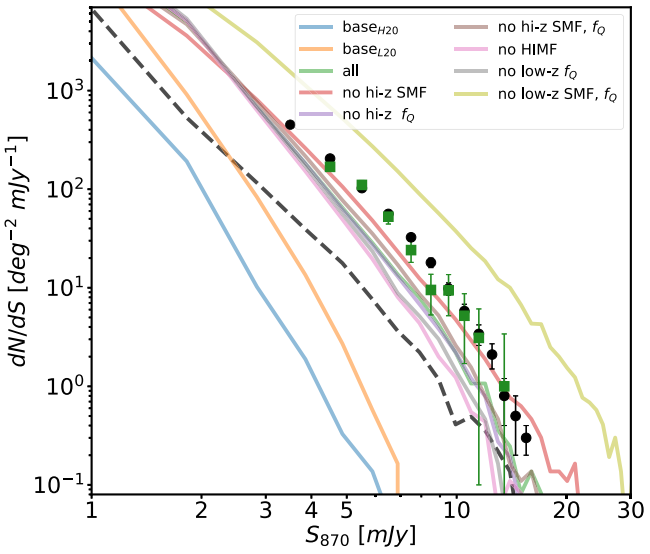

Figure 7. The predicted sub-mm (870 μm) number counts from a mock catalogue constructed for every configuration listed in Table 1, compared to observed number counts derived by Geach et al. (2017) (black dots) and Stach et al. (2019) (green squares). As a comparison, the Araya-Araya et al. (2024) number counts from the Henriques et al. (2015) version of `L-Galaxies` is included (black dashed line). Most of our calibrated models match the observed S_{870} number counts to within a factor of a few, across an order of magnitude in sub-mm flux density. However, when $n_{\rm SMG}$ is not an observational constraint ('base$_{\rm H20}$' and 'base$_{\rm L20}$'), the number counts are critically underpredicted (see orange and blue lines). When neither the low-z SMF nor $f_{\rm Q}$ is used as a constraint, sub-mm number counts are overpredicted (see lime green line).

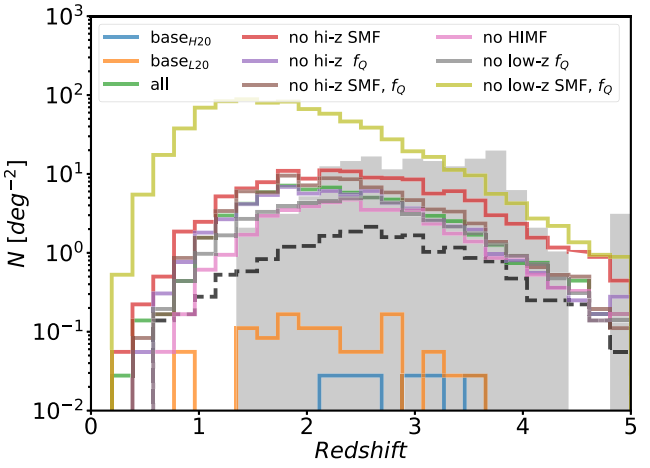

Figure 8. The predicted redshift distribution of bright SMGs ($S_{870} \geq 5.2$ mJy) from a mock catalogue constructed for every configuration listed in Table 1 (coloured lines), compared to observational data from Dudzevičiūtė et al. (2020) (grey filled histogram). As a comparison, the bright SMG redshift distribution predicted by Araya-Araya et al. (2024) using the Henriques et al. (2015) version of `L-Galaxies` is included (black dashed histogram). Most of the configurations present similar redshift distributions.

at $z \lesssim 1.4$, this discrepancy could be due to the small sky area of the observational data and survey selection effects.

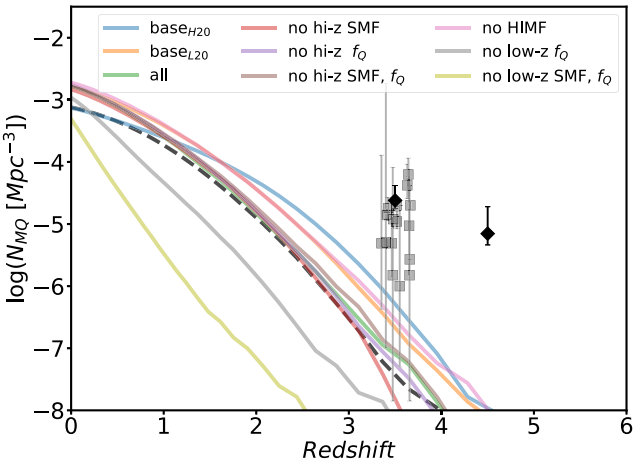

Figure 9. The predicted evolution of the number density of massive ($\log(M_\star/{\rm M}_\odot) \geq 10.6$) quiescent (sSFR $< 0.2/t_{\rm obs}(z)$, where $t_{\rm obs}(z)$ is the age of the universe at redshift z) galaxies obtained for all configuration listed in Table 1. The grey squares are previous observational results compiled by Valentino et al. (2023), while the black diamonds are new measurements from Valentino et al. (2023). We note that Valentino et al. (2023) did not detect any MQs at $5 < z < 6$. As a comparison, the number density from the Henriques et al. (2015) version of `L-Galaxies` is included (black dashed line). Except for configuration 'no HIMF', when $n_{\rm SMG}$ is not an observational constraint ('base$_{\rm H20}$' and 'base$_{\rm L20}$'), the number density of MQs at high-z is consistent with the lower limits of the observational data. Most of the configurations that match the observed S_{870} number counts underpredict the number density, highlighting the longstanding tension in modelling both populations.

4.1.5 Number density of massive quiescent galaxies

In this work, we have used the quiescent fraction of massive galaxies at two different redshifts ($z = 0.4$ and/or $z = 2.8$) as an input calibration constraint for some configurations. As we show in Fig. 4, observational work suggests that the fraction of quiescent galaxies decreases significantly towards high redshift. In Fig. 9, we present the evolution of the number density of massive ($\log(M_\star/{\rm M}_\odot) \geq 10.6$) quiescent galaxies predicted by our various model configurations. In this figure, we use the QG definition adopted by Carnall et al. (2020): sSFR $< 0.2/t_{\rm obs}(z)$, where $t_{\rm obs}(z)$ is the age of the universe at the redshift z. As shown by Carnall et al. (2020) and Valentino et al. (2023), this definition is virtually equivalent to others commonly used in the literature. Note that this definition is different to that adopted to calibrate the model. The observational data in Fig. 9 is drawn from Valentino et al. (2023), who compiled data from several studies quantifying the number density of MQs at $3 \leq z \leq 4$. This compilation includes results from Muzzin et al. (2013), Straatman et al. (2014), Davidzon et al. (2017), Schreiber et al. (2018), Cecchi et al. (2019), Girelli, Bolzonella & Cimatti (2019), Merlin et al. (2019), Carnall et al. (2020), Shahidi et al. (2020), Gould et al. (2023), Weaver et al. (2023), and Carnall et al. (2023). Note that we add a small scatter ($\Delta z = \pm 0.25$) on the median redshift ($z = 3.5$) of the observational data for visualization purposes.

The most notable result from Fig. 9 is that none of our configurations can match the median number density of MQs inferred observationally. Among the configurations, 'no low-z SMF, $f_{\rm Q}$', which significantly overpredicts the sub-mm number counts, severely underpredicts the number density of high-redshift MQs. The configuration with the best performance in this metric is 'base$_{\rm H20}$', achieving number densities comparable to the lower bounds of the observations. However, its predictions for sub-mm number counts

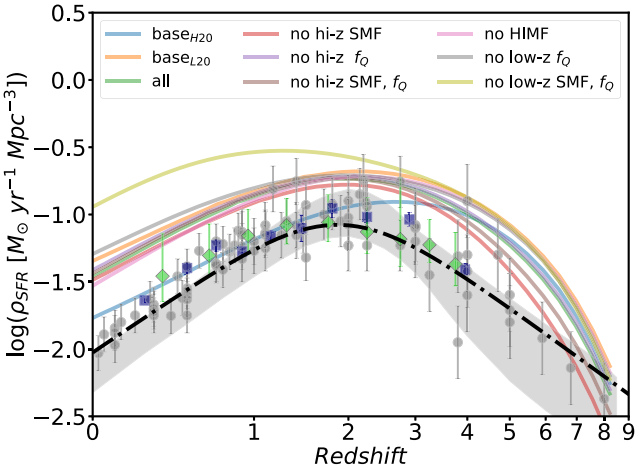

Figure 10. The predicted CSFRD for every configuration listed in Table 1 compared to Behroozi et al. (2013) (grey dots), Madau & Dickinson (2014) function (black dash–dotted line), Leslie et al. (2020) (green diamonds), Zavala et al. (2021) (grey area), and Cochrane et al. (2023b) (navy squares) results from observations. Overall, our calibrated models overpredict the SFR density, even when n_{SMG} is not used as a constraint. All predicted CSFRDs peak at a similar redshift, which is approximately consistent with observational results, except for configurations 'base$_{H20}$' (which peaks earlier) and 'no low-z SMF; f_Q' (which peaks later).

show the worst agreement with the observational data. Interestingly, the configuration 'no hi-z f_Q' (where f_Q at $z = 2.8$ was not used as a constraint) provides better agreement with the observational data than 'no low-z f_Q' (where f_Q at $z = 0.4$ was not used as a constraint). This suggests that using f_Q at low redshift as a constraint has a greater impact on reproducing the evolution of the quiescent population than applying it only at high redshift. Nevertheless, calibrating the model using the low-redshift f_Q does not ensure consistency at higher redshifts, as all configurations struggle to reproduce the observed number density of MQs.

Overall, the predicted number density of MQs from configurations where the number density of SMGs was used as a constraint can match only the lower limits of the observational data. The unique exception is the 'no HIMF' configuration, whose predictions are similar to those of the base$_{H20}$' and 'base$_{L20}$' configurations. Based on Figs 7 and 9, our results demonstrate that at high redshift, a higher number density of SMGs corresponds to a lower number density of MQs, and vice versa. These findings clearly highlight the tension in modelling both the SMG and massive quiescent galaxy populations. Interestingly, the 'no HIMF' configuration serves as a 'in between the tension' model, producing predictions that do not significantly underrepresent either galaxy population.

4.1.6 Cosmic SFR density

One of the most fundamental measures in extragalactic astronomy is the evolving CSFRD, which traces the history of star formation and serves as a critical tool for testing galaxy evolution models. Our predictions for the CSFRD are presented in Fig. 10, where we compare them to the best-fitting function from Madau & Dickinson (2014) (converted to a Chabrier 2003 IMF) and the observational data from Behroozi, Wechsler & Conroy (2013), Leslie et al. (2020), Zavala et al. (2021), and Cochrane et al. (2023b). All predicted CSFRDs are higher in normalization than the observational measurements, across all epochs. The CSFRD from configuration 'base$_{H20}$' shows better agreement with observational constraints at intermediate redshifts ($0.5 \lesssim z \lesssim 2$), but it overestimates the SFRD both near $z = 0$ and at higher redshifts, with a peak occurring around $z \sim 3$. Notably, the SFRD predicted by configuration 'base$_{H20}$' deviates from the results of Henriques et al. (2020), particularly in the redshift at which the CSFRD peaks ($z \sim 3$ versus $z \sim 2$ in Henriques et al. 2020). This discrepancy may be attributed to differences in the redshift ranges of the observational constraints used during calibration (e.g. $z = 0.4$ and $z = 2.8$ instead of $z = 0$ and $z = 2$), which can significantly influence the shape of the resulting CSFRD. For instance, configuration 'base$_{H20}$' yields high-redshift f_Q predictions that are more consistent with observations, whereas Henriques et al. (2020) tends to underestimate this quantity. Consequently, in order to simultaneously match the high-redshift SMF and f_Q, configuration 'baseH20' required the peak of star formation to occur earlier than in the Henriques et al. (2020) model. On the other hand, configuration 'base$_{L20}$' predicts a CSFRD similar to those obtained from configurations where the number density of SMGs is included. This suggests that the difference in shape between the 'base$_{H20}$' and the new configurations is primarily driven by our changes to the SMFs used in the calibration. Most configurations predict similar SFRDs below the peak redshift, except for 'no low-z SMF, f_Q', which is approximately 0.5 dex higher than the others. Additionally, the CSFRD peak for this configuration occurs at $z \sim 1.2$, aligning with the peak in the SMG redshift distribution for this configuration. In contrast, the other configurations have peaks near $z \sim 2$, consistent with observational data. Slightly larger differences occur at higher redshifts, where the difference between the upper and lower SFRD predictions at $z \sim 8$ is ~ 0.4 dex.

4.1.7 SMBH mass function

Another important galaxy property is the SMBH mass. In `L-Galaxies`, the strength of AGN feedback, which injects energy into the hot gas atmosphere, suppressing cooling, depends on the SMBH mass. This process regulates star formation, as only cold gas can form new stars. Note that the SMBH mass function is not used as an observational constraint when calibrating the models.

We present our predictions for the SMBH mass function at $z = 0$ in Fig. 11, also comparing to the best-fitting function from Tucci & Volonteri (2017) and observational results from Vika et al. (2009) and Shankar et al. (2020). The predicted SMBH mass functions display two main shapes. The 'no low-z SMF, f_Q', 'no HIMF', 'no hi-z SMF, f_Q', and 'all' configurations exhibit a clear peak in black hole mass, below which the number densities of SMBHs decrease rapidly. The location of the peak depends on the configuration. For example, the 'no low-z SMF, f_Q' configuration, which overestimates the SMBH mass function, peaks at $\log(M_{BH}/M_\odot) \sim 7.5$. Configurations 'no hi-$z$ SMF, f_Q' and 'all' have similar peaks at $\log(M_{BH}/M_\odot) \sim 6.7$. In contrast, the 'no HIMF' configuration reasonably reproduces the observed SMBH mass function, with its peak at $\log(M_{BH}/M_\odot) \sim 7.2$. The remaining configurations underpredict the observationally inferred black hole number densities at intermediate and low masses, with black hole mass functions more similar in shape to SMFs (i.e. not peaked). It is worth noting that a more detailed formalism for the growth of SMBHs in `L-Galaxies` (a modification of Henriques et al. 2015, as presented in Izquierdo-Villalba et al. 2020), results in SMBH mass functions that are in good agreement with observations.

4.1.8 Summary of galaxy property predictions

We quantify and summarize the comparison between the predicted observables and the corresponding observational data for each model configuration in Table 2. This is done by computing the median logarithmic difference (in dex) between model predictions and observations. The colours green, light green, yellow, light red, and red indicate levels of agreement corresponding to $< 0.1, 0.3, 0.5, 1.0$, and > 1.0 dex, respectively. For observables with multiple estimates in the literature—such as the sub-mm number counts, the number density of MQs at $z \sim 3.5$, the SMBH mass function at $z = 0$, and the CSFRD—we adopt the median of the reported values.

As shown in Fig. 7, while the predicted sub-mm number counts are generally consistent with the observed range within uncertainties, most configurations yield only moderate agreement with the median values. In contrast, all models substantially underpredict the number density of MQs at $z \sim 3.5$ relative to the median observational estimate, with consistency achieved only with the reported lower limits (see Fig. 9). Among all configurations, 'no HIMF' provides the best overall agreement with both the sub-mm number counts and the number density of MQs at high redshift.

4.2 Physical interpretation

As we have demonstrated, the best-fitting models for each configuration produce different predictions for the SMF, $f_{\rm Q}$, HIMF, sub-mm number counts, the evolution of the number density of MQs, the CSFRD, and the SMBH mass function. In this subsection, we present key results that provide insights into the physical processes modelled by each configuration.

4.2.1 Star formation

In `L-Galaxies`, there are two physical drivers of star formation: the molecular hydrogen surface density (a secular mechanism, implemented as a Kennicutt–Schmidt-type scaling relation) and merger-induced starbursts. We discuss each of these mechanisms in turn, starting with the secular mode. The Henriques et al. (2020) version of `L-Galaxies`, which tracks H_2 in spatially resolved rings, assumes that the star formation density is proportional to the surface density of H_2 (Fu et al. 2013), with an inverse dependence on the dynamical time. The proportionality constant, $\alpha_{\rm SF}$, is a free parameter in the model. In Fig. 12 (top panel), we show the best-fitting $\alpha_{\rm SF}$ for each configuration. We find that the '$\rm base_{H20}$' configuration exhibits the highest efficiency in converting H_2 into stars, followed by '$\rm base_{L20}$'. These two models were calibrated without including the number density of SMGs. The remaining configurations, which were calibrated using $n_{\rm SMG}$, display similar $\alpha_{\rm SF}$ values, with the lowest efficiency seen in the 'no low-z SMF, $f_{\rm Q}$' configuration.

The second star formation mechanism is the 'collisional starburst' formalism from Somerville, Primack & Faber (2001), which describes the conversion of cold gas into stars triggered by galaxy mergers. This mechanism has two associated free parameters, $\alpha_{\rm SF,burst}$ and $\beta_{\rm SF,burst}$:

$$M_{\star,\rm burst} = \alpha_{\rm SF,burst} \left(\frac{M_1}{M_2}\right)^{\beta_{\rm SF,burst}} M_{\rm cold}, \quad (1)$$

where M_1 and M_2 ($M_2 > M_1$) are the total baryonic mass of the two merging galaxies, and $M_{\rm cold}$ is the sum of the cold gas masses (i.e. the sum of H I, H_2, and metals in the ISM). We show in Fig. 12 (bottom panel) the scaling relation in equation (1) as a function of the galaxy mass ratio for the best-fitting models of each configuration. Unlike the secular mode, configuration '$\rm base_{H20}$' demonstrates the lowest efficiency in converting cold gas into stars during mergers, similar to 'no low-z SMF, $f_{\rm Q}$'. In these models, bursts only occur for mergers with mass ratios above ∼1:3. The efficiency then rises, reaching ∼20 per cent and ∼55 per cent of the cold gas converted into stars for '$\rm base_{H20}$' and 'no low-z SMF, $f_{\rm Q}$', respectively. The 'no low-z $f_{\rm Q}$' configuration is also inefficient in this regime, increasing the burst efficiency gradually with mass ratio, achieving ∼15 per cent in equal-mass mergers.

These three models contrast with the others, which exhibit strong burst efficiencies even at low mass ratios –converting ∼20–40 per cent of cold gas into stars during mergers of mass ratio 1:1000, and up to ∼60–90 per cent for 1:1 mergers. Configurations '$\rm base_{L20}$' and 'no hi-z SMF, $f_{\rm Q}$' share identical scalings, as their best-fitting starburst parameters are the same. 'No HIMF', 'no hi-z SMF', and 'no hi-z $f_{\rm Q}$' also show similar behaviour, with minor differences at low mass ratios. These are the most efficient configurations in this channel, achieving up to ∼90 per cent gas conversion in major mergers. The 'all' configuration shows a similar trend to 'no HIMF', but with a lower normalization.

Interestingly, configuration 'no low-z SMF, $f_{\rm Q}$', which predicts the highest SMF number densities at low redshifts and strongly overpredicts sub-mm number counts, is the least efficient at forming stars through both secular and burst modes. This suggests its elevated SMF and low quiescent fractions result from ineffective feedback. We explore feedback in the following subsections. Also noteworthy is the difference between '$\rm base_{H20}$' and '$\rm base_{L20}$': the former, calibrated with lower SMF normalization, forms stars primarily through secular processes, with bursts occurring only for mass ratios $\gtrsim 0.5$. In contrast, the '$\rm base_{L20}$' model – calibrated with the SMFs from Leja et al. (2020) – requires both high secular and burst efficiencies, similar to most configurations that include the SMG number density as a constraint.

4.2.2 Feedback from supernovae and stellar winds

Especially during the late stages of a star's life, significant amounts of material and energy are released into the ISM through SNe and stellar winds. This process is critical to galaxy evolution, as it heats the cold gas required for star formation and enriches the ISM with metals. In L-Galaxies, SN feedback heats the cold gas, transferring it to the hot gas phase. Any remaining energy further heats the hot gas, suppressing cooling and potentially driving outflows. In the Henriques et al. (2020) version of `L-Galaxies`, this energy is released at the end of a star's life (see Yates et al. 2013), rather than immediately, as assumed under the instantaneous recycling approximation commonly used in other simulations. This spreads out the injection of energy (and metals) over time, especially for SNe-Ia.

SN feedback is governed by two key efficiencies: reheating cold gas from the disc into the hot atmosphere ($\epsilon_{\rm disk}$), and ejecting gas from the hot atmosphere into the circumgalactic medium ($\epsilon_{\rm halo}$). Both follow the same functional form, parametrized by three free parameters (η_x, V_x, and β_x):

$$\epsilon_x = \eta_x \times \left[0.5 + \left(\frac{V_{\rm max}}{V_x}\right)^{-\beta_x}\right], \quad (2)$$

where x denotes either reheating or ejection, and $V_{\rm max}$ is the maximum circular velocity of the host dark matter halo. Since $\epsilon_{\rm halo}$ represents the fraction of SN energy used for ejection, values above

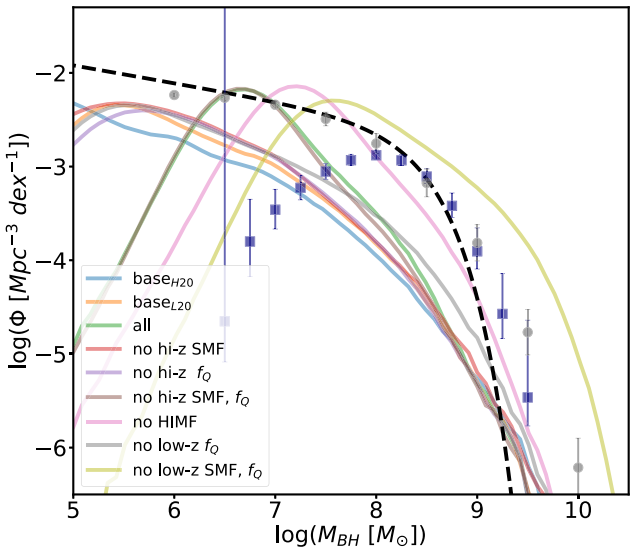

Figure 11. The SMBH mass function at $z = 0$, predicted by each of the configurations listed in Table 1. We overplot observational results from Vika et al. (2009) (navy squares), Tucci & Volonteri (2017) (dashed black line), and Shankar et al. (2020) (grey circles) for comparison. The SMBH mass function from configuration 'no HIMF' (which is reasonably consistent with the S_{870} number counts and number density of MQs at high-z) is the most consistent with observational data. We identify two main SMBH mass function shapes: peaked distributions and Schechter-like distributions. These shapes are explained by the black hole growth model (see Fig. 14; top panel).

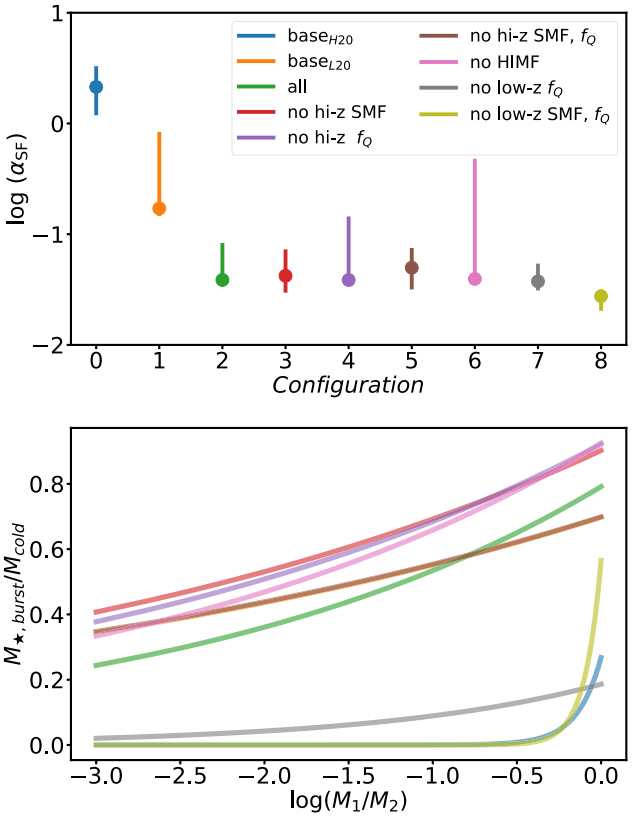

Figure 12. *Top:* The best-fitting parameter associated with the efficiency in converting H_2 into stars (secular star formation). The error bar indicates the 16th and 84th percentiles of the final 2000 MCMC runs of the 96 chains. Models calibrated with n_{SMG} present similar and lower efficiency in forming stars from H_2 surface density, compared to configurations 'base$_{H20}$' and 'base$_{L20}$'. *Bottom:* The best-fitting scaling relation (equation 1) that describes the fraction of cold gas converted into stars driven by mergers as a function of the mass ratio of every configuration. Except for configuration 'base$_{H20}$', configurations that critically underpredict the quiescent population ('no low-z f_Q' and 'no low-z SMF, f_Q') have a low fraction of stars formed in merger-induced starbursts. All other models present an elevated starburst efficiency. This can also deplete the subsequent star formation due to the small amount of remaining cold gas to form new stars.

1 are not allowed in the model, even if the functional form permits them. In this sense, the scaling may not fully capture the efficiency of SN-driven gas ejection. In the original Henriques et al. (2020) model, ϵ_{halo} saturates at 1 across all V_{max}.

We show the scaling relations for reheating and ejection as a function of V_{max} for each configuration in Fig. 13 (top and bottom panels, respectively). The reheating efficiencies fall into three distinct groups. Configurations 'base$_{H20}$' and 'baseL20' show nearly linear relations with the weakest V_{max} dependence, but with the highest reheating efficiencies at $\log(V_{max}) \gtrsim 1.9$ and $\gtrsim 2.1$, respectively. Efficient reheating in massive haloes is likely important for quenching massive galaxies. In contrast, 'no low-z SMF, f_Q' and 'no hi-z SMF' exhibit steep inverse dependencies on V_{max}, differing only by an offset. These are most effective at reheating gas in low-mass haloes but inefficient in massive systems. This helps explain why the 'no low-z SMF, f_Q' configuration overpredicts the SMFs

Table 2. The median logarithmic difference (dex) between model predictions of each configuration and observations. The colours green, light green, yellow, light red, and red indicate levels of agreement corresponding to <0.1, 0.3, 0.5, 1.0, and >1.0 dex, respectively. We assign a value of −99.9 when, on average across the data points of a given observable, the model fails to predict even a single galaxy.

Config	SMF		f_Q		HIMF	S_{870} counts	n_{MQ}	SMBH MF	CSFRD
	$z = 0.4$	$z = 2.8$	$z = 0.4$	$z = 2.8$	$z = 0$	$z \lesssim 6$	$z = 3.5$	$z = 0$	$z < 6$
0: base$_{H20}$	0.05	-0.4	-0.07	-0.58	-0.01	-99.9	-0.90	-1.26	0.09
1: base$_{L20}$	0.08	-0.45	-0.12	-0.55	0.06	-2.79	-1.30	-1.27	0.49
2: all	0.08	-0.41	-0.16	-2.31	0.04	-0.42	-1.81	-1.12	0.45
3: no hi-z SMF	0.11	-0.26	-0.16	-2.52	0.11	-0.14	-2.19	-1.2	0.42
4: no hi-z f_Q	0.08	-0.38	-0.23	-2.35	0.05	-0.47	-1.99	-1.31	0.48
5: no hi-z SMF, f_Q	0.08	-0.37	-0.15	-2.11	0.08	-0.35	-1.69	-1.22	0.43
6: no HIMF	0.07	-0.35	-0.08	-1.23	-0.02	-0.66	-1.22	-0.38	0.45
7: no low-z f_Q	0.08	-0.42	-1.09	-99.9	0.05	-0.63	-2.79	-0.88	0.49
8: no low-z SMF, f_Q	0.24	-0.33	-2.16	-99.9	0.10	0.71	-99.9	0.34	0.64

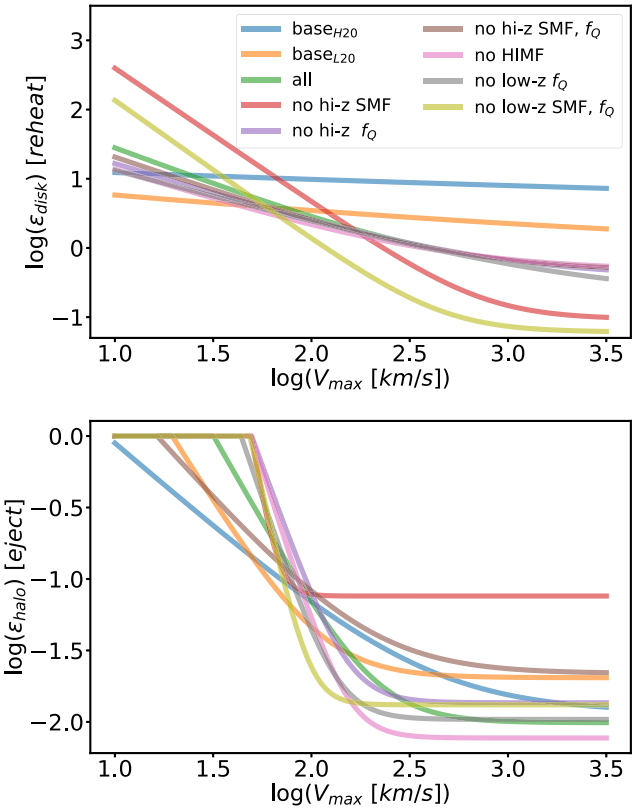

Figure 13. *Top:* The best-fitting scaling relation (equation 2) that describes the efficiency of heating the cold gas and reheating the hot gas atmosphere, $\epsilon_{\rm disk}$, as a function of the maximum halo rotational velocity, $V_{\rm max}$ – a proxy of halo mass. The model that best matches the observed S_{870} number counts ('no hi-z SMF') and the model that overpredicts them ('no low-z SMF, f_Q') both show a heating efficiency that is strongly dependent on halo mass: these models show the least efficient heating in massive haloes but the most efficient heating in low-mass haloes. On the other hand, the two models that critically underpredict the sub-mm number counts ('base$_{\rm H20}$' and 'base$_{\rm L20}$') present a weak dependence in $V_{\rm max}$, having the strongest efficiency of heating cold gas for high-mass haloes. *Bottom:* The best-fitting scaling relation (equation 2) that describes the fraction of available SN energy to eject gas (in outflows) from the galaxy's hot gas atmosphere, $\epsilon_{\rm halo}$, as a function of the maximum halo rotational velocity, $V_{\rm max}$. The nine models have similar efficiency for ejecting gas in intermediate and high-mass systems, except for configuration 'no hi-z SMF'. This could be the reason for this configuration predicting the lowest SFRD at higher redshifts.

and CSFRD at low redshift—not due to enhanced star formation efficiency, but due to weak feedback allowing more gas retention. The other configurations show broadly similar reheating trends with only small differences.

The second relation (Fig. 13, bottom) tracks the fraction of SN energy used for ejecting hot gas. At high $V_{\rm max}$ ($\log(V_{\rm max}/{\rm km\,s^{-1}}) \gtrsim 2$), most configurations yield similar ejection efficiencies ($\log(\epsilon_{\rm halo}) \sim -2$ to -1.5). At low velocities ($V_{\rm max} \lesssim 50\,{\rm km/s}$), $\epsilon_{\rm halo}$ saturates to 1 in most models. This scale corresponds to haloes with $M_{\rm vir} \lesssim 10^{11}\,{\rm M}_\odot$. The 'no hi-$z$ SMF' model displays the highest ejection efficiency across almost all halo masses. Strong feedback in low- and intermediate-mass haloes, especially common at early times, likely explains this configuration's suppressed high-z SFRD.

On the other hand, 'no low-z SMF, f_Q' also shows high reheating and ejection efficiency in low-mass haloes and the lowest overall star formation efficiency (Fig. 12). None the less, it predicts a high SFRD at low redshift, driven by a larger number of intermediate-mass star-forming galaxies ($M_\star \sim 10^{8.5}$–$10^{10}\,{\rm M}_\odot$), as discussed in Section 4.2.1.

4.2.3 *AGN feedback*

As shown in Fig. 13, SN and stellar wind feedback become inefficient in massive galaxies, making AGN feedback crucial for regulating star formation in these systems. The 2020 version of `L-Galaxies` adopts the Croton et al. (2006) framework, which includes two AGN feedback modes: quasar mode, which governs SMBH growth during galaxy mergers, and radio mode, which injects energy into the hot gas halo to suppress cooling.

In the quasar mode, SMBHs grow by accreting cold gas during mergers. The accreted mass is given by

$$\Delta M_{\rm BH,Q} = \frac{f_{\rm BH}(M_{\rm sat}/M_{\rm cen})M_{\rm cold}}{1+(V_{\rm BH}/V_{200c})^2}, \quad (3)$$

where $M_{\rm cen}$ and $M_{\rm sat}$ are the baryonic masses of the central and satellite galaxies, $M_{\rm cold}$ is their combined cold gas mass, and V_{200c} is the host halo's virial velocity. The $f_{\rm BH}$ and $V_{\rm BH}$ are free parameters.

The top panel of Fig. 14 shows $\Delta M_{\rm BH,Q}/M_{\rm cold}$ versus V_{200c} for each best-fitting configuration, assuming $M_{\rm sat}/M_{\rm cen} = 1$. Two main trends emerge: configurations 'no low-z SMF, $f_{\rm Q}$', 'no HIMF', 'all', and 'no hi-z SMF, $f_{\rm Q}$' show consistently high accretion efficiency with weak dependence on halo mass. These also produce a peak-shaped SMBH mass function at $z = 0$ (Fig. 11). The shared flat scaling suggests a link between this accretion mode and the resulting SMBH distribution.

In contrast, other configurations show a strong halo mass dependence. Notably, while 'no HIMF' and 'no low-z SMF, $f_{\rm Q}$' exhibit similar SMBH growth scaling, their mass functions differ markedly – likely due to the latter's larger cold gas reservoirs, stemming from weaker SN (and AGN) feedback. The remaining models display comparable scaling trends, with strong V_{200c} dependence and similar SMBH mass functions. Although 'base$_{\rm H20}$' is ∼1 dex less efficient, its SMBH mass function at $z = 0$ remains comparable (within ∼0.1 dex), underscoring the non-linear nature of SMBH–galaxy co-evolution.

The second AGN process in `L-Galaxies`, radio mode, models feedback explicitly. Here, the SMBH accretes hot gas, and the injected energy suppresses further cooling. The accretion rate depends on the hot gas and SMBH masses, modulated by a free parameter, $k_{\rm AGN}$, which governs feedback efficiency. Best-fitting $k_{\rm AGN}$ values for each configuration are shown in Fig. 14 (bottom panel). Most configurations yield similar $k_{\rm AGN}$ values, except 'no low-z SMF, $f_{\rm Q}$', where it is ∼3 orders of magnitude lower. Despite rapid SMBH growth, the low AGN efficiency in this case fails to effectively quench star formation.

Configurations with the highest $k_{\rm AGN}$ also show peak-shaped SMBH mass functions and weak halo-mass dependence in cold gas accretion – highlighting the role of AGN feedback in quenching. However, the similar $k_{\rm AGN}$ values across most models, despite differing numbers of MQs (Fig. 9), suggest that other processes – such as strong SN feedback in 'base$_{\rm H20}$' and 'base$_{\rm L20}$' – also significantly contribute to quenching.

4.3 Model degeneracy

The best-fitting model for each configuration corresponds to the set of free parameters that yield the highest likelihood. However, given

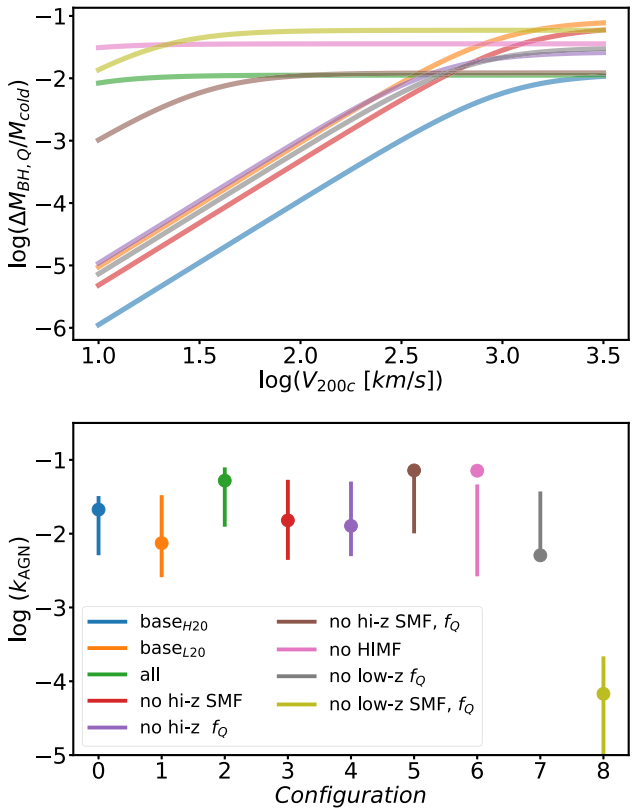

Figure 14. *Top:* The best-fitting scaling relation (equation 3) that describes the black hole growth from cold gas accretion in merger events (assuming $M_{\rm sat}/M_{\rm cen} = 1$, for the purpose of illustration) as a function of V_{200c} – a proxy of halo mass. Models that predict a peak-shaped SMBH mass function at $z = 0$, including the model that better matches the observational results ('no HIMF'), present an almost halo mass-independent cold gas accretion. The SMF-like SMBH mass function is driven by the strong dependence of the accreted cold gas by the SMBH on V_{200c}. *Bottom:* The best-fitting AGN efficiency parameter, $k_{\rm AGN}$, for every configuration. As in the top panel of Fig. 12, the error bar indicates the 16$^{\rm th}$ and 84$^{\rm th}$ percentiles of the final 2000 MCMC runs of the 96 chains. All models, except configuration 'no low-z SMF, $f_{\rm Q}$', have similar AGN efficiency in injecting energy to the hot gas, reducing the cooling rate. Thus, the AGN feedback across the models becomes almost entirely dependent on the SMBH mass, being underestimated due to the underpredictions of the SMBH mass function, except for configuration 'no HIMF'. Although configuration 'no low-z SMF, $f_{\rm Q}$' overpredicts the SMBH mass function, the AGN feedback, in this case, is less impactful due to the low AGN efficiency, $k_{\rm AGN}$.

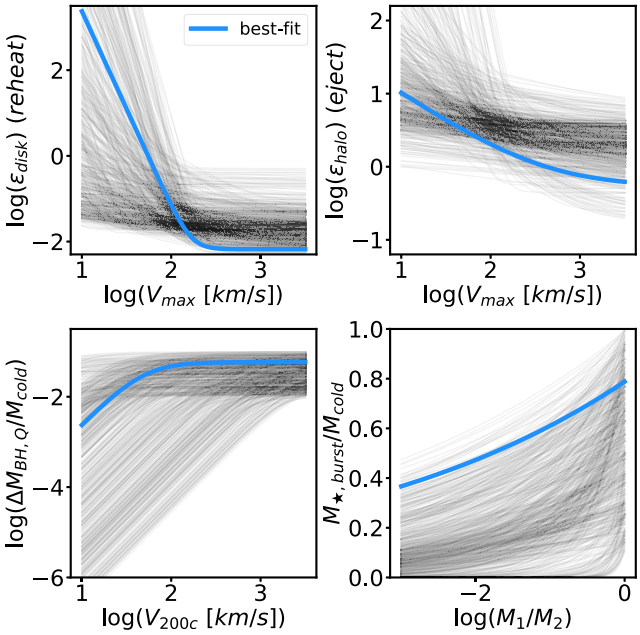

Figure 15. Physical scaling relations that describe the SN feedback, $\epsilon_{\rm disk}$ and $\epsilon_{\rm disk}$ (top panels; equation 2), SMBH growth (bottom left panel; equation 3), and stellar mass formed in merger-induced starbursts (bottom right; equation 1) for all sets of free parameters with total likelihood within 1 dex of the best-fitting (transparent black lines) for configuration 'no HIMF'. These results evidence the high level of model degeneracy, where the models with similar likelihood, in some cases, range many orders of magnitude. In most cases, the scaling relations from the configuration best-fitting do not occupy the most populated area drawn from similar likelihood models.

the complex hyperparameter space of `L-Galaxies`, degeneracies are expected. In this section, we analyse how the physical scaling relations presented in the previous section vary across models with similar likelihoods.

Fig. 15 illustrates four physical scaling relations for parameter sets with likelihoods within 1 dex of the maximum likelihood, based on the final 2000 MCMC runs across the 96 chains (see Section 3.5) of the 'no HIMF' configuration. The physical scalings shown in Fig. 15 are the efficiency of reheating, $\epsilon_{\rm disk}$ (top left panel; corresponding to the top panel of Fig. 13), the efficiency of ejecting gas, $\epsilon_{\rm halo}$ (top right panel; corresponding to the bottom panel of Fig. 13), SMBH growth from cold gas, $\Delta M_{\rm BH,Q}/M_{\rm cold}$ (bottom left panel; corresponding to the top panel of Fig. 14), and the fraction of cold gas converted into stars during merger-induced starbursts, $M_{\star,\rm burst}/M_{\rm cold}$ (bottom right panel; corresponding to the bottom panel of Fig. 12).

As shown in Fig. 15, the physical scaling relations for *good* likelihood models exhibit significant scatter, spanning approximately two dex across the ranges of $V_{\rm max}$, V_{200c}, and M_1/M_2. Also, the physical scalings derived from the best-fitting model of the 'no HIMF' configuration generally do not coincide with the most densely populated regions (representing models with similar scalings), except in the case of SMBH growth from cold gas. These findings highlight the high level of degeneracy in the galaxy formation model when calibrated against the observational constraints of the 'no HIMF' configuration (as is also the case for the other configurations). This suggests that the physical insights that can be gained from such fitted parameters may be limited.

Another manifestation of degeneracy identified in this work is that many configurations – calibrated using different sets of observational constraints and resulting in different best-fitting parameters – can none the less yield similar predictions for key observables, such as the SMFs, the fraction of quiescent galaxies, and sub-mm number counts, among others. This highlights a key limitation of traditional calibration methods, such as classical 'hand-tuning', which do not fully explore the high-dimensional parameter space of galaxy formation models. Such approaches risk converging on local minima and may overlook alternative viable solutions.

5 DISCUSSION

The tension between observations and theoretical models in simultaneously modelling DSFGs and high-redshift MQs remains

unresolved. Many models struggle to reproduce even one of these extreme populations, particularly the MQ population, as shown by Lagos et al. (2025). However, given the current approach to setting the free parameters of astrophysical processes – often done manually – it is not entirely evident that galaxy formation models fundamentally fail to reproduce these populations. In this work, we test the ability of the `L-Galaxies` SAM to address this tension by performing a robust calibration designed to reproduce observed populations of both DSFGs and MQs. We chose `L-Galaxies` for this study because its MCMC-based calibration framework allows us to explore different sets of observational constraints systematically. Our results yield a model that represents progress towards resolving this discrepancy, though some limitations remain. The key findings of this work are discussed in this section.

5.1 The impact of varying observational constraints

The adopted calibration framework was designed to explore how the calibrated model (defined by the set of tuned parameters) varies, given different sets of observational constraints (see Table 1). For instance, the configurations named 'base_{H20}' and 'base_{L20}' use essentially the same sets of constraints, differing only in the source of the data set used to constrain SMFs and massive quiescent galaxy fractions. The 'base_{H20}' configuration uses the same data as Henriques et al. (2020) (a compilation from the literature), whereas 'base_{L20}' was calibrated with the SMFs and quiescent fractions (f_{Q}) from Leja et al. (2020) and Leja et al. (2022), respectively. The key difference between these data sets lies in the SMFs, with Leja et al. (2020) predicting systematically higher number densities across the stellar mass range, particularly at the massive end. As expected, this leads to differences in the SMFs predicted by the calibrated models (Fig. 3) and significantly impacts the predicted CSFRD (Fig. 10), with a discrepancy of approximately 0.5 dex at low redshift. While both configurations reproduce the quiescent population reasonably well and in a similar manner (Fig. 4), they severely underpredict the sub-mm number counts (Fig. 7). In these configurations, number density of SMGs (n_{SMG}) was not included as an observational constraint. Our results motivate the incorporation of n_{SMG} as a constraint.

In the remaining configurations, we address this issue by including sub-mm number densities within a systematic MCMC calibration framework. This work represents the first time such a systematic and extensive exploration of the hyperparameter space in a SAM has included SMGs as a constraint. We calibrated these configurations using the number density of bright ($S_{870} \geq 5.2$ mJy) SMGs at $z = 2.8$ to capture the observed sub-mm number counts, while also fitting for SMFs and f_{Q} across different redshifts. Among all configurations, only the 'no hi-z SMF' successfully reproduces the observed S_{870} number counts, whereas the others tend to slightly underpredict them. On the other hand, omitting low-redshift constraints (configuration 'no low-z SMF, f_{Q}') leads to an *overprediction* of the sub-mm number counts. The main differences in the predictions arise in the quiescent population, with all configurations underrepresenting these galaxies to some degree. When f_{Q} at low redshift is not included as an observational constraint, the best-fitting model significantly underestimates the number of quiescent galaxies (configurations 'no low-z SMF, f_{Q}' and 'no low-z f_{Q}'). Conversely, when the high-redshift SMF and f_{Q} are not used as constraints, the resulting models are similar (configurations 'no hi-z SMF', 'no hi-z f_{Q}', and 'no hi-z SMF, f_{Q}').

Interestingly, when the high-redshift SMF is not used as a constraint (configuration 'no hi-z SMF'), the best-fitting model most accurately reproduces the observed number counts but performs the worst in predicting the number density of MQs at high redshift – even more so than when the high-redshift f_{Q} is not included. Additionally, we find that when both SMF and f_{Q} are used simultaneously as constraints (configuration 'all'), the predicted galaxy properties remain statistically similar to those obtained when neither constraint is applied (configuration 'no hi-z SMF, f_{Q}').

Finally, although the HIMF (at $z = 0$) was not included as a constraint in the 'no HIMF' configuration, the model still successfully reproduces it. Among the best-performing models, this particular configuration is slightly weaker in predicting both the submillimetre number counts and the number density of MQs at high redshift. Nevertheless, it still achieves a reasonable agreement for both, making it the most successful model overall in simultaneously reproducing DSFGs and MQs.

5.2 Matching the sub-mm number counts

Matching the submillimetre number counts without invoking an IMF modification remains a challenge for many of the most widely used cosmological simulations, such as `EAGLE` (Cowley et al. 2019), `IllustrisTNG` (Hayward et al. 2021), and `L-Galaxies` (Araya-Araya et al. 2024), among others. Only a few simulations have been able to closely reproduce the observed number counts, including `Illustris` (Hayward et al. 2021), `SIMBA` (Lovell et al. 2021), `FLAMINGO` (Kumar et al. 2025), and the `SHARK` (v1.0) (Lagos et al. 2019) SAM.

In this study, we demonstrated that incorporating the number density of bright SMGs at a single redshift as a constraint significantly improves model predictions of the sub-mm number counts.[5] Indeed, all configurations that included this constraint successfully matched (or closely matched) the observational data, across an order of magnitude in sub-mm flux density. Among these, configuration 'no hi-z SMF' provides the best predictions for the sub-mm number counts, even at the bright end. These new calibrated models present an opportunity for future theoretical studies of bright SMGs.

Here, we explore the key characteristics that favour good matches to the observed SMG population, based on the best-fitting parameters presented in Section 4.2. The configuration that best-reproduces the observed number counts ('no hi-z SMF') has similar best-fitting parameters to most of the other models, for parameters relating to star formation (both secular and merger-driven), SMBH growth, and AGN feedback efficiency. The main difference lies in the SN feedback model, specifically in the scaling relation for (re)heating gas (Fig. 13, top panel). The two configurations that predict the highest number of SMGs ('no hi-z SMF' and 'no low-z SMF, f_Q') exhibit similar functional forms for this efficiency, which plays a crucial role in suppressing star formation. The best-fitting scaling relations indicate a strong inverse dependence of the (re)heating efficiency on V_{max} (proxy of dark matter halo mass). For these configurations, the efficiency is highest in low-mass systems and lowest in high-mass haloes. As a result, more cold gas remains available for star formation in massive galaxies (the mass range of most SMGs) compared to the

[5]It is important to note that this does not imply models calibrated without using the number density of SMGs cannot reproduce the sub-mm number counts – as demonstrated by models such as `SHARK` (v1.0) and `SIMBA`. This is possible because multiple combinations of parameters can yield predictions consistent with the observed sub-mm number counts. due to the degenerate space.

other models, whether this star formation proceeds through secular or merger-induced star formation.

Given the high star formation efficiencies involved in these two calibrated models for high stellar mass galaxies, AGN feedback would be required for effective quenching. However, the 'no hi-z SMF' configuration actually *underpredicts* the SMBH mass function at $z = 0$ (Fig. 11), indicating that SMBH growth in this model is lower than required by observations. Consequently, AGN feedback is less effective than required to reproduce the quiescent population, and number densities of MQs in this model are significantly lower than observed. On the other hand, the 'no low-z SMF, $f_{\rm Q}$' configuration exhibits a similar SN feedback scaling relation to 'no hi-z SMF' but critically overpredicts the sub-mm number counts. In this case, the AGN feedback efficiency (radio mode) is approximately two orders of magnitude lower. As a result, there is no strong regulatory mechanism to suppress star formation in intermediate and massive galaxies. These cases highlight the persistent difficulty of reproducing both SMG and MQ populations simultaneously.

5.3 Reproducing the massive quiescent population at high-z

Galaxy formation models fail to reproduce the high number density of MQs at $z \gtrsim 3$ from observations, as shown in Merlin et al. (2019), Szpila et al. (2025), Vani et al. (2025), and Lagos et al. (2025), among others; this is of considerable interest given the increasing numbers of such galaxies being characterized by *JWST*. Indeed, this issue also happens for `L-Galaxies`. However, it is important to note that there are large discrepancies between current observational estimates of the number density, with limits ranging ~ 2 dex (Valentino et al. 2023). This is, at least in part, due to the different methods to estimate galaxy properties, selection criteria, and the available data used in those works. Cosmic variance and the difficulties of estimating number densities from extremely small samples also pose a challenge for these works. For instance, Alberts et al. (2024) obtained comparable number densities of MQs to Carnall et al. (2023) and Valentino et al. (2023), despite adopting a ~ 1 dex lower stellar mass cut and explicitly studying an overdense region.

Most of our re-calibrated models underestimate both the fraction and number density of quiescent galaxies (Figs 4 and 9, respectively), especially at high redshift.[6] Among our configurations, only three are in reasonable agreement with the lower limits of the observed number density of MQs, comparable with some models presented by Lagos et al. (2025). These configurations are '$\rm base_{H20}$', '$\rm base_{L20}$', and 'no HIMF', the first two of which provide the worst matches to the sub-mm number counts. In contrast, the configuration 'no HIMF' closely matches the sub-mm number counts within the uncertainties (similar to the results of Lovell et al. 2021, who studied SMGs in the `SIMBA` simulation) and hence provides a promising avenue for future work.

We identify two different combinations of physical mechanisms that act to quench galaxies in these three models. First, configurations '$\rm base_{H20}$' and '$\rm base_{L20}$' present a similar shape for the scaling relations that set the efficiency of (re)heating gas. These two models present the highest efficiency in massive haloes ($V_{\rm max} \gtrsim 110\,{\rm km\,s^{-1}}$) compared to the rest of the configurations. Configuration '$\rm base_{H20}$' has the highest efficiency and, consequently, predicts the highest number density of MQs. As discussed in the last subsection, models that favour the production of DSFGs require a lower (re)heating efficiency for high-mass systems, so it is natural for these two configurations to critically underpredict the sub-mm number counts. The SMBH mass functions at $z = 0$ for these configurations show that the SMBHs grew less than the observed; in these configurations, the AGN feedback may be too weak, requiring stronger SN feedback to match the fraction of quiescent galaxies (which is used as a constraint).

On the other hand, the best-fitting model of configuration 'no HIMF' closely matches (still slightly underpredicting) the observationally derived SMBH mass function at $z = 0$. For that configuration, an 'in-between' scaling relation for (re)heating gas is obtained, yielding reasonably good agreement with the observed sub-mm number counts. These results suggest that the SMBH mass function (and ideally, its evolution) could be a key observable to calibrate galaxy formation models, as it would help in constraining AGN feedback, breaking the degeneracies seen among the main quenching mechanisms (Fig. 15).

5.4 Limitations and caveats

Although we found a model that matches the sub-mm number counts reasonably well and simultaneously agrees with the observed lower limits for the number density of MQs (configuration 'no HIMF') using the Henriques et al. (2020) version of the `L-Galaxies` SAM, it still presents limitations. For instance, all configurations struggle to capture the massive end of the SMF at high redshift (Fig. 3), even for 'no low-z SMF, $f_{\rm Q}$', which was specifically designed to match it (using the SMF at $z = 2.8$ as a constraint). This trend is also observed in the `SHARK` SAM (Lagos et al. 2024). Moreover, all best-fitting models overpredict the CSFRD by ~ 0.5 dex compared to observational data, except for configuration '$\rm base_{H20}$'. This suggests that the overprediction of the CSFRD is likely due to the Leja et al. (2020) SMFs, which were used to calibrate these models. Consequently, these results highlight the importance and impact of the observational data used to constrain the models. As mentioned earlier, observational works estimate galaxy properties using a specific method or technique, and these estimates can vary depending on the approach taken. This variability makes the comparison between observations and simulations challenging. A possible solution could be to select observationally analogous simulated galaxy populations based on forward-modelled observer-frame magnitudes/fluxes, thus avoiding the complex dependence on the techniques used to select galaxies and estimate their properties (see e.g. Cochrane et al. 2024). However, this approach involves other assumptions in the forward modelling.

In this work, we assumed that 40 per cent of the metals in the cold gas reservoirs of galaxies are in the form of dust, as the version of `L-Galaxies` by Henriques et al. (2020) does not track the evolution of dust. More recently, a detailed model for dust production and destruction has been incorporated into the 2020 version of `L-Galaxies` (Yates et al. 2024), which could be employed to refine predictions of sub-mm emission within the simulation. Nevertheless, our assumption is unlikely to significantly affect our results, as adopting a constant dust-to-metal ratio is a reasonable approximation for massive, metal-rich galaxies such as those comprising the SMG population (Rémy-Ruyer et al. 2014).

Another crucial finding is the high level of degeneracy (Fig. 15) observed. Although some sets of free parameters yield similar likelihoods, they may represent entirely different treatments of the

[6] One possible solution is more accurate modelling of the properties of orphan galaxies – galaxies whose host dark matter haloes have been fully accreted or disrupted by more massive systems – in SAMs, as demonstrated by Harrold et al. (2024) for the Henriques et al. (2015) version of `L-Galaxies`. This approach significantly improves the consistency with the observed quiescent galaxy SMF.

astrophysical mechanisms. This arises from the hyperparameter space of `L-Galaxies` (and galaxy formation models in general), which is composed of 15 free parameters. As a result, it is expected that the algorithm will identify multiple reasonable 'good' solutions. In this context, robust calibration should be considered as an essential step for galaxy formation models.

6 SUMMARY

In this work, we used the MCMC mode of the `L-Galaxies` SAM to robustly calibrate its free parameters. Our main goal was to investigate whether this approach can address the long-standing tension between modelled DSFGs (also known as submillimetre galaxies, SMGs) and MQs at high redshifts. To address this, we implemented nine sets of observational constraints, including the widely used SMFs, the fraction of quiescent galaxies ($f_{\rm Q}$), and the HIMF, as well as including for the first time the number density of SMGs ($n_{\rm SMG}$). These nine combinations of observational constraints, referred to as configurations (see Table 1), produced nine distinct models. Our main predictions and interpretations are as follows:

(i) The SMF is a key observable used to calibrate galaxy formation models. Here, we updated the SMFs used to calibrate the Henriques et al. (2020) version of `L-Galaxies` to the Leja et al. (2020) results. Our predicted SMFs (Fig. 3) are consistent with the observational data at low redshift. However, the models struggle to reproduce the massive end at higher redshifts, even when high-redshift SMFs are used as constraints.

(ii) Similarly, despite incorporating $f_{\rm Q}$ as a function of stellar mass at high redshifts as a constraint, the models fail to reproduce this observable (see Fig. 4). When $f_{\rm Q}$ is not used as a constraint at low redshifts, the quiescent population is significantly underrepresented. As expected, configurations excluding $n_{\rm SMG}$ as a constraint exhibit better consistency with the observed $f_{\rm Q}$.

(iii) Models calibrated with $n_{\rm SMG}$ accurately reproduce the observed S_{870} number counts (Fig. 7). In contrast, models that do not include this constraint critically underpredict this observable, emphasizing the importance of using $n_{\rm SMG}$ in the calibration process, to capture the SMG population.

(iv) Models that align better with S_{870} number counts tend to underpredict the number density of MQs at high redshift (Fig. 9), highlighting the persistent tension between these two populations. However, one configuration –calibrated using SMFs and $f_{\rm Q}$ at $z = 0.4$ and 2.8, and $n_{\rm SMG}$ at $z = 2.8$, while excluding the HIMF ('no HIMF') – achieves reasonable agreement with both populations.

(v) Most models predict an elevated cosmic star formation density compared to observational data (Fig. 10). This discrepancy appears to stem from the use of Leja et al. (2020) SMFs for calibration, rather than from the inclusion of $n_{\rm SMG}$.

(vi) Predicted SMBH mass functions at $z = 0$ (Fig. 11) show that most models underpredict the number density compared to observations. This leads to weaker AGN feedback, a critical mechanism for regulating star formation in massive galaxies. The 'no HIMF' configuration presents the best agreement with observational data.

(vii) Analysing star formation across configurations (Fig. 12), we found that models calibrated with $n_{\rm SMG}$ exhibit reduced star formation efficiency from secular processes, compensated for by increased merger-induced starbursts. This merger-driven star formation may deplete cold gas, limiting subsequent secular star formation in descendant galaxies.

(viii) Examining SN feedback (Fig. 13) reveals that models that better reproduce $f_{\rm Q}$ exhibit higher efficiencies in heating cold gas within massive haloes, likely driving their predictions for $f_{\rm Q}$. However, gas ejection efficiencies (outflows) for massive galaxies are similar across models, with significant variations only at the low-mass end.

(ix) For black hole growth and AGN feedback (Fig. 14), the fitted models show two distinct trends in SMBH growth via merger-driven cold gas accretion. One trend shows accretion nearly independent of halo mass, resulting in a peaked SMBH mass function at $z = 0$. The other trend exhibits a strong halo mass dependency, producing an SMBH mass function similar to the SMF. The former trend aligns better with observations. Insufficient SMBH masses imply underpowered AGN feedback, except in the 'no HIMF' configuration.

(x) Finally, we assessed the degeneracy in the 'no HIMF' configuration (Fig. 15), which provides a reasonable prediction of submillimetre number counts and is consistent with the lower limits of the number density of high-redshift MQs. Our robust analysis reveals that models with different underlying physics can yield similar likelihood values, owing to significant degeneracies within the hyperparameter space of `L-Galaxies`. This highlights that the commonly used observational constraints are insufficient to uniquely determine a single preferred physical model. In this context, additional, previously unconsidered constraints – such as SMBH mass functions – may prove valuable in breaking these degeneracies.

Our results provide a comprehensive analysis of calibration outcomes for `L-Galaxies` and emphasize the importance of robust calibration techniques in exploring the hyperparameter space of galaxy formation models. Additionally, we identify a promising model that represents a step forward in resolving the tension between modelled SMGs and high-redshift massive quiescent populations.

ACKNOWLEDGEMENTS

This work was initiated during the CCA Pre-doctoral Program. PA-A thanks the Coordenaçõ de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES), for supporting his PhD scholarship (project 88882.332909/2020-01). RKC was funded by support for programme #02321, provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–03127. RKC is grateful for support from the Leverhulme Trust via the Leverhulme Early Career Fellowship. The Flatiron Institute is supported by the Simons Foundation. CCH thanks Neal Katz for insightful discussions about SAMs. LSJ acknowledges the support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; project 308994/2021-3) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; project 2011/51680-6). MCV acknowledges the FAPESP for supporting his PhD (2021/06590-0).

DATA AVAILABILITY

The full source code for the original L-GALAXIES 2020 version is publicly available via GitHub, with installation and running instructions provided on the L-Galaxies website. Additional simulation data products derived by the authors and presented here can be obtained upon request by sending an email to paraya-araya@usp.br.

REFERENCES

Alberts S. et al., 2024, ApJ, 975, 85
Angulo R. E., Hilbert S., 2015, MNRAS, 448, 364

Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Araya-Araya P., Vicentin M. C., Sodré Laerte J., Overzier R. A., Cuevas H., 2021, MNRAS, 504, 5054
Araya-Araya P. et al., 2024, ApJ, 977, 204
Ayromlou M., Kauffmann G., Yates R. M., Nelson D., White S. D. M., 2021, MNRAS, 505, 492
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Baldry I. K. et al., 2012, MNRAS, 421, 621
Barger A. J., Cowie L. L., Sanders D. B., Fulton E., Taniguchi Y., Sato Y., Kawara K., Okuda H., 1998, Nature, 394, 248
Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton Univ. Press, Princeton, NJ
Blain A. W., Chapman S. C., Smail I., Ivison R., 2004, ApJ, 611, 725
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calvi R., Castignani G., Dannerbauer H., 2023, A&A, 678, A15
Carnall A. C. et al., 2020, MNRAS, 496, 695
Carnall A. C. et al., 2023, MNRAS, 520, 3974
Casey C. M., 2016, ApJ, 824, 36
Casey C. M., Narayanan D., Cooray A., 2014, Phys. Rep., 541, 45
Cecchi R., Bolzonella M., Cimatti A., Girelli G., 2019, ApJ, 880, L14
Chabrier G., 2003, PASP, 115, 763
Chapman S. C., Lewis G. F., Scott D., Richards E., Borys C., Steidel C. C., Adelberger K. L., Shapley A. E., 2001, ApJ, 548, L17
Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, ApJ, 622, 772
Charlot S., Bruzual G., 2007, in press
Chen C.-C. et al., 2016, ApJ, 831, 91
Cochrane R. K. et al., 2021, MNRAS, 503, 2622
Cochrane R. K., Hayward C. C., Anglés-Alcázar D., Somerville R. S., 2023a, MNRAS, 518, 5522
Cochrane R. K. et al., 2023b, MNRAS, 523, 6082
Cochrane R. K., Anglés-Alcázar D., Cullen F., Hayward C. C., 2024, ApJ, 961, 37
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Cooper O. R., Casey C. M., Zavala J. A., Champagne J. B., da Cunha E., Long A. S., Spilker J. S., Staguhn J., 2022, ApJ, 930, 32
Cowie L. L., Barger A. J., Wang W. H., Williams J. P., 2009, ApJ, 697, L122
Cowley W. I., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., Lagos C. d. P., 2019, MNRAS, 487, 3082
Croton D. J. et al., 2006, MNRAS, 365, 11
Daddi E. et al., 2009, ApJ, 694, 1517
Daddi E. et al., 2010, ApJ, 713, 686
Dannerbauer H. et al., 2014, A&A, 570, A55
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Davidzon I. et al., 2017, A&A, 605, A70
Downes D. et al., 1999, A&A, 347, 809
Dudzevičiūtė U. et al., 2020, MNRAS, 494, 3828
Dunlop J. S. et al., 2004, MNRAS, 350, 769
Dunlop J. S. et al., 2017, MNRAS, 466, 861
Eales S., Lilly S., Gear W., Dunne L., Bond J. R., Hammer F., Le Fèvre O., Crampton D., 1999, ApJ, 515, 518
Fontanot F., Monaco P., Silva L., Grazian A., 2007, MNRAS, 382, 903
Fu J. et al., 2013, MNRAS, 434, 1531
García-Vergara C., Hodge J., Hennawi J. F., Weiss A., Wardlow J., Myers A. D., Hickox R., 2020, ApJ, 904, 2
Geach J. E. et al., 2017, MNRAS, 465, 1789
Genel S. et al., 2014, MNRAS, 445, 175
Girelli G., Bolzonella M., Cimatti A., 2019, A&A, 632, A80
Gouin C., Aghanim N., Dole H., Polletta M., Park C., 2022, A&A, 664, A155
Gould K. M. L. et al., 2023, AJ, 165, 248
Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710
Harrold J. E., Almaini O., Pearce F. R., Yates R. M., 2024, MNRAS, 532, L61
Haynes M. P. et al., 2011, AJ, 142, 170
Hayward C. C., Kereš D., Jonsson P., Narayanan D., Cox T. J., Hernquist L., 2011, ApJ, 743, 159
Hayward C. C., Narayanan D., Kereš D., Jonsson P., Hopkins P. F., Cox T. J., Hernquist L., 2013, MNRAS, 428, 2529
Hayward C. C. et al., 2021, MNRAS, 502, 2922
Henriques B. M. B., Thomas P. A., 2010, MNRAS, 403, 768
Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, 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. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2020, MNRAS, 491, 5795
Herard-Demanche T. et al., 2025, MNRAS, 537, 788
Hill R., Polletta M. d. C., Bethermin M., Dole H., Kneissl R., Scott D., 2025, A&A, 698, A204
Hodge J. A., da Cunha E., 2020, R. Soc. Open Sci., 7, 200556
Hughes D. H. et al., 1998, Nature, 394, 241
Ilbert O. et al., 2013, A&A, 556, A55
Izquierdo-Villalba D., Bonoli S., Dotti M., Sesana A., Rosas-Guevara Y., Spinoso D., 2020, MNRAS, 495, 4681
Izquierdo-Villalba D., Sesana A., Bonoli S., Colpi M., 2022, MNRAS, 509, 3488
Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2
Kimmig L. C., Remus R.-S., Seidel B., Valenzuela L. M., Dolag K., Burkert A., 2025, ApJ, 979, 15
Kumar A. et al., 2025, A&A, 698, A236
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573
Lagos C. d. P. et al., 2019, MNRAS, 489, 4196
Lagos C. d. P. et al., 2024, MNRAS, 531, 3551
Lagos C. d. P. et al., 2025, MNRAS, 536, 2324
Laigle C. et al., 2016, ApJS, 224, 24
Leja J., van Dokkum P. G., Franx M., Whitaker K. E., 2015, ApJ, 798, 115
Leja J., Johnson B. D., Conroy C., van Dokkum P. G., Byler N., 2017, ApJ, 837, 170
Leja J., Carnall A. C., Johnson B. D., Conroy C., Speagle J. S., 2019a, ApJ, 876, 3
Leja J. et al., 2019b, ApJ, 877, 140
Leja J., Speagle J. S., Johnson B. D., Conroy C., van Dokkum P., Franx M., 2020, ApJ, 893, 111
Leja J. et al., 2022, ApJ, 936, 165
Leslie S. K. et al., 2020, ApJ, 899, 58
Li C., White S. D. M., 2009, MNRAS, 398, 2177
Lim C.-F. et al., 2020, ApJ, 895, 104
Long A. S. et al., 2024, preprint (arXiv:2408.14546)
Lovell C. C., Geach J. E., Davé R., Narayanan D., Li Q., 2021, MNRAS, 502, 772
Lu Y., Mo H. J., Weinberg M. D., Katz N., 2011, MNRAS, 416, 1949
Lu Y. et al., 2014, ApJ, 795, 123
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Manning S. M. et al., 2022, ApJ, 925, 23
Maraston C., 2005, MNRAS, 362, 799
Marinacci F. et al., 2018, MNRAS, 480, 5113
Marrone D. P. et al., 2018, Nature, 553, 51
McAlpine S. et al., 2016, Astron. Comput., 15, 72
McAlpine S. et al., 2019, MNRAS, 488, 2440
Merlin E. et al., 2019, MNRAS, 490, 3309
Michałowski M. J. et al., 2017, MNRAS, 469, 492
Miller T. B. et al., 2018, Nature, 556, 469

Murphy G. G., Yates R. M., Mohamed S. S., 2022, MNRAS, 510, 1945
Muzzin A. et al., 2013, ApJ, 777, 18
Nanayakkara T. et al., 2024, Sci. Rep., 14, 3724
Narayanan D. et al., 2021, ApJS, 252, 12
Neri R., Downes D., Cox P., Walter F., 2014, A&A, 562, A35
Oteo I. et al., 2018, ApJ, 856, 72
Planck Collaboration et al., 2014, A&A, 571, A16
Rémy-Ruyer A. et al., 2014, A&A, 563, A31
Richards E. A., 1999, ApJ, 513, L9
Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651
Safarzadeh M., Lu Y., Hayward C. C., 2017, MNRAS, 472, 2462
Schaye J. et al., 2023, MNRAS, 526, 4978
Schreiber C. et al., 2018, A&A, 618, A85
Shahidi A. et al., 2020, ApJ, 897, 44
Shankar F. et al., 2020, Nat. Astron., 4, 282
Simpson J. M. et al., 2014, ApJ, 788, 125
Simpson J. M. et al., 2017, ApJ, 839, 58
Skelton R. E. et al., 2014, ApJS, 214, 24
Smail I., Ivison R. J., Blain A. W., 1997, ApJ, 490, L5
Smith D. J. B., Hayward C. C., Jarvis M. J., Simpson C., 2017, MNRAS, 471, 2453
Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504
Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, MNRAS, 423, 1992
Spinoso D., Bonoli S., Valiante R., Schneider R., Izquierdo-Villalba D., 2023, MNRAS, 518, 4672
Springel V., 2005, MNRAS, 364, 1105
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2018, MNRAS, 475, 676
Stach S. M. et al., 2018, ApJ, 860, 161
Stach S. M. et al., 2019, MNRAS, 487, 4648
Stach S. M. et al., 2021, MNRAS, 504, 172
Straatman C. M. S. et al., 2014, ApJ, 783, L14
Szpila J., Davé R., Rennehan D., Cui W., Hough R. T., 2025, MNRAS, 537, 1849
Tacconi L. J. et al., 2010, Nature, 463, 781
Toft S. et al., 2014, ApJ, 782, 68
Tomczak A. R. et al., 2014, ApJ, 783, 85
Tomczak A. R. et al., 2016, ApJ, 817, 118
Trayford J. W. et al., 2017, MNRAS, 470, 771
Tucci M., Volonteri M., 2017, A&A, 600, A64
Valentino F. et al., 2020, ApJ, 889, 93
Valentino F. et al., 2023, ApJ, 947, 20
Vani A., Ayromlou M., Kauffmann G., Springel V., 2025, MNRAS, 536, 777
Vika M., Driver S. P., Graham A. W., Liske J., 2009, MNRAS, 400, 1451
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Wagg J., Hughes D. H., Aretxaga I., Chapin E. L., Dunlop J. S., Gaztañaga E., Devlin M., 2007, MNRAS, 375, 745
Walter F. et al., 2012, Nature, 486, 233
Wang G. C. P. et al., 2021, MNRAS, 508, 3754
Weaver J. R. et al., 2023, A&A, 677, A184
Yates R. M., Henriques B., Thomas P. A., Kauffmann G., Johansson J., White S. D. M., 2013, MNRAS, 435, 3500
Yates R. M., Henriques B. M. B., Fu J., Kauffmann G., Thomas P. A., Guo Q., White S. D. M., Schady P., 2021, MNRAS, 503, 4474
Yates R. M., Hendriks D., Vijayan A. P., Izzard R. G., Thomas P. A., Das P., 2024, MNRAS, 527, 6292
Zavala J. A. et al., 2021, ApJ, 909, 165
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30
da Cunha E. et al., 2015, ApJ, 806, 110

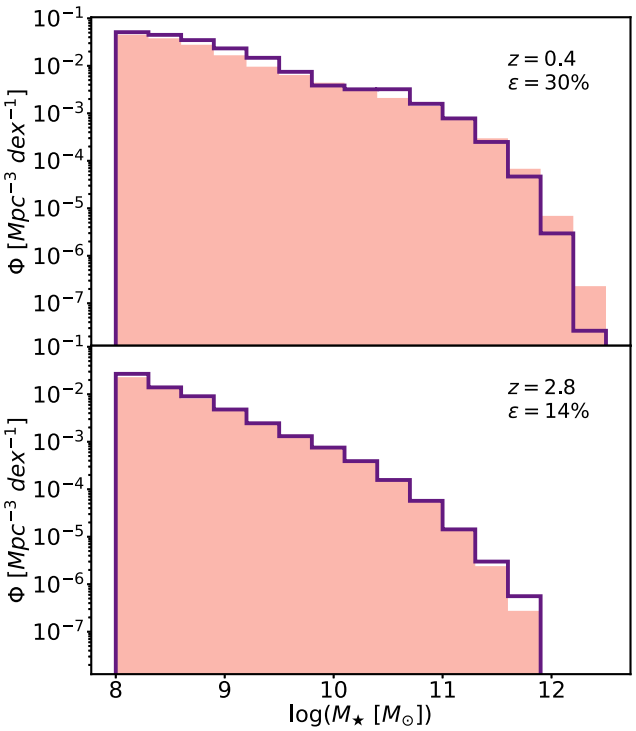

Figure A1. The SMF obtained from the fiducial model (filled histogram) compared to that derived from the sample of merger trees (solid line) at $z = 0.4$ (top panel) and $z = 2.8$ (bottom panel). The ϵ value denotes the average relative error between the SMFs (sampled and full volume).

APPENDIX A: OBSERVABLES FROM THE SAMPLE OF MERGER TREES

In Section 3.2, we present our method for obtaining a sample of merger trees, the predictions of which must closely represent those of the full simulation volume. The predictions from the sampled merger trees are derived by weighting a given observable according to the fraction of similar haloes not included in the sample. Here, we compare the full-volume and sampled SMFs, as well as the fractions of quiescent galaxies as a function of stellar mass, at $z = 0.4$ and $z = 2.8$, shown in Figs A1 and A2, respectively. To quantify the similarities between the sample and full-volume predictions, we estimate the average relative error, ϵ, as defined by equation (A1):

$$\epsilon = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\Phi_{\mathrm{sampled},i} - \Phi_{\mathrm{full},i}\right|}{\Phi_{\mathrm{full},i}}, \tag{A1}$$

where $\Phi_{\mathrm{sampled},i}$ and $\Phi_{\mathrm{full},i}$ represent a given prediction (e.g. the SMF or f_{Q}) in the i-th bin – such as a stellar mass bin – from the sample of merger trees and the full volume, respectively. N is the total number of bins. As can be seen from these figures, the SMF constructed from the sample of merger trees matches that obtained from the full volume well.

As discussed in Section 3.2, the merger tree sample was also selected to be representative of the number density of SMGs, another key observable in this work. Although the sampled set comprises only 430 merger trees out of a total of approximately 20 million, we obtain comparable predictions for both observables used to calibrate the model. At $z = 2.8$, the fiducial model predicts an SMG number density ($S_{870} \geq 5.2$ mJy) of $n_{\mathrm{SMG}} = 9 \times 10^{-6}\,\mathrm{Mpc}^{-3}$, while the sampled merger tree set yields $n_{\mathrm{SMG}} = 8.8 \times 10^{-6}\,\mathrm{Mpc}^{-3}$.

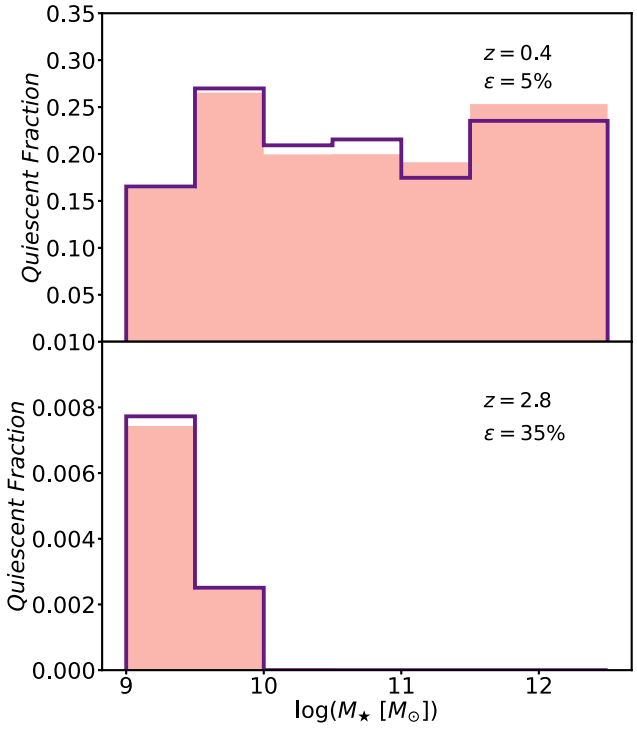

Figure A2. The fraction of quiescent galaxies as a function of stellar mass (f_Q) obtained from the fiducial model (filled histogram) compared to that derived from the sample of merger trees (solid line) at $z = 0.4$ (top panel) and $z = 2.8$ (bottom panel). The ϵ value denotes the average relative error between both f_Q values (sampled and full volume).

APPENDIX B: BEST-FITTING PARAMETERS FOR EACH CALIBRATION CONFIGURATION

We present the best-fitting values for the 15 free parameters across the nine configurations listed in Table 1 in Fig. B1. The error bars in Fig. B1 correspond to the 16th and 84th percentiles from the final 2000 MCMC samples across 96 chains (see Section 3.5).

The parameters α_{SF}, $\alpha_{SF,burst}$, and $\beta_{SF,burst}$ are associated with secular star formation (as shown in the top panel of Fig. 12) and merger-induced starbursts (free parameters of equation 1; bottom panel of Fig. 12), respectively. The AGN efficiency parameter, k_{AGN}, is the same as presented in the bottom panel of Fig. 14, while f_{BH} and V_{BH} are the free parameters governing the black hole growth scaling relation (equation 3; top panel of Fig. 14).

The parameters η_{reheat}, V_{reheat}, and β_{reheat} define the scaling relation in equation (2) describing gas (re)heating due to SN feedback (Fig. 13, top panel). Similarly, the parameters η_{eject}, V_{eject}, and β_{eject}, which govern the ejection of gas from the hot gas atmosphere, share the same scaling form as the (re)heating process (Fig. 13; bottom panel).

The remaining parameters, not discussed in detail in this paper, include γ_{reinc}, which sets the reincorporation timescale of ejected gas; $\alpha_{friction}$, a correction factor for dynamical friction based on the Binney & Tremaine (1987) formula; and finally, M_{RP}, the halo mass threshold above which ram pressure stripping is considered.

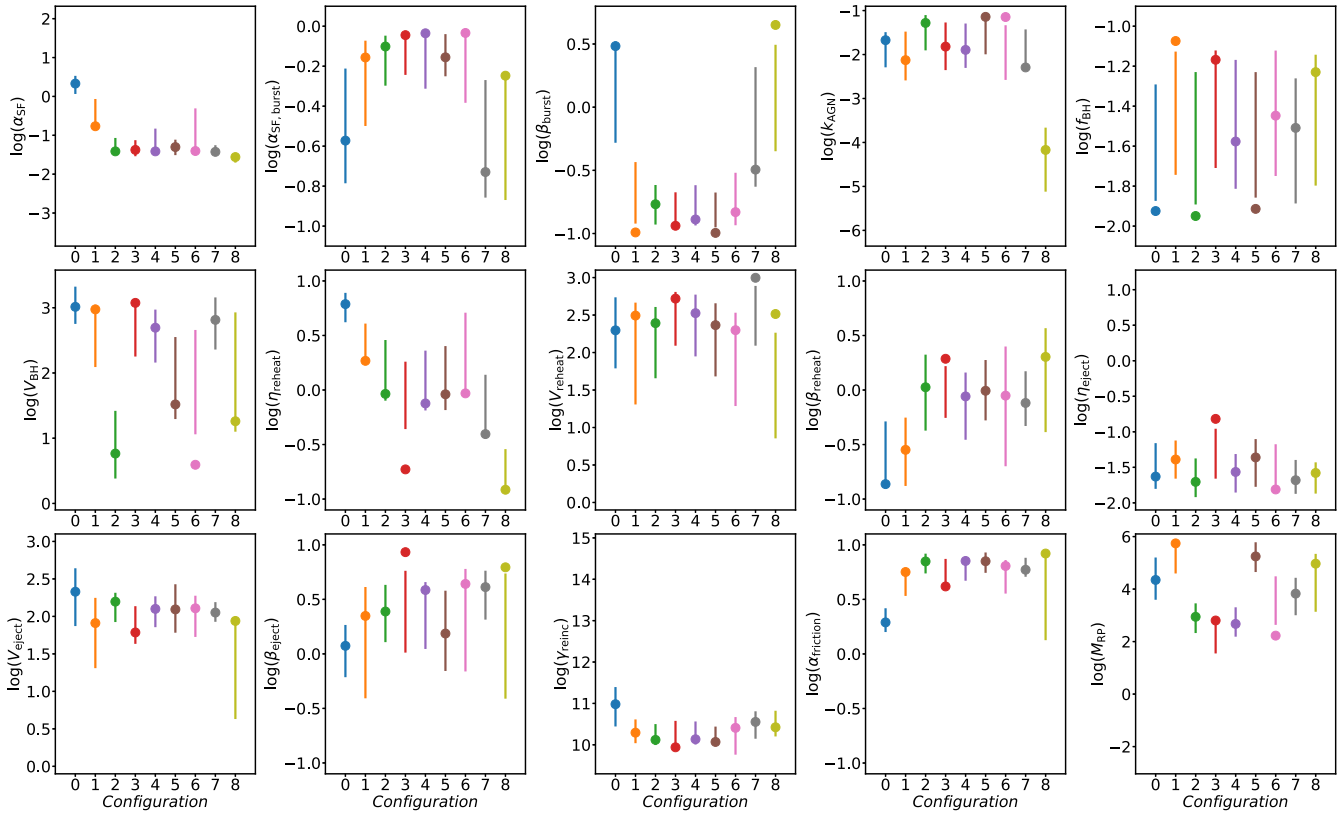

Figure B1. The best-fitting parameters obtained for each configuration are listed in Table 1. The error bars represent the 16th and 84th percentiles of the final 2000 MCMC runs across 96 chains. The parameters α_{SF} (first panel in the top row) and k_{AGN} (fourth panel in the top row) are shown in the top panel of Fig. 12 and the bottom panel of Fig. 14, respectively.

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