Withstanding the cold: energy feedback in simulations of galaxies that include a cold interstellar medium

Chaikin, E.

5 COLIBRE: calibrating subgrid feedback in cosmological simulations that include a cold gas phase

Authors: **Evgenii Chaikin** et al.
To be submitted to Monthly Notices of the Royal Astronomical Society

Abstract

We present the calibration of the new subgrid model COLIBRE for hydrodynamical simulations of galaxy formation. COLIBRE builds upon the EAGLE model with significant improvements on various fronts. In particular, COLIBRE includes the multiphase structure of the interstellar medium, gas cooling in non-equilibrium, a live dust model coupled to the chemistry, and more sophisticated prescriptions for star formation and feedback from star formation and active galactic nuclei (AGN). To calibrate COLIBRE, we employ the new astrophysical code SWIFT. We run Latin hypercubes of $\sim 10^2$ simulations that vary up to four subgrid parameters in cosmological volumes of (50 comoving Mpc)^3 at a gas (dark-matter) mass resolution of $1.47\times10^7\,\mathrm{M}_\odot$ ($1.94\times10^7\,\mathrm{M}_\odot$). We train Gaussian process emulators on these runs to predict the $z = 0$ galaxy stellar mass function (GSMF) and size-stellar mass relation (SSM) as functions of the model parameters, which we then fit to observations. The trained emulators not only provide the best-fitting parameter values but also enable us to investigate how different aspects of the COLIBRE supernova (SN) and AGN feedback affect the predicted GSMF and SSM. In particular, we demonstrate that while the $z = 0$ observed GSMF and SSM can be matched separately with a relatively simple SN feedback model, reproducing both necessitates a more sophisticated prescription for SN feedback. We show that the COLIBRE model not only reproduces the calibration observables, but also matches a rich variety of other galaxy properties to which the model has not been calibrated, including the properties of cold gas.

5.1 Introduction

In the last few decades, numerical simulations of galaxy formation have become an indispensable tool for advancing our understanding of the physics of galaxy formation (see Crain & van de Voort, 2023, for a recent review). The rapid growth of computational facilities opened the possibility of simulating large, $\gtrsim (0.1\ \mathrm{Gpc})^3$ cosmological volumes with self-consistent modelling of baryonic processes (e.g. Schaye et al., 2010; Dubois et al., 2014; Schaye et al., 2015; Pillepich et al., 2018; Davé et al., 2019; Bird et al., 2022; Pakmor et al., 2023; Schaye et al., 2023), and reaching spatial resolutions sufficient to study the properties of the multiphase interstellar medium (ISM) of galaxies in smaller volumes situated in a cosmological environment (e.g. Dubois et al., 2021; Feldmann et al., 2023). A major part of this success stems from the progress in computational methods, which has greatly increased the efficiency with which large computational machines can be exploited. In particular, modern astrophysical codes demonstrate impressive performance in standard scaling tests extending up to $\sim 10^5$ compute cores (e.g. Springel et al., 2021; Schaller et al., 2023).

5

At the same time, the advent of advanced observational facilities such as the Atacama Large Millimeter/submillimeter Array (ALMA; Wootten & Thompson, 2009) and JWST (Gardner et al., 2006), has allowed us to study spatially resolved properties of galaxies with unprecedented sensitivity and accuracy, both in the local and high-redshift Universe. Dense, cold interstellar gas can be probed by ALMA at sub-pc resolution either through CO rotation-line emission at low z (e.g. Ramos Almeida et al., 2022) or via [CII] 158 μm line at high z (e.g. Gurman et al., 2023). The properties of warmer gas of high−z objects can be studied at comparable resolution with JWST, using emission lines such as [OIII] or Hβ (e.g. Chen et al., 2023; Giménez-Arteaga et al., 2023). Clearly, for a fair comparison between theory and observations, both JWST and ALMA data demand numerical simulations that reach comparable (or higher) spatial resolutions, and that self-consistently model the multiphase interstellar gas.

Simulations of galaxy formation from the past 10 years have shown remarkable success in matching observational data and producing galaxies with realistic properties (see e.g. Crain & van de Voort, 2023, and references therein). Various observed relations are successfully reproduced by the simulations, including the observed galaxy stellar mass functions (GSMF) and luminosity functions at different redshifts, the galaxy size-stellar mass relation (SSM), the galaxy star formation main sequence, the galaxy stellar mass-metallicity relation and many others (e.g. Schaye et al., 2015; Pillepich et al., 2018; Davé et al., 2019). However, those successful simulations mostly neglected the modelling of the cold neutral gas, which is believed to play a key role in the cosmic baryon cycle (e.g. Péroux & Howk, 2020).

In fact, the large, high-resolution cosmological-volume simulations from the

past decade such as HorizonAGN (Dubois et al., 2014), eagle (Schaye et al., 2015), IllustrisTNG (Pillepich et al., 2018), Simba (Davé et al., 2019) all applied 'a temperature floor' to the interstellar gas and/or artificially enhanced the gas pressure assuming an effective equation of state, so that the (dense) gas cannot cool below $\sim 10^4$ K. The reason for this is twofold. First, simulating the cold phase is numerically difficult because of the small time-steps and small Jeans lengths and masses that are readily reached in the dense, cold phase, which can make the computations unfeasibly slow. Second, modelling the cold phase cannot be accomplished without accounting for the intricate physical processes that are relevant to the cold phase. These processes include (self-)shielding of gas from the extragalactic UV background, the physics of formation and dissociation of molecules and cooling emission therefrom, and the formation and evolution of dust grains, including their live interactions with the cold gas phase (e.g. Tacconi et al., 2020).

Among the most recent simulations of cosmological volumes that allow the gas to enter the cold phase are NewHorizon (Dubois et al., 2021) and FIREbox (Feldmann et al., 2023). The FIREbox simulation used the Fire2 galaxy formation model (Hopkins et al., 2018) and was run in a cosmological volume of (22.1 comoving Mpc)^3 down to redshift $z = 0$ with gas and dark-matter particle masses of, respectively, $m_{\rm gas} = 6.3 \times 10^4\,{\rm M}_\odot$ and $m_{\rm DM} = 3.3 \times 10^5\,{\rm M}_\odot$. As the Fire2 simulations, FIREbox was run with the Gizmo mesh-less finite-mass hydrodynamic and gravity solver (Hopkins, 2015). The element ionization states were calculated based on tabulated results from the simulations with the photoionization code cloudy (Ferland et al., 1998), including a shielding correction for cosmic UV background and local sources. The molecular-to-neutral gas fraction was computed on-the-fly by employing an analytical expression from Krumholz & Gnedin (2011). FIREbox did not model the self-consistent evolution of dust grains but accounted for dust collisional heating/cooling and photo-electric heating using the analytical expressions from Meijerink & Spaans (2005) and Wolfire et al. (2003), assuming a constant dust-to-metal ratio and a constant dust temperature of 30 K (for further details, see Hopkins et al. 2018). The model reproduces multiple scaling relations of low- and intermediate-mass galaxies, including the mass-metallicity relation for the stellar component and gas phase, as well as the relations between galaxy HI and H_2 masses and stellar mass. The drawbacks of FIREbox include the limited statistical power of the simulation due to the relatively small volume, the lack of AGN feedback, and that the galaxies are generally too massive (the GSMF in FIREbox is systematically higher than the observed GSMF).

The domain of the NewHorizon simulation is a zoom-in region of ∼(16 comoving Mpc)^3 taken from the larger, (142 comoving Mpc)^3 volume of the HorizonAGN simulations (Dubois et al., 2014). The NewHorizon simulation was run to redshift $z = 0.25$ with a modified version of the HorizonAGN model, using

the adaptive mesh refinement code RAMSES (Teyssier, 2002). Inside the zoom-in region, the dark-matter particle mass is equal to $1.2 \times 10^6\,\mathrm{M}_\odot$ and the gas spatial resolution can reach ≈ 34 pc in the densest environments. The cooling of metal-enriched gas in NewHorizon is based on tabulated rates from Sutherland & Dopita (1993) at temperatures above $\approx 10^4$ K and Dalgarno & McCray (1972) below $\approx 10^4$ K, which allows the gas to cool to 0.1 K. Primordial species are assumed to be in ionization equilibrium in the presence of a homogeneous redshift-dependent UV background, whose intensity is exponentially suppressed at densities $n_\mathrm{H} \gtrsim 0.01$ cm^{-3} due to self-shielding. The prescription for star formation assumes a variable star formation efficiency based on a cloud turbulent Mach number, following Padoan & Nordlund (2011); Hennebelle & Chabrier (2011). The NewHorizon model incorporates mechanical feedback from CC SNe (Kimm & Cen, 2014) and black hole (BH) spin-dependent feedback from AGN. It neglects feedback from stellar winds, HII regions, and type-Ia supernovae (SNe), and does not include a prescription for the evolution of interstellar dust. The NewHorizon model reproduces – among others – the observed Kennicutt-Schmidt (KS) star formation law (Kennicutt, 1998a; Kennicutt et al., 2007), the cosmic stellar density, and the stellar mass-metallicity relation. The drawbacks of NewHorizon include the limited simulated volume and the lack of data at redshifts lower than $z = 0.25$. Furthermore, similar to FIREbox, most galaxies in NewHorizon are over-massive, which results in a discrepancy with the observed GSMF[1].

The results from NewHorizon and FIREbox demonstrate that simulations of galaxy formation that include a cold ISM are challenging but possible. In this paper, we present the calibration of a new galaxy formation model that includes a cold gas phase: COLIBRE (Schaye et al., in preparation). In order to calibrate the model, we exploit machine-learning techniques, following the approach taken by Kugel et al. (2023) for the FLAMINGO simulations (Schaye et al., 2023). Based on a modest number of actual simulations, we train Gaussian process emulators that reconstruct the GSMF and SSM relations from the COLIBRE simulations as smooth, continuous functions of a small number of subgrid parameters. We then fit these emulators to the observed GSMF and SSM at $z = 0$ and obtain the best-fitting values for the subgrid parameters. Furthermore, we consider variations of the fiducial COLIBRE model that use simplified versions of SN feedback, and discuss how changes in various subgrid parameters affect the galaxy properties. This work is structured as follows. In Section 5.2 we describe the most relevant aspects of the COLIBRE galaxy formation model. In Section 5.3 we present the details of the emulation. In Section 5.4 we outline our strategy for calibrating the COLIBRE model. In Section 5.5 we present the results of the calibration, and in Section 5.6 we summarize our conclusions.

[1] The authors argue, however, that by accounting for the selection effects and environment, these discrepancies can be reduced.

5.2 Simulations

All simulations in this work were run with the open-source astrophysical code SWIFT (Schaller et al., 2023). At the core of the SWIFT code is a task-based parallelism method. The equations of hydrodynamics are solved with the smoothed particle hydrodynamics (SPH) method using the energy-density SPH scheme SPHENIX (Borrow et al., 2022). The gravity is solved with the use of the Fast Multiple Method (Greengard & Rokhlin, 1987) for short-range forces and a particle-mesh method solved in Fourier space for long-range forces. We utilize the linear response method (Ali-Haïmoud & Bird, 2013) to model the evolution of massive neutrinos in the linear approximation, and their effect on dark matter (DM) and baryons.

In the SPH scheme SPHENIX, we adopt the same values of the artificial viscosity and energy conduction parameters as in Borrow et al. (2022). We take the quartic spline as our SPH kernel and set the resolution parameter $\eta = 1.2348$, which gives the effective number of neighbours within kernel support of $N_{\rm ngb} = 64.9$. We also use $\eta = 1.2348$ for BH particles[2], while for stellar particles we set η to 1.1642.

Gas particle time-steps are limited by the standard Courant-Friedrichs-Lewy (CFL) condition with the CFL parameter of $C_{\rm CFL} = 0.2$, as well as by the local gravitational acceleration $\boldsymbol{a}_{\rm grav}$ (such that $\Delta t_{\rm max,grav} \propto 1/\sqrt{|\boldsymbol{a}_{\rm grav}|}$). Additionally, we do not allow the time-steps of gas particles to exceed the time-steps of any of their SPH neighbours by more than a factor of 4. Moreover, if at a given time-step in the simulation, some gas particles are inactive[3] because of their longer time-steps, while one or several of their gas neighbours are active because they are on shorter time-steps, then the inactive particles are 'woken up' and their time-steps are recomputed according to the present conditions (for details, see Durier & Dalla Vecchia, 2012). We also recompute particle time-steps when particles receive energy in SN and/or AGN feedback, which ensures that the injected energy is not quickly dissipated due to numerically enhanced radiative losses, as shown by Durier & Dalla Vecchia (2012).

During the simulation, gas particle masses may increase if the particles are affected by stellar mass loss (§5.2.6) or decrease if (a fraction of) their mass is accreted onto supermassive black holes (SMBHs, see §5.2.10). We do not allow the particle mass to exceed four times the initial particle mass. This is accomplished by splitting gas particles violating the mass threshold into two new particles of equal mass. At the same time, we forbid the particle mass from becoming less

[2] Although neither BH nor stellar particles experience hydrodynamic forces, they follow the SPH neighbour search algorithm to locate their gas neighbours, which is necessary for the modelling of AGN and stellar feedback processes.

[3] The division into active and inactive gas particles is a consequence of the time-step hierarchy adopted in SWIFT: because of differences in the local properties, gas particles may have different time-steps that vary by 2^n where n is some integer number.

than ≈ 50 per cent of the initial mass. This is achieved by not letting SMBHs accrete mass from gas particles whose mass is smaller than the lowest allowed value.

In this work, we use the simulation output at redshifts $z = 0$ and 2. To identify (sub-)haloes in the simulation snapshots, we employ the publicly available structure finder VELOCIraptor[4] (Cañas et al., 2019; Elahi et al., 2019).

5.2.1 Initial conditions

Initial conditions (ICs) of our simulations are produced by the monofonIC code (Hahn et al., 2020; Michaux et al., 2021) using third-order Lagrangian perturbation theory (LPT). We follow the 2-field prescription from Hahn et al. (2021) to generate ICs for baryons and DM and take $z = 127$ as the starting redshift of the simulations. The Gaussian random fields underlying the ICs are taken from the public multiscale Gaussian white noise field *panphasia* (Jenkins, 2013). The values of the phases are indicated in table B1 of Schaye et al. (2015). In all our runs, we use the '3x2pt + all external constraints' cosmology from Abbott et al. (2022): $\Omega_{m,0} = 0.306$, $\Omega_{b,0} = 0.0486$, $\sigma_8 = 0.807$, $h = 0.681$, $n_s = 0.967$. We assume a single massive neutrino species with a mass of 0.06 eV.

The simulations in this work were run in a cosmological volume of (50 comoving Mpc)^3, which is smaller than the volume of the colibre production runs (see Schaye et al., in preparation). In the ICs, for each gas particle there are four corresponding dark-matter particles (for the details on particle positions in the ICs, see, e.g., Richings et al., 2021). We set the number of DM particles to be four times greater than that of the gas in order to minimize the effects of spurious collisional heating of galactic stellar discs. Namely, owing to two-body collisions between stellar and DM particles, the dynamically hotter DM will transfer energy to the dynamically colder stars in an attempt to reach equipartition between the two components, which has a negative impact on the morphology and kinematics of the stellar component in simulated galaxies (Ludlow et al., 2019, 2021, 2023). In total, we use 376^3 particles for the gas and 4×376^3 particles representing DM, which yields a particle mass for the gas (DM) of $1.47\times10^7\,\mathrm{M}_\odot$ ($1.94\times10^7\,\mathrm{M}_\odot$). The associated Plummer-equivalent gravitational softening length for gas (DM) particles, $\varepsilon_{\mathrm{soft,gas}}$ ($\varepsilon_{\mathrm{soft,DM}}$), is equal to the minimum of 3.58 (4.18) comoving kpc and 1.4 (1.64) proper kpc.

As colibre captures the multiphase structure of the ISM, we set the minimum SPH smoothing length, h_{min}, to be very small (but non-zero) to avoid potential runaway collapse of cold, dense gas, following the findings of Ploeckinger et al. (2023). We use $h_{\mathrm{min}}/\varepsilon_{\mathrm{soft,gas}} \ll 10^{-5}$.

[4] https://velociraptor-stf.readthedocs.io/en/latest/

5.2.2 Radiative cooling

The radiative cooling and heating rates for hydrogen and helium (and free electrons therefrom) are computed with the non-equilibrium thermochemistry solver CHIMES (Richings et al., 2014a,b). Additionally, we explicitly track nine heavy elements that contribute most to the cooling rate: C, N, O, Ne, Mg, Si, S, Ca, and Fe. Their rates are provided by a modified version of the Ploeckinger & Schaye (2020) cooling tables (Ploeckinger et al., in preparation), which were generated by CHIMES under the assumption of ionization equilibrium. The tables also account for cooling due to free-free emission and molecular cooling (including from CO, H_2O, OH, HD), while the molecular cooling from H_2 is computed in non-equilibrium by CHIMES. Additionally, we include dust-associated heating and cooling processes using a live dust model (§5.2.4) coupled to CHIMES and account for the inverse Compton cooling off CMB photons and cosmic ray heating. We do not impose an artificial pressure floor or a temperature floor. We allow the gas to cool down to a temperature of 10 K.

Ploeckinger & Schaye (2020) used the photoionization code CLOUDY (Ferland et al., 2017) to tabulate element-by-element radiative cooling and heating rates of 11 individual elements (9 heavy elements, hydrogen, and helium) as functions of the gas density, metallicity, temperature, and redshift. Their equilibrium tables include a modified version of the uniform, redshift-dependent UV and X-ray background from Faucher-Giguère (2020) (see appendix B of Ploeckinger & Schaye (2020) for details on the modification), a cosmic ray ionization background, and an interstellar radiation field (ISRF) whose shape is constrained by the ISFR in the Milky Way Galaxy (Black, 1987). The intensities of the cosmic ray background and ISRF scale as $N_{\rm Jeans}^{1.4}$ where $N_{\rm Jeans}$ is the Jeans column density of the gas[5]. Furthermore, the model includes the effects of dust, such as metal depletion onto dust grains, assuming a constant dust-to-gas ratio of $D/G = 5.6{\times}10^{-3}$ at hydrogen column densities above $10^{20.56}$ cm^{-2}, and a scaling with $N_{\rm Jeans}^{1.4}$ at lower densities.

While Ploeckinger & Schaye (2020) considered a Jeans column density contributed only by gas thermal motions, the updated tables from Ploeckinger et al. (in preparation) also account for contributions from turbulent motions, which become important in the cold ISM. A turbulent velocity of $v_{\rm turb} = 6$ km s^{-1} is assumed and the total Jeans column density, $N_{\rm Jeans}$, is computed as the maximum of the thermal and turbulent Jeans column densities. We do not allow this column density to be greater than the minimum of $N_{\rm Jeans,max} = 10^{24}\,{\rm cm}^2$ and $n_{\rm H}\,l_{\rm max}$ where the length scale $l_{\rm max} = 100\,{\rm kpc}$ and $n_{\rm H}$ is the gas hydrogen number density.

The original version of the Ploeckinger & Schaye (2020) model incorporates

[5]The power of 1.4 comes from the observed Kennicutt-Schmidt relation (Kennicutt, 1998b), $\Sigma_{\rm SFR} \propto \Sigma_{\rm gas}^{1.4}$, where $\Sigma_{\rm SFR}$ and $\Sigma_{\rm gas}$ are the galaxy-averaged star formation rate surface density and gas surface density, respectively.

the effects of self-shielding of cold, dense gas by assuming that the gas shielding column is equal to half of the thermal Jeans column density. In the new version, self-shielding effects are approximated in a similar manner but the thermal Jeans column density is replaced with the total $N_{\rm Jeans}$, which includes contributions from turbulent motions. The other changes between the tables employed in this work and Ploeckinger & Schaye (2020) include adjustments in the scaling relations of the cosmic ray and ISRF rates with $N_{\rm Jeans}$. Further details will be provided in Ploeckinger et al. (in preparation).

We stress that the redshift-dependent UV and X-ray background, cosmic ray ionization background, and ISRF are all applied self-consistently both to individual metal elements, which are evolved assuming ionization equilibrium (but use the non-equilibrium electron density to compute the cooling rates), and to H and He whose chemical abundances and ionization states are calculated in non-equilibrium with the CHIMES solver.

5.2.3 Chemistry

5

We track the abundance of 12 individual species: H, He, C, N, O, Ne, Mg, Si, Fe, Sr, Ba, and Eu[6]. At the start of the simulations, the hydrogen mass fraction is equal to $X_{\rm H} = 0.756$, and the remaining baryonic mass is contained in He. The abundances of heavier elements become non-zero shortly after the first stars have formed, as a result of stellar chemical enrichment (see §5.2.6).

Once stellar production of metals has begun, metals will be gradually released into the ISM where they are expected to get mixed and redistributed, as governed by the turbulent motions in the gas phase (e.g. Klessen & Lin, 2003). However, in galaxy simulations that are conducted with SPH codes, the mixing of metals is notoriously difficult to accomplish. First of all, the standard SPH equations lack intrinsic mixing terms (e.g. Wadsley et al., 2008). Additionally, given the typical resolution of galaxy simulations performed in a cosmological context, the simulations will likely fail to resolve gas turbulent motions already at ~ 10 pc scales, thus suppressing turbulent mixing at these and smaller scales.

To overcome these issues, in COLIBRE we solve a set of diffusion equations for the individually tracked species. The 12 species are diffused among the neighbouring SPH gas particles following the element diffusion prescription from Correa et al. (in preparation). In short, over one time-step, the diffusion affects the mass fraction of some element X carried by gas particle i as

$$X_i^{\rm new} = X_i^{\rm old} + \sum_j K_{ij} \left(X_i^{\rm old} - X_j^{\rm old}\right) \Delta t_{ij}\,, \tag{5.1}$$

[6]This set of elements is not identical to that used in the prescription for gas radiative cooling (§5.2.2). In particular, the abundances of Ca and S are used to compute radiative cooling rates but these elements are not tracked in COLIBRE chemistry. Instead, their abundances are assumed to have solar mass ratios relative to Si.

where the sum runs over the gas neighbours of particle i, which are enumerated by the index j, and the superscripts 'old' and 'new' stand for the mass fraction of element X immediately before and after the diffusion update, respectively. The diffusion coefficients K_{ij} are calculated according to a turbulent mixing model (e.g. Shen et al., 2010); they depend on the gas particle density, the local velocity shear, and the dimensionless diffusion coefficient C, which we set to 0.01. The time-step Δt_{ij} is equal to the minimum of the time-steps of the interacting particles i and j, ensuring the conservation of diffused mass in the case where the gas particles have different time-steps.

5.2.4 Dust

The COLIBRE simulations incorporate a subgrid model for the formation and evolution of interstellar dust, which will be described in detail by Trayford et al. (in preparation). Briefly, the model traces two types of dust grains: graphites and silicates, with the silicates further divided into Mg and Fe flavours. Dust grains are produced in the AGB phase of stellar evolution and in CC SNe. Dust grains grow by accreting mass from the gas phase, and they lose mass in hot gas due to thermal sputtering. Furthermore, we account for grain destruction due to shocks from individual SNe that are unresolved at our resolution, following Asano et al. (2013). Additionally, the model incorporates two processes that modify the sizes of the grains while preserving their (total) mass: grain shattering and coagulation. We assume that all dust grains have spherical shapes and track two grain sizes: grains with radii of 0.01 and 0.1 μm.

The dust grains can readily grow within molecular cloud environments, with local densities exceeding 10^3 cm^{-3} (e.g. Hirashita, 2000; Stepnik et al., 2003). Because such densities and associated spatial scales are poorly resolved in our simulations, we apply a density-dependent boost factor to the rates of the dust-growth processes: accretion from the gas phase and coagulation. The form of the boost factor $\beta_{\rm dust}$ is as follows,

$$\log_{10}\beta_{\rm dust}(n_{\rm H}) = \begin{cases} 0 & n_{\rm H} < n_{\rm b,min} \\ \frac{(\log_{10} n_{\rm H} - \log_{10} n_{\rm b,min})\log_{10}\beta_{\rm dust,max}}{\log_{10} n_{\rm b,max} - \log_{10} n_{\rm b,min}}; & n_{\rm b,min} \leq n_{\rm H} < n_{\rm b,max} \\ \log_{10}\beta_{\rm dust,max} & n_{\rm H} \geq n_{\rm b,max}, \end{cases} \tag{5.2}$$

where the parameters $n_{\rm b,min}$ and $n_{\rm b,max}$ specify, respectively, the density above which we apply a boost and the density above which the boost saturates, with the saturation level controlled by the parameter $\beta_{\rm dust,max}$. Throughout this work, we adopt the following values of the boost parameters: $n_{\rm b,min} = 0.1$ cm^{-3}, $n_{\rm b,max} = 100$ cm^{-3}, and $\beta_{\rm dust,max} = 100$. Variations in the dust-boost factor will be studied in Trayford et al. (in preparation).

The live dust abundances are coupled to the CHIMES solver, accounting for the distribution of grain sizes. The dust is used by CHIMES to calculate the formation rate of molecular hydrogen on dust grains and other reactions facilitated by dust, and to compute the dust-associated heating and cooling processes, including dust radiative cooling and photoelectric heating. Additionally, the live dust abundances are used to set the metal abundances in the gas phase.

5.2.5 Star formation

The COLIBRE prescription for star formation is detailed in Nobels et al. (2023). A gas element is labelled as 'star-forming' if the gas is locally unstable against gravitational collapse.

Mathematically, the instability condition is expressed by requiring that the (absolute) gravitational binding energy of a gas cloud represented by the gas element exceeds its kinetic energy due to thermal and turbulent motions,

$$\alpha \equiv \frac{\sigma_{\rm turb}^2 + \sigma_{\rm th}^2}{G\rho^{1/3}\langle m_{\rm ngb}\rangle^{2/3}} < 1\,, \tag{5.3}$$

in which ρ is the mass density of the gas element, $\langle m_{\rm ngb}\rangle$ is the average mass in the SPH kernel of the gas element, $\sigma_{\rm turb}$ is the three-dimensional turbulent velocity dispersion, $\sigma_{\rm th}$ is the thermal velocity dispersion, and G is the gravitational constant. The value of $\langle m_{\rm ngb}\rangle$ is calculated as $\langle m_{\rm ngb}\rangle = \langle N_{\rm ngb}\rangle m_{\rm gas}$ where $m_{\rm gas}$ is the mass of the gas element for which the star formation criterion is used, and the effective number of gas neighbours in the kernel $\langle N_{\rm ngb}\rangle \approx 65$. Physically, the SPH smoothing length of the gas element can be regarded as the radius of the (collapsing) cloud.

The thermal velocity dispersion, $\sigma_{\rm th}$, is given by

$$\sigma_{\rm th} = \sqrt{\frac{3P}{\rho}}\,, \tag{5.4}$$

where P is the thermal pressure of the gas element. The turbulent velocity dispersion $\sigma_{\rm turb}$ of some gas element labelled by the index i is computed as

$$\sigma_{{\rm turb},i}^2 = \frac{1}{\rho_i}\sum_j m_j |\boldsymbol{v}_i - \boldsymbol{v}_j|^2\, W(|\boldsymbol{r}_i - \boldsymbol{r}_j|, h_i)\,, \tag{5.5}$$

where the sum runs over the neighbours of gas element i, which are labelled by the index j, $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are the vectors describing the positions of gas elements i and j, respectively, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are the velocities of gas elements i and j, and W is the SPH kernel function centred on gas element i, which has a smoothing length h_i.

The finite resolution of our simulations prevents us from directly following gas collapse into stars. Instead, we convert star-forming gas particles into stellar

particles stochastically. To compute the probability of a star-forming gas element becoming a stellar particle, we use the Schmidt (1959) law,

$$p_{\rm sf} = \frac{\dot{m}_{\rm sf}}{m_{\rm gas}} \Delta t = \frac{\varepsilon}{t_{\rm ff}} \Delta t \,, \tag{5.6}$$

where $\dot{m}_{\rm sf}$ is the star formation rate (SFR) of the gas element, Δt is the size of the time-step at which the calculation is performed, $t_{\rm ff} = [3\pi/(32G\rho)]^{1/2}$ is the free-fall time-scale, and we assume a star formation efficiency per free-fall time of $\varepsilon = 0.01$, motivated by the star formation efficiencies inferred from observations of giant molecular clouds (e.g. Krumholz & Tan, 2007; Lee et al., 2016).

Nobels et al. (2023) demonstrated that this prescription for star formation reproduces the observed KS laws for neutral, atomic, and molecular gas.

5.2.6 Stellar evolution and chemical enrichment

A stellar particle physically represents a population of many stars that formed simultaneously from a single gas cloud and therefore possess the same properties. We assume that all stellar particles are characterized by a Chabrier (2003) stellar initial mass function (IMF) with the minimum and maximum masses of 0.1 and 10^2 $M_\odot$, respectively. Besides standard properties such as position and velocity, which are inherited from the parent gas particle, a stellar particle is characterized by its age, initial metallicity, and initial mass. The age is counted since the moment the stellar particle has entered the main sequence of stellar evolution, and the initial mass is the mass at the zero-age main sequence (ZAMS). Once formed, stellar particles become involved in various forms of stellar feedback, among which are stellar winds, radiation pressure, HII regions, core-collapse supernovae (CC SNe) and type-Ia supernovae, and chemical enrichment of the gas phase.

To improve the temporal sampling of stellar feedback processes, we limit the time-steps of stellar particles with age $t_{\rm age} < 40$ Myr to be smaller than 1 Myr. The exact values of the time-steps are determined by gravitational forces exerted on the stellar particles. In contrast, when the age of a stellar particle becomes greater than 10^2 Myr, its feedback-related processes are sampled in time only once every $0.05\, t_{\rm age}$ Myr, which is done in order to limit the computational expense. We verified that this sampling strategy has a negligible impact on our results.

Stellar particles continuously inject metal-enriched gas into their surrounding ISM as part of CC and type-Ia SN feedback as well as during the Asymptotic Giant Branch (AGB) phase, following Wiersma et al. (2009); Schaye et al. (2015). We compute the net stellar yields of a stellar particle of age $t_{\rm age}$ in a given time-step Δt by integrating the Chabrier (2003) IMF over the ZAMS masses of individual stars that died at ages between $t_{\rm age}$ and $t_{\rm age}+\Delta t$. We use the metallicity-dependent stellar lifetime tables from Portinari et al. (1998) to relate ZAMS masses to $t_{\rm age}$, in

which $t_{\rm age}$ physically corresponds to the sum of the time-scales of the stellar H- and He-burning stages. We assume that all mass (including metals) ejected over the lifetime of an individual star that died at age $t_{\rm age}$, is released instantaneously once the stellar particle's age reaches $t_{\rm age}$.

We take CC SN yields from Nomoto et al. (2013) who provide the data for ZAMS masses between 13 and 40 $M_\odot$ and metallicities between 0 and 0.05. We assume that all stars with ZAMS masses between 8 and 40 $M_\odot$ evolve into CC SNe. However, because Nomoto et al. (2013)'s lowest ZAMS stellar mass is $13\,M_\odot$, we linearly extrapolate the yields down to $8\,M_\odot$, at a fixed stellar metallicity. If a stellar particle is born with an initial metallicity greater than 0.05, we evaluate its yields at $Z = 0.05$.

Yields from type-Ia SN feedback are taken from Kobayashi et al. (2020) assuming that type-Ia SNe originate from explosions of near-Chandrasekhar mass white dwarfs. Given the weak dependence of the yields on metallicity, we evaluate all type-Ia SN yields at $Z = 0.02$. The times of detonation of individual type-Ia SNe are calculated according to a delay time distribution (DTD, see Eq. 5.15 in §5.2.9).

We model metal enrichment during the AGB phase of stellar evolution by adopting nucleosynthesis yields from Karakas (2010); Doherty et al. (2014); Fishlock et al. (2014); Karakas & Lugaro (2016); Cinquegrana & Karakas (2022). We connect these datasets by creating a single look-up table with 'net stellar yield' extending over the mass range from 1 to 12 $M_\odot$ and metallicity ranging between 10^{-4} and 0.1. Further details on metal enrichment will be provided by Correa et al. (in preparation).

5.2.7 Early stellar feedback

Our stellar evolution model includes three early stellar feedback processes from massive stars: stellar winds, direct radiation pressure, and HII regions. To determine the energies, momenta, and ionizing flux injected into the gas phase by these feedback processes, we use the Binary Population and Spectral Synthesis (BPASS) tables (Eldridge et al., 2017; Stanway & Eldridge, 2018) version 2.2.1 with a Chabrier (2003) IMF whose minimum and maximum stellar masses are 0.1 and $10^2\,M_\odot$, respectively.

The details of the numerical implementation and impact of these early feedback processes will be presented in Ploeckinger et al. (in preparation). Briefly, stellar wind and direct radiation pressure feedback are realized by star particles stochastically kicking their gas neighbours with a kick velocity of 50 km s^{-1}. The number of kicks in the stellar-wind feedback is based on the cumulative momentum provided by the BPASS tables given the stellar particle's age and metallicity at birth. The radiation pressure feedback is determined by the photon energy spectrum, which is also fetched from the BPASS tables. To calculate the photon

momentum exerted onto the gas, we use the wavelength-dependent optical depth taken from the modified tables of Ploeckinger & Schaye (2020) (see §5.2.2), which is based on the local Jeans column density. Finally, in HII regions young star particles stochastically ionize and heat their surrounding gas to a temperature of $T = 10^4$ K, following a Strömgren sphere approximation. The probability of a gas particle becoming an HII region depends on the density of the ambient medium and on the BPASS ionizing photon flux from the stellar particle(s). While a gas particle is part of an HII region, it is not allowed to turn into a stellar particle even if it satisfies the star formation criterion. Young stellar particles select a new set of their gas neighbours as HII regions every 2 Myr, provided the conditions for the Strömgren sphere are fulfilled.

5.2.8 Core collapse supernova feedback

The COLIBRE model for feedback from CC SNe is a modified version of the thermal-kinetic model of Chaikin et al. (2023).

The amount of energy in CC SN feedback released by a stellar particle in a time-step from t to $t + \Delta t$ is calculated as

$$\Delta E_{\rm CCSN} = 10^{51}\,{\rm erg}\, f_{\rm E}\, m_* \int_{m_{\rm d}(t+\Delta t)}^{m_{\rm d}(t)} \Phi(m)\,{\rm d}m\,, \tag{5.7}$$

in which $\Phi(m)$ is the Chabrier (2003) IMF and $m_{\rm d}(t)$ denotes the mass of the star(s) that explode as core-collapse SNe at age t. We use the metallicity-dependent stellar lifetime tables from Portinari et al. (1998) to compute $m_{\rm d}(t)$. The function $m_{\rm d}(t)$ is non-zero only for ZAMS masses between $m_{\rm min,CCSN} = 8$ and $m_{\rm max,CCSN} = 10^2\,{\rm M}_\odot$, which roughly correspond to the stellar ages of ≈ 40 and 3 Myr, respectively.

Unlike Chaikin et al. (2023), we assume that the energy of a single SN in units of 10^{51} erg, $f_{\rm E}$, depends on the thermal pressure of the parent gas particle, $P_{\rm birth}$, measured immediately before it turned into the stellar particle under consideration. The relation between $f_{\rm E}$ and $P_{\rm birth}$ has the following form[7],

$$f_{\rm E}(P_{\rm birth}) = f_{\rm E,min} + \frac{f_{\rm E,max} - f_{\rm E,min}}{1 + \exp\left(-\dfrac{\log_{10} P_{\rm birth} - \log_{10} P_{\rm birth,0}}{\sigma_P}\right)}\,, \tag{5.8}$$

where $f_{\rm E,min}$ and $f_{\rm E,max}$ are, respectively, the minimum and maximum energies that can be injected by a single SN, in units of 10^{51} erg, $P_{\rm birth,0}$ is a normalization constant, which we will call the pivot birth pressure, and the parameter σ_P

[7] This function can also be expressed in the functional form used in EAGLE for their metallicity-dependent SN energy, $f_{\rm E} = f_{\rm E,min} + (f_{\rm E,max} - f_{\rm E,min})/[1 + (P_{\rm birth}/P_{\rm birth,0})^{-n_P}]$, in which $n_P = 1/(\sigma_P \log_{10} {\rm e})$. However, in addition to the dependence on the metallicity, the SN energy in EAGLE, $f_{\rm E}$, was a function of the birth density of star particles.

defines the width of the transition from $f_{\rm E,min}$ to $f_{\rm E,max}$. The functional form of equation (5.8) implies that the SN feedback of stellar particles born in higher gas pressure environments will be more energetic. In our fiducial setting, we take $f_{\rm E,min} = 0.1$, $f_{\rm E,max} = 3.5$, and $\sigma_P = 0.3$, while the best value of $P_{\rm birth,0}$ will be predicted by emulators (see below). Additionally, in §5.5.4, we will show how variations in $f_{\rm E,min}$, $f_{\rm E,max}$, and σ_P affect the simulated galaxies, and discuss how the fiducial values of these three parameters were chosen.

Physically, the dependence of the SN energy on $P_{\rm birth}$ can be interpreted as non-universality of the stellar IMF, which is not unrealistic. In fact, a variable IMF has been suggested by multiple observational studies (e.g. Thomas et al., 2011; Cappellari et al., 2012; Martín-Navarro et al., 2015; Li et al., 2017) and a pressure-dependent IMF has been employed in numerical simulations to successfully reproduce the observational trends (e.g. Barber et al., 2018, 2019). Values of $f_{\rm E}$ greater than one can be regarded as accounting for hypernovae (e.g. Woosley et al., 1999), and/or as compensation for some degree of numerical overcooling in high-density (and typically also high-pressure) gas (e.g. Stinson et al., 2006; Dalla Vecchia & Schaye, 2012).

The energy $\Delta E_{\rm CCSN}$ is injected stochastically into the gas within the SPH kernel of the stellar particle. As in Chaikin et al. (2023), the parameter $f_{\rm kin}$ is used to split the energy $\Delta E_{\rm CCSN}$ between the two channels of energy injection: thermal and kinetic. A fraction $f_{\rm kin}\,\Delta E_{\rm CCSN}$ is injected kinetically, while the remainder, $(1-f_{\rm kin})\Delta E_{\rm CCSN}$, is distributed within the gas in thermal form.

Thermal channel of energy injection

The thermal channel of CC SN feedback utilizes the stochastic model of Dalla Vecchia & Schaye (2012), which was employed in the EAGLE simulations (Schaye et al., 2015). In the Dalla Vecchia & Schaye (2012) model, gas particles receive SN energy from nearby stellar particles with a certain probability, $p_{\rm SN,heat}$. The amount of the injected energy, $\Delta E_{\rm SN}$, is chosen such that following the injection, the gas particle's temperature is increased by a fixed, pre-defined amount, $\Delta T_{\rm SN}$. Mathematically, the relationship between $\Delta E_{\rm SN}$ and $\Delta T_{\rm SN}$ is expressed as

$$\Delta E_{\rm heat}(m_{\rm gas}, \Delta T_{\rm SN}) = \frac{k_{\rm B}\Delta T_{\rm SN}}{(\gamma-1)}\frac{m_{\rm gas}}{\mu m_{\rm p}}, \tag{5.9}$$

where $m_{\rm p}$ is the proton mass, $k_{\rm B}$ is the Boltzmann constant, $m_{\rm gas}$ indicates the mass of the gas particle that is being heated, $\gamma = 5/3$ is the ratio of specific heats for an ideal monatomic gas, and $\mu = 0.59$ is the mean molecular weight of a fully ionized gas. The probability that a given stellar particle will heat one of its gas neighbours in some time-step from t to $t+\Delta t$, $p_{\rm SN,heat}$, is calculated as the ratio of the available SN energy in the time-step, $(1-f_{\rm kin})\Delta E_{\rm CCSN}$, to the energy required to heat the gas mass contained within the stellar kernel, $\Delta E_{\rm heat}(m_{\rm ngb}, \Delta T_{\rm SN})$,

$$p_{\rm SN,heat} = (1 - f_{\rm kin}) \frac{\Delta E_{\rm CCSN}(t, \Delta t, m_*, f_{\rm E})}{\Delta E_{\rm heat}(m_{\rm ngb}, \Delta T_{\rm SN})}, \tag{5.10}$$

where $m_{\rm ngb}$ is the sum of the masses of the gas neighbours found within the kernel of the stellar particle. Once $p_{\rm SN,heat}$ has been computed, we start drawing random numbers r from the interval $0 \leq r < 1$. We draw the random numbers $N_{\rm ngb}$ times where $N_{\rm ngb}$ is the number of the stellar particle's gas neighbours. We then check how many times (out of $N_{\rm ngb}$) the random numbers happened to be smaller than $p_{\rm SN,heat}$. The number of such outcomes determines the number of energy injections the stellar particle will carry out in this time-step. To decide which gas particles within the stellar particle's kernel will receive these energy injections, we adopt the isotropic neighbour selection algorithm from Chaikin et al. (2022) with the maximum number of rays set to 8.

We note that in EAGLE, thermal energy injections were distributed among gas neighbours with a mass-weighted neighbour selection scheme, as opposed to the isotropic method from Chaikin et al. (2022) who showed that the former scheme is biased towards injecting SN energy into high-density gas, and for this reason leads to more radiative energy losses than the isotropic algorithm. Whereas EAGLE used a constant heating temperature $\Delta T_{\rm SN}$, in COLIBRE $\Delta T_{\rm SN}$ depends on the gas density.

A density-dependent heating temperature

Specifically, the CC SN feedback in the EAGLE simulations used a constant heating temperature of $\Delta T_{\rm SN} = 10^{7.5}$ K, with the detailed motivation provided by Dalla Vecchia & Schaye (2012). In short, values greater than $\sim 10^{7.5}$ K would lead to undersampling of SN feedback because the number of energy injections distributed within the surrounding gas by a single stellar particle over its lifetime, $\langle N_{\rm heat,tot} \rangle$, would be less than 1. The value of $\langle N_{\rm heat,tot} \rangle$ is computed as

$$\begin{aligned} \langle N_{\rm heat,tot} \rangle &= \frac{(1 - f_{\rm kin}) E_{\rm CCSN,tot}(m_*, f_{\rm E})}{\Delta E_{\rm heat}(m_{\rm gas}, \Delta T_{\rm SN})}, \\ &= 0.91 (1 - f_{\rm kin}) f_{\rm E} \left(\frac{m_*}{m_{\rm gas}} \right) \left(\frac{\Delta T_{\rm SN}}{10^{7.5}\,{\rm K}} \right)^{-1}, \end{aligned} \tag{5.11}$$

where m_* is the initial mass of the stellar particle, $m_{\rm gas}$ is the (average) mass of gas particles in the simulation, and $E_{\rm CCSN,tot}(m_*, f_{\rm E}) = 10^{51}\,{\rm erg}\, f_{\rm E}\, m_* \int_{m_{\rm min,CCSN}}^{m_{\rm max,CCSN}} \Phi(m)\,{\rm d}m$ is the total CC SN energy released by the stellar particle over the course of its lifetime. Assuming that $m_* \approx m_{\rm gas}$, $f_{\rm E} \sim 1$, and $f_{\rm kin} \ll 1$ and requiring that each star particle on average deposits at least one energy injection in its lifetime (i.e. $\langle N_{\rm heat,tot} \rangle \gtrsim 1$), gives a constraint on the heating temperature $\Delta T_{\rm SN} \lesssim 10^{7.5}$ K.

On the other hand, $\Delta T_{\rm SN}$ needs to be high enough to prevent the injected energy from being radiated away before doing work, which would lead to inefficient SN feedback, which is a consequence of the limited resolution (e.g. Stinson et al., 2006; Dalla Vecchia & Schaye, 2012). Assuming that at high temperatures radiative cooling is dominated by bremsstrahlung, Dalla Vecchia & Schaye (2012) showed that the maximum density at which the feedback can remain efficient is

$$n_{\rm H,crit} = 2\,{\rm cm}^{-3}\left(\frac{\Delta T_{\rm SN}}{10^{7.5}\,{\rm K}}\right)^{3/2}\left(\frac{f_t}{10}\right)^{-3/2}\times \left(\frac{m_{\rm gas}}{1.5\times10^7\,{\rm M}_\odot}\right)^{-1/2}\left(\frac{\langle N_{\rm ngb}\rangle}{65}\right)^{-1/2}\times \left(\frac{\mu}{0.6}\right)^{-9/4}\left(\frac{g(X_{\rm H})}{0.14}\right)^{3/2}, \tag{5.12}$$

where f_t is the ratio of the radiative cooling time-scale of the heated gas element to the sound-crossing time-scale across the element and the function $g(X_{\rm H}) = X_{\rm H}^{2/3}(1+X_{\rm H})^{-1}(1+3X_{\rm H})^{-1}$ with $X_{\rm H}$ being the hydrogen mass fraction[8].

In contrast to EAGLE, in COLIBRE we exploit equation (5.12) and allow the heating temperature to vary within a certain range of values, $\Delta T_{\rm SN,min} < \Delta T_{\rm SN} < \Delta T_{\rm SN,max}$, monotonically increasing with the gas density. In our fiducial model, we set $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$ to $10^{6.5}$ and $10^{7.5}$ K, respectively. The usage of values lower than $10^{7.5}$ K will greatly increase the sampling of SN feedback events in low-mass galaxies where the number of stellar particles (and hence SN energy injections) may be small. Moreover, lower $\Delta T_{\rm SN}$ will make SN feedback less destructive in gas environments with relatively low densities, which potentially alleviates the problem of overly large SN-driven bubbles identified in the EAGLE simulations (Bahé et al., 2016).

More precisely, in COLIBRE we assume that the value of the heating temperature, $\Delta T_{\rm SN}$, depends on the average (physical) gas density at the location of the star particle, $\rho_{\rm SN}$, which is estimated in the time-step when the star particle does SN feedback. We compute $\rho_{\rm SN}$ as

$$\rho_{\rm SN} = \sum_{i=1}^{N_{\rm ngb}} m_i\, W(|\boldsymbol{r}_* - \boldsymbol{r}_i|, h_*), \tag{5.13}$$

where the sum runs over all gas particles within the stellar kernel, m_i is the mass of gas particle i, $\boldsymbol{r}_*$ and $\boldsymbol{r}_i$ are the coordinates of the stellar particle and gas par-

[8] From separate tests, we found that at the resolution of our simulations ($m_{\rm gas} = 1.47\times10^7\,{\rm M}_\odot$), the requirement of $f_t = 10$ proposed by Dalla Vecchia & Schaye (2012) is sufficient but not necessary: somewhat lower values of f_t are acceptable too, as long as $f_t \gtrsim 2$. For example, for $f_t = 2$, the critical density for $\Delta T_{\rm SN} = 10^{7.5}$ K becomes ≈ 20 cm^{-3}. For comparison, the median density in our simulations at which CC SNe take place is ~ 1 cm^{-3}.

ticle i, respectively, and W is the SPH kernel function with the stellar particle's smoothing length h_*. After having computed $\rho_{\rm SN}$, we convert it to a hydrogen number density $n_{\rm H,SN}$ assuming primordial abundances, with the hydrogen mass fraction of $X_{\rm H} = 0.756$, and calculate $\Delta T_{\rm SN}$ as

$$\Delta T_{\rm SN}(n_{\rm H,SN}) = \Delta T_{\rm SN,0} \left(\frac{n_{\rm H,SN}}{n_{\rm SN,0}} \right)^{2/3}, \tag{5.14}$$

in which $\Delta T_{\rm SN,0}$ and $n_{\rm SN,0}$ are free parameters and the slope of 2/3 is motivated by the cooling argument from Dalla Vecchia & Schaye (2012), our equation (5.12). Because of the degeneracy between $n_{\rm SN,0}$ and $\Delta T_{\rm SN,0}$ ($\Delta T_{\rm SN,0} \propto n_{\rm SN,0}^{2/3}$), we fix $\Delta T_{\rm SN,0}$, setting it to $10^{6.5}$ K and only consider $n_{\rm SN,0}$ in the following.

Kinetic channel of energy injection

The remaining part of CC SN energy, which is not used up in the thermal channel, $f_{\rm kin}\,\Delta E_{\rm CCSN}$, is released in kinetic form, following a modified version of the stochastic kinetic model of Dalla Vecchia & Schaye (2008). The full details of our algorithm for SN kinetic feedback are presented in Chaikin et al. (2023). Briefly, stellar particles inject kinetic energy with a probability

$$p_{\rm kick,pair} = f_{\rm kin} \frac{\Delta E_{\rm CCSN}(t, \Delta t, m_*, f_{\rm E})}{2 \Delta E_{\rm kick}(m_{\rm ngb}, \Delta v_{\rm kick})},$$

where $\Delta E_{\rm kick}(m_{\rm ngb}, \Delta v_{\rm kick}) = m_{\rm ngb}\,\Delta v_{\rm kick}^2/2$ and $\Delta v_{\rm kick}$ is the desired kick velocity. Chaikin et al. (2023) showed that low values of $\Delta v_{\rm kick}$, such as 50 km s^{-1}, help drive turbulence in the neutral ISM and improve the agreement with the observed spatially resolved relation between HI velocity dispersion and SFR surface density in nearby galaxies. Based on these findings, in this work, we adopt $\Delta v_{\rm kick} = 50$ km s^{-1}. The effect of varying $\Delta v_{\rm kick}$ between 10 and 10^3 km s^{-1} can be found in Chaikin et al. (2023).

Similarly to the thermal SN feedback, once we know $p_{\rm kick,pair}$, we draw a random number $N_{\rm ngb}$ times from an interval of $0 \leq r < 1$. The number of kick events that the stellar particle will distribute in the time-step from t to $t + \Delta t$ is equal to the number of times the condition $r < p_{\rm kick,pair}$ is found. In each kick event, the stellar particle kicks *two* of its gas neighbours in opposite directions, which is necessary to conserve linear momentum. Additionally, the model ensures that angular momentum and energy in SN feedback are exactly conserved[9] and the injected energy is distributed statistically isotropically. For further details, we refer the reader to Chaikin et al. (2023).

[9]The exact conservation of energy is realized by accounting for the relative motion between stellar particles and their gas neighbours. Owing to the relative velocity corrections, gas particles may experience velocity kicks that are greater or smaller than the desired kick velocity $\Delta v_{\rm kick}$.

5.2.9 Type-Ia supernovae

We implement type-Ia SN feedback as a purely thermal ($f_{\rm kin} = 0$) isotropic stochastic feedback following the 'isotropic' algorithm from Chaikin et al. (2022). We assume that the heating temperature in type-Ia SN feedback scales with the gas density in the same way as for CC SN feedback, following equation (5.14), where the values of $\Delta T_{\rm SN,min}$, $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,0}$, $n_{\rm H,0}$, and the maximum number of rays are set to those from CC SN feedback.

To calculate the energy budget for type-Ia SN feedback executed by one stellar particle, we use a DTD,

$$\mathrm{DTD}(t) = \frac{\nu}{\tau} \exp\left(-\frac{t - t_{\rm delay}}{\tau}\right) \Theta(t - t_{\rm delay}), \tag{5.15}$$

in which $\nu = 1.6 \times 10^{-3}\,\mathrm{M}_\odot^{-1}$ is the total number of type-Ia SNe that will ever occur per unit initial stellar mass, $\tau = 2$ Gyr is the type-Ia SN time-scale, and $\Theta(x)$ is the Heaviside step function. Nobels et al. (in preparation) will show that this form of DTD results in good agreement with the observed rates of type-Ia SNe.

As was the case for CC SNe, the energy of type-Ia SN released by one stellar particle corresponds to the combined energy from many individual type-Ia SNe that are not resolved in our simulations. We set the time $t_{\rm delay}$ to 40 Myr, which marks the delay since the birth of the stellar particle before the first unresolved, individual SN has gone off and contributed its energy to the stellar particle's total energy. The energy from all individual type-Ia SNe in a time-step $[t, t + \Delta t)$ is calculated by integrating equation (5.15) from t to $t + \Delta t$ and assuming an energy per individual type-Ia SN of 10^{51} erg.

Energy-wise, type-Ia feedback is subdominant to that from CC SNe, and its presence has only little effect on the galaxy properties that are relevant for the calibration of the COLIBRE simulations (see Appendix 5.A for more details). Unless stated otherwise, all discussions about SN feedback in the following text will refer entirely to CC SN feedback.

5.2.10 Black holes

In galaxy simulations, supermassive black holes (SMBHs) are represented by collisionless BH particles, which can grow in mass by accreting surrounding gas and/or by merging with other BH particles (e.g. Springel et al., 2005; Booth & Schaye, 2009).

Seeding

To seed SMBHs in the simulation, we run an on-the-fly friends-of-friends (FoF) group finder, using a linking length of 0.2 times the mean inter-particle separation (e.g. Di Matteo et al., 2008). The FoF algorithm is executed every $\Delta \log_{10} a =$

1.00751, starting at $a = 0.05$, where a is the cosmic scale factor. SMBHs are seeded in haloes whose FoF mass is greater than $10^{10}\,\mathrm{M}_{\odot}$, that contain at least 256 particles (of any type), and that do not already harbour a BH particle.

During seeding, we look for the densest gas particle in the FOF halo and convert it into a BH particle, which inherits the gas particle's dynamical mass, velocity, and position. For all mass-dependent processes that are modelled in a subgrid fashion, such as gas accretion onto SMBHs and energy feedback, we use the BH subgrid mass, as opposed to the dynamical mass of the particle, in order to allow BH masses smaller than the particle mass (e.g. Springel et al., 2005; Booth & Schaye, 2009). The subgrid mass is initially equal to the seed mass, $M_{\mathrm{BH,seed}}$, which is one of the free parameters of the simulation.

Gas accretion

The (instantaneous) mass accretion rate onto a BH particle is computed using a modified Bondi-Hoyle-Lyttleton formula (Krumholz et al., 2006),

$$\dot{m}_{\mathrm{BH}} = 4\pi G^2 \frac{m_{\mathrm{BH}}^2 \rho_{\mathrm{gas}}}{c_{\mathrm{sound}}^3} \left[\frac{(1+\mathscr{M}^2)^4}{1.1^2+\mathscr{M}^2} + \frac{1}{(0.34\,f_{\star})^2} \right]^{-1/2}, \tag{5.16}$$

where $\mathscr{M}^2 = (\sigma_{\mathrm{turb}}/c_{\mathrm{sound}})^2 + (v_{\mathrm{gas}}/c_{\mathrm{sound}})^2$ is the Mach number squared, $f_{\star} = 1/[1+(\omega\, r_{\mathrm{Bondi}}/c_{\mathrm{sound}})^{0.9}]$ is the correction due to vorticity in the gas flow with ω being the vorticity, and $r_{\mathrm{Bondi}} = G m_{\mathrm{BH}}/c_{\mathrm{sound}}^2$ is the Bondi radius with m_{BH} being the subgrid mass of the BH particle. The absolute value of the gas bulk velocity $\boldsymbol{v}_{\mathrm{gas}}$, $v_{\mathrm{gas}} \equiv |\boldsymbol{v}_{\mathrm{gas}}|$, the gas turbulent velocity dispersion σ_{turb}, and the vorticity $\omega \equiv |\nabla \times \boldsymbol{v}_{\mathrm{gas}}|$ are all computed as mass-weighted averages over all gas neighbours within the kernel of the BH particle. Finally, we calculate the gas mass density, ρ_{gas}, in the standard SPH way by applying Eq. (5.13) to the gas neighbours within the BH kernel.

Out of the total gas mass accreted by a BH particle over a given time-step from t to $t + \Delta t$, $\dot{m}_{\mathrm{BH}}\Delta t$, the BH receives a fraction

$$\Delta m_{\mathrm{BH}} = (1-\varepsilon_{\mathrm{r}})\dot{m}_{\mathrm{BH}}\Delta t\,, \tag{5.17}$$

where ε_{r} is the radiative efficiency. The remaining mass, $\varepsilon_{\mathrm{r}}\dot{m}_{\mathrm{BH}}\Delta t$, is assumed to have been converted into energy that escapes the SMBH as radiation (see §5.2.10). If the updated subgrid mass of the BH particle, $m_{\mathrm{BH}}^{\mathrm{new}} = m_{\mathrm{BH}} + \Delta m_{\mathrm{BH}}$, is greater than its dynamical mass at the beginning of the time-step, $m_{\mathrm{BH}}^{\mathrm{dyn}}$, then the value of $m_{\mathrm{BH}}^{\mathrm{dyn}}$ is increased to $m_{\mathrm{BH}}^{\mathrm{new}}$. To ensure conservation of mass, the mass deficit, $m_{\mathrm{BH}}^{\mathrm{new}} - m_{\mathrm{BH}}^{\mathrm{dyn}}$ is 'nibbled' from the mass of the gas particles that reside within the BH kernel, following the method of Bahé et al. (2022). Briefly, as a result of

nibbling, some gas neighbour i loses δm_i of its mass, which is calculated as

$$\delta m_i = \frac{m_{\rm BH}^{\rm new} - m_{\rm BH}^{\rm dyn}}{1-\varepsilon_{\rm r}} \frac{W(|\boldsymbol{r}_{\rm BH} - \boldsymbol{r}_i|, h_{\rm BH})\, m_i}{\sum_j W(|\boldsymbol{r}_{\rm BH} - \boldsymbol{r}_j|, h_{\rm BH}) m_j}, \tag{5.18}$$

while the dynamical mass of the BH is increased by $(1-\varepsilon_{\rm r})\delta m_i$ due to the contribution of gas neighbour i. In Eq. (5.18), the sum is computed over all gas neighbours in the SPH kernel of the BH particle, $\boldsymbol{r}_j$ and m_j denote, respectively, the position and mass of gas neighbour j, $h_{\rm BH}$ is the smoothing length of the BH particle, and $\boldsymbol{r}_{\rm BH}$ describes the BH particle's position.

Conversely, if $m_{\rm BH}^{\rm new}$ is less than $m_{\rm BH}^{\rm dyn}$, then we assume that the difference $m_{\rm BH}^{\rm dyn} - m_{\rm BH}^{\rm new} > 0$ represents a subgrid gas reservoir around the SMBH and all accreted mass comes therefrom. We then only reduce $m_{\rm BH}^{\rm dyn}$ by $\varepsilon_{\rm r}\Delta m_{\rm BH}$ to account for the energy that has been converted into radiation. No gas particle's mass is nibbled in this case.

Finally, in COLIBRE the mass accretion rate, $\dot{m}_{\rm BH}$, is capped above 10 times the mass accretion rate at the Eddington luminosity, $\dot{m}_{\rm Edd}$,

$$\dot{m}_{\rm BH,max} = 10\,\dot{m}_{\rm Edd} = 10\,\frac{4\pi G\, m_{\rm BH}\, m_{\rm p}}{\varepsilon_{\rm r}\sigma_{\rm T} c}, \tag{5.19}$$

where $\sigma_{\rm T}$ is the Thompson cross section for electron scattering and c is the speed of light. In practice, we find that setting $\dot{m}_{\rm BH,max}$ to $\dot{m}_{\rm Edd}$, which is commonly done in galaxy simulations (e.g. Schaye et al., 2015; Pillepich et al., 2018; Dubois et al., 2021), results in negligible differences in galaxy and SMBH properties, as super-Eddington accretion events are rare and short-lived.

BH mergers

Following Bahé et al. (2022), two BH particles will merge if the distance between them, $\Delta r_{\rm BH}$, is less than three gravitational softening lengths, $\Delta r_{\rm BH} < 3\varepsilon_{\rm soft,gas}$; the less massive BH is within the kernel of the more massive BH; and if their relative velocity $\Delta v_{\rm BH}$ satisfies $\Delta v_{\rm BH} < \sqrt{2G(M+m)/\Delta r}$ where M and m are the dynamical masses of the larger and smaller merging BH particle, respectively. Once the merger criteria are simultaneously satisfied, the BHs are instantaneously merged. During the merging, all properties of the BH particle with the lower subgrid mass are transferred to the more massive particle, after which the less massive BH particle is removed from the simulation.

BH repositioning

SMBHs are subject to significant dynamical friction, which causes them to lose their orbital energy and spiral in towards the centre of the host galaxy (e.g. Ostriker, 1999). Because galaxy simulations of representative volumes lack the resolution to properly capture the effects of dynamical friction, SMBHs have to be

pushed towards the centre of the galaxy with an ad hoc prescription (e.g. Di Matteo et al., 2008; Bahé et al., 2022). In COLIBRE we follow the method of Bahé et al. (2022) where every time-step Δt, BH particles search for the gas particle that has the lowest gravitational potential and are immediately 're-positioned' at its coordinates. The gas particles that are considered for BH re-positioning have to be both within the BH particles' SPH smoothing kernels and within 3 gravitational softening lengths from the BH. As recommended by Bahé et al. (2022), when selecting the gas neighbour with the lowest gravitational potential we exclude the contribution of the BH particle to the potential.

Feedback from active galactic nuclei

Following Booth & Schaye (2009) and similar to EAGLE (Schaye et al., 2015), we implement active galactic nuclei (AGN) feedback from SMBHs as a purely thermal AGN feedback.

In a given time-step of length Δt, we first compute the energy of the radiation emitted around the SMBH that is coupled to the gas, $\Delta E_{\rm AGN}$, based on the SMBH's instantaneous gas accretion rate $\dot{m}_{\rm BH}$ (Eq. 5.16). The value of $\Delta E_{\rm AGN}$ is given by

$$\Delta E_{\rm AGN} = \varepsilon_{\rm f} \varepsilon_{\rm r} \dot{m}_{\rm BH} c^2 \Delta t\,, \tag{5.20}$$

where $\varepsilon_{\rm f}$ is the coupling efficiency. For both $\varepsilon_{\rm r}$ and $\varepsilon_{\rm f}$, we use a value of 0.1, which yields a plausible AGN luminosity function and realistic SMBH masses.

As is the case with stellar feedback, injecting the energy $\Delta E_{\rm AGN}$ into surrounding gas may result in numerical overcooling if $\Delta E_{\rm AGN}$ is relatively small. Following Booth & Schaye (2009), we wait until a sufficiently large amount of energy has been accumulated by the BH. Numerically, this is achieved by having each BH particle carry an energy reservoir, $E_{\rm AGN}^{\rm reservoir}$, which is empty upon BH seeding but whose energy is increased every time-step by the value of $\Delta E_{\rm AGN}$ for that time step. Once the energy in the reservoir has exceeded some threshold energy $\Delta E_{\rm AGN,thr}$, we inject the energy $\Delta E_{\rm AGN,thr}$ into one of the gas particles within the BH's kernel and subtract an equivalent amount of energy from the reservoir. We define $\Delta E_{\rm AGN,thr}$ as the energy that results in a temperature increase of the heated gas neighbour by $\Delta T_{\rm AGN}$, $\Delta E_{\rm AGN,thr} \equiv \Delta E_{\rm heat}(\langle m_{\rm gas}\rangle, \Delta T_{\rm AGN})$, where $\langle m_{\rm gas}\rangle$ is the average gas particle mass in the BH's kernel and the expression for $\Delta E_{\rm heat}(\langle m_{\rm gas}\rangle, \Delta T_{\rm AGN})$ is given by equation (5.9). We use $\Delta T_{\rm AGN} = 10^9$ K, which is the same value as used in the fiducial EAGLE simulation. If the BH particle accretes rapidly and/or its time-step is very long, the energy in the reservoir $E_{\rm AGN}^{\rm reservoir}$ may temporarily exceed $N_{\rm AGN}\Delta E_{\rm AGN,thr}$ where $N_{\rm AGN}$ is some integer number greater than one. In this case, the energy $\Delta E_{\rm AGN,thr}$ is injected into $N_{\rm AGN}$ gas neighbours, and $E_{\rm AGN}^{\rm reservoir}$ is reduced by $N_{\rm AGN}\Delta E_{\rm AGN,thr}$.

To select the gas neighbours that will receive the energy, we employ the 'Minimum Distance' algorithm from Chaikin et al. (2022). In the 'Minimum Distance'

algorithm, if a BH particle needs to distribute $N_{\rm AGN} \geq 1$ energy injections among its neighbours, then the $N_{\rm AGN}$ closest neighbours each receive one energy injection. As in Bahé et al. (2022), we set the maximum number of rays carried by BH particles to $N_{\rm BH,rays} = 50$, meaning that a BH particle can simultaneously heat no more than 50 of its neighbours. In rare events where $N_{\rm AGN}$ is greater than $N_{\rm BH,rays}$, we increase $\Delta T_{\rm AGN}$ by $N_{\rm AGN}/N_{\rm BH,rays}$ and heat the $N_{\rm BH,rays}$ closest neighbours using the updated value of $\Delta T_{\rm AGN}$. Additionally, if the number of gas particles within the BH's kernel, $N_{\rm ngb}$, is smaller than $\min(N_{\rm AGN}, N_{\rm BH,rays})$, then $\Delta T_{\rm AGN}$ is raised by $\min(N_{\rm AGN}, N_{\rm BH,rays})/N_{\rm ngb}$ and all $N_{\rm ngb}$ particles receive the energy corresponding to the updated $\Delta T_{\rm AGN}$.

5.3 Emulators

We use Gaussian process emulators to validate the choice of the subgrid parameters in the COLIBRE model and to demonstrate that simplified models with reduced numbers of subgrid parameters cannot provide an equally good fit to the target observational data. We construct Gaussian process emulators using the python package SWIFT-EMULATOR (Kugel & Borrow, 2022).

Generally, finding the values of subgrid parameters for which the simulation best reproduces a certain set of observational data is a cumbersome process. Given the number of 'knobs to tune' in a galaxy formation model, the search for the best-fitting values may require running thousands of simulations for various values of the subgrid parameters. Such a blind search would clearly be inefficient for COLIBRE because of the high dimensionality of the model's parameter space. Instead, we will follow the method[10] from Kugel et al. (2023), who employed Gaussian process emulators to calibrate the large (but lower resolution and using a simpler galaxy formation model) cosmological simulations FLAMINGO (Schaye et al., 2023; Kugel et al., 2023).

We set up a number of simulations (~ 100) that sample a part of the COLIBRE parameter space of interest here at unique points, utilizing the Latin hypercube sampling technique (see §5.4.3). We use these simulations to train emulators in order to 'interpolate' to other points in the parameter space. That is, the emulators will provide a continuous reconstruction of the parameter space, without requiring us to run any additional simulations. With the trained emulators, the search for the best-fitting values of subgrid parameters will be simplified to the minimization of the error between the emulator predictions and the target observational data.

[10] However, unlike Kugel et al. (2023), in our analysis we do not consider the bias parameters for stellar mass and cosmic variance, which Kugel et al. (2023) found to have a negligible effect on the calibration of the FLAMINGO simulations.

5.3.1 Setup

Consider a smooth, surjective mapping f between some input scalar variable x and an output y where the mapping f depends on some parameters[11] represented by a vector $\boldsymbol{\theta}$. Assuming that we know the true relation $y = f(x, \boldsymbol{\theta})$ only at some *finite* number of points N, $\{x_n, \boldsymbol{\theta}_n, y_n = f(x_n, \boldsymbol{\theta}_n)\}_{n=1}^{N}$, our goal will be to make use of this limited information to interpolate the relation $y = f(x, \boldsymbol{\theta})$ to the unknown parts of the space along $\boldsymbol{\theta}$ and x. If the data $\{x_n, \boldsymbol{\theta}_n, y_n\}_{n=1}^{N}$ come from a numerical simulation, as will be the case for us, then the interpolation process is termed *emulation*.

We will write a hat above 'f' to distinguish an emulator of $y = f(x, \boldsymbol{\theta})$, $y = \hat{f}(x, \boldsymbol{\theta})$, from the true relation. The size of vector $\boldsymbol{\theta}$ is equal to the number of parameters on which the emulator depends. In the following, we will write $N_{\rm param}$ as a short-hand notation for the length of $\boldsymbol{\theta}$.

A Gaussian process with zero mean is fully specified by its covariance function (e.g. Rasmussen & Williams, 2006). As in Kugel et al. (2023), we construct the covariance function by using the squared exponential kernel,

$$k(\boldsymbol{X}, \boldsymbol{X}') = \exp\left[-\frac{(\boldsymbol{X} - \boldsymbol{X}')^{\rm T}\mathbf{C}(\boldsymbol{X} - \boldsymbol{X}')}{2}\right], \tag{5.21}$$

where the vectors $\boldsymbol{X} = (x, \boldsymbol{\theta})$ and $\boldsymbol{X}' = (x', \boldsymbol{\theta}')$ correspond to two different points in the $N_{\rm param} + 1$-dimensional parameter space, and $\mathbf{C}$ is a diagonal matrix that sets the length scale for each dimension of the parameter space. We do not need to opt for more sophisticated kernels because the relations we are going to emulate are smooth and their behaviour is relatively straightforward to predict once the emulators have been trained.

5.3.2 The emulated relations

We emulate two relations that will be used to calibrate the COLIBRE model:

- *Galaxy stellar mass function* (GSMF) at $z = 0$. Here the input variable x is the galaxy stellar mass, M_*, measured in 3D apertures of size 50 kpc, and the output variable y is the number of galaxies per unit volume, $\mathrm{d}n$, per logarithmic bin of stellar mass, $\mathrm{d}\log_{10} M_*$. Because the stellar mass M_* can span many orders of magnitude, to increase the accuracy of the emulation we will work in log scale, adopting $x = \log_{10} M_*$ as opposed to $x = M_*$. Likewise, for the output y, we will take $y \equiv f(x) = \log_{10}(\mathrm{d}n/\mathrm{d}\log_{10} M_*)$, as opposed to $\mathrm{d}n/\mathrm{d}\log_{10} M_*$.

[11] As we will show later in §5.4.2, these extra parameters will be combinations of various subgrid parameters of the COLIBRE model.

- *Size-stellar mass relation* (SSM) *at* $z = 0$. Here the input is again $x = \log_{10} M_*$, while the output is the median projected stellar half-mass radius of the simulated galaxies whose stellar mass is M_*.

Our stellar mass, M_*, is computed in 3D spherical apertures with a radius of 50 proper kpc. By conducting mock observations of galaxies from the EAGLE simulations, de Graaff et al. (2022) found that this choice of aperture yields results similar to the masses inferred from fitting Sérsic profiles, which is a method commonly used for observations. For both emulated relations, we take the x values from the simulations and arrange them in bins of equal size of $\Delta \log_{10}(M_*/\mathrm{M}_\odot) = 0.2$. For the SSM relation, we then compute the median projected galaxy half-mass size in each bin, while for the GSMF we count the number of objects in each mass bin and divide it by the simulated volume and by the logarithmic width of the bin.

We account for the errors on y values as follows. For the SSM, we set the error, Δy, to be half of the difference between the 84^{th} and 16^{th} percentiles of the distribution of the y values in each stellar-mass bin. For the GSMF the error is given by the Poisson noise.

5

- Additionally, we emulate the $z = 0$ relation between the *stellar mass to halo mass ratio* and halo mass (SMHM). This relation will *not* be used to calibrate the COLIBRE model because the SMHM is poorly constrained observationally. Instead, we will use it for diagnostic purposes. Our halo mass, M_{vir}, is defined using the spherical overdensity threshold of Bryan & Norman (1998). The x and y values fed to the emulator are $x = \log_{10} M_{\mathrm{vir}}$ and $y = \log_{10}(M_*/M_{\mathrm{vir}})$. The y values are the median SMHM values computed in halo-mass bins with a 0.2-dex width and the error on the y values equals half the difference between the 84^{th} and 16^{th} percentiles of the SMHM values of individual simulated galaxies falling in each bin. In the computation of the SMHM, we consider only central sub-haloes.

5.3.3 Target observational data

We will now describe the observational data to which the GSMF and SSM emulators from §5.3.2 will be fit.

Galaxy stellar mass function at $z = 0$

The GSMF provides one of the most stringent constraints on the evolution of stellar mass in the Universe: it determines not only the total stellar mass formed in the Universe but also the relative abundance of low- and high-mass galaxies. We constrain the simulated $z = 0$ GSMF by matching it to the $z = 0$ GSMF from Driver et al. (2022). The Driver et al. (2022) GSMF was derived from the Galaxy And Mass Assembly, Data Release 4 (GAMA DR4) survey, which provides more

than 240000 galaxy spectra in the redshift range from $z \approx 0$ to 1. The Driver et al. (2022) GSMF is not only highly robust but it is also presented exactly at $z = 0$ as the authors correct for the redshift evolution.

Galaxy size-stellar mass relation at $z = 0$

Reproducing the observed GSMF does not guarantee that other properties of the simulated galaxies, besides stellar mass, will be realistic. For example, while the GSMF provides constraints on the total stellar masses of galaxies, it says nothing about how that stellar mass is distributed within the galaxies. Indeed, Crain et al. (2015) showed that depending on the subgrid model adopted in the numerical simulation, simulated galaxies may be described by the same GSMF, but differ drastically in their stellar half-mass radii.

We take the galaxy stellar mass-size relation from Hardwick et al. (2022) as a second constraint on our simulations. The Hardwick et al. (2022) data come from the eXtended GALEX Arecibo SDSS Survey (xGASS) survey (Catinella et al., 2018), which contains ~ 1200 galaxies in the redshift range $0.01 < z < 0.05$ with a flat distribution of stellar masses between $M_* \approx 10^9$ and $10^{11.5}\,\mathrm{M}_\odot$. Hardwick et al. (2022) provide the median galaxy half-mass radii in stellar-mass bins of 0.2 dex width. The half-mass sizes are estimated based on Sérsic profile fits, and are smaller by ≈ 0.1 dex than the corresponding half-light sizes in the r-band, in agreement with de Graaff et al. (2022) who found a similar difference between mass- and light-weighted galaxy sizes in the EAGLE simulations, using the SDSS pipeline for the analysis.

5.3.4 The search for the best-fitting parameter values

Each emulated relation $y = \hat{f}(x, \boldsymbol{\theta})$ changes continuously and smoothly with x and $\boldsymbol{\theta}$. Once we have constructed $y = \hat{f}(x, \boldsymbol{\theta})$ by training the emulator on the simulation data $\{x_n, \boldsymbol{\theta}_n, y_n\}_{n=1}^N$, our goal is to find the values of the parameters $\boldsymbol{\theta}$ that result in the best agreement between the emulator's predictions and observational data. To quantify how well Gaussian process emulators can fit the observational data, we will use Bayesian analysis.

Prior

We start by setting up a prior on the emulator parameters $\boldsymbol{\theta}$, $\mathcal{P}_{\mathrm{prior}}(\boldsymbol{\theta})$. We assume that each parameter θ_i has a uniform prior within some range from $\theta_{i,\mathrm{min}}$ to $\theta_{i,\mathrm{max}}$, and drops to zero otherwise,

$$\mathcal{P}_{i,\mathrm{prior}}(\theta_i) = \begin{cases} 1, & \text{if } \theta_{i,\mathrm{min}} \leq \theta_i \leq \theta_{i,\mathrm{max}} \\ 0, & \text{otherwise}. \end{cases} \tag{5.22}$$

The total prior for $\boldsymbol{\theta}$ is then the product $\mathcal{P}_{\rm prior}(\boldsymbol{\theta}) = \prod_i^{N_{\rm param}} \mathcal{P}_{i,\rm prior}$.

We opt for such a prior because of our limited knowledge about the parameters $\boldsymbol{\theta}$. The vector $\boldsymbol{\theta}$ contains subgrid parameters of our galaxy formation model. Probing N random realizations of vector $\boldsymbol{\theta}$ requires running N independent simulations, each of which may take a long time ($\sim$ weeks) to complete. Therefore, given the finite number of simulations that we can afford to run, our training data contain a rather small number of values for each subgrid parameter θ_i. The values of θ_i are distributed within a certain interval whose lower and upper ends define, respectively, $\theta_{i,\rm min}$ and $\theta_{i,\rm max}$ in equation (5.22). We have to set the prior to zero outside the domain sampled by the actual simulations because the errors of a Gaussian process emulator become large if the emulator is used for extrapolation.

Likelihood

We compute the total log-likelihood function, $\ln \mathcal{L}(\boldsymbol{\theta})$, as the sum of individual log-likelihood functions for the emulated GSMF and SSM,

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \ln \mathcal{L}_{\rm GSMF}(\boldsymbol{\theta}) + \ln \mathcal{L}_{\rm SSM}(\boldsymbol{\theta}), \tag{5.23}$$

which means that the GSMF and SSM relation contribute equally to the total likelihood.

The likelihood of each emulated relation is computed assuming that the statistical errors in the emulator's prediction and the observational data are Gaussian distributed and independent,

$$\ln \mathcal{L}_{\rm R}(\boldsymbol{\theta}) = -\frac{\langle N_{\rm obs}\rangle}{N_{\rm R,obs}} \frac{1}{2} \sum_{n=1}^{N_{\rm R,obs}} \left[\frac{\hat{f}_{\rm R}(x_{{\rm R},n}, \boldsymbol{\theta}) - y_{{\rm R},n}}{\sqrt{\sigma_{{\rm R},n}^2 + \varepsilon_{\rm emu}^2}} \right]^2, \tag{5.24}$$

where the subscript 'R' stands for GSMF or SSM. Next, $x_{{\rm R},n}$, $y_{{\rm R},n}$, and $\sigma_{{\rm R},n}$ are, respectively, the x values, y values, and errors on the y values of the observational data used to constrain the emulated relation 'R' (see §5.3.3), and $N_{\rm R,obs}$ is the number of observational data points over which the sum is computed. $\hat{f}_{\rm R}(x_{{\rm R},n}, \boldsymbol{\theta})$ is the prediction of the emulator of the relation 'R' evaluated at $x_{{\rm R},n}$ and for the parameter vector $\boldsymbol{\theta}$. The SSM and GSMF likelihood functions are normalized by $N_{\rm R,obs}/\langle N_{\rm obs}\rangle$ where $\langle N_{\rm obs}\rangle = (N_{\rm GSMF,obs} + N_{\rm SSM,obs})/2$ is the average number of data points contained in the observational data for the GSMF and SSM emulators. This normalization ensures that having different numbers of data points for the GSMF and SSM datasets has no impact on the relative contributions of the GSMF and SSM to the total likelihood. Note that since the emulators are built using the simulation data in log-log space (see §5.3.2), the x and y values of the observational data in Eq. (5.24), as well as the errors on the observed y values, are logarithmic too.

Finally, $\varepsilon_{\rm emu}$ is the uncertainty in the emulator predictions. As will be described in §5.4.1, we build emulators not only for the full COLIBRE model but also for three variations that simplified prescriptions for SN feedback. To estimate $\varepsilon_{\rm emu}$, we train the emulators on all but one simulation from the training data of a given model, and ask the emulator to predict the GSMF and SSM for the simulation that was left out. We repeat this procedure for each simulation in the training data of each model and record the differences between the emulator predictions and the actual simulation data. This gives us $2\,N_{\rm runs}^{\rm all\,models}$ vectors with emulator errors where entries of each vector correspond to different stellar mass bins and $N_{\rm runs}^{\rm all\,models} = 136$ is the total number of simulations in the training data from the four models (see §5.4.3). The factor of 2 accounts for the fact that we collect the errors for two emulated relations: GSMF and SSM. We concatenate all vectors into a single list and compute its standard deviation. We found that the emulator errors for the GSMF and SSM are both close to 0.06 dex, so we will set $\varepsilon_{\rm emu}$ to this value for both the GSMF and SSM relation.

Posterior

The log posterior is the sum of the log likelihood and the log prior,

$$\ln \mathcal{P}_{\rm posterior}(\boldsymbol{\theta}) = \ln \mathcal{L}(\boldsymbol{\theta}) + \ln \mathcal{P}_{\rm prior}(\boldsymbol{\theta}), \tag{5.25}$$

from which we obtain the values of the parameters of the best-fitting model, $\boldsymbol{\theta}_{\rm best}$, as

$$\ln \mathcal{P}_{\rm posterior}(\boldsymbol{\theta}_{\rm best}) = \max\Big(\ln \mathcal{P}_{\rm posterior}(\boldsymbol{\theta})\Big). \tag{5.26}$$

To find the maximum of the posterior distribution, $\ln \mathcal{P}_{\rm posterior}(\boldsymbol{\theta})$, we use the Markov chain Monte Carlo (MCMC) python package EMCEE (Foreman-Mackey et al., 2013). We run MCMC for 5000 steps using 30 independent walkers that start at random positions in the parameter space. In the analysis of the posterior distribution, we remove the first 200 steps for each walker to avoid the 'burn-in' phase. In order to generate proposal steps for the random walk through the parameter space, we employ the 'stretch move' algorithm (Goodman & Weare, 2010) with a stretch scale parameter of 2. Lastly, we note that it is not necessary to normalize the posterior to find the best-fitting parameter values, as can be seen from equation (5.26).

5.4 Calibration with emulators

This section describes the calibration strategy of the COLIBRE simulations, which makes use of the method of emulators detailed above. The objective of the calibration is to maximize the agreement between the simulation and target observational data, which is achieved by adjusting the subgrid parameters of the model, or the model itself.

5.4.1 Models with simplified supernova feedback

As the starting point of the calibration, we take a model of galaxy formation with a significantly simplified version of SN feedback, compared to the fiducial COLIBRE algorithm presented in §5.2.8. We will call this model the Basic model. The other aspects of the galaxy formation physics in the Basic model will be the same as described in Section 5.2.

We do not commence with calibrating the full COLIBRE algorithm for SN feedback because it is not obvious *a priori* whether using a more complex model will lead to a better fit of the simulation to the observational data. Only failing to match the target observational data with the simplified model will indicate that a more sophisticated model might be a necessity.

Besides the Basic model, we will consider two other simplified prescriptions for SN feedback, ThermalKinetic and ThermalKineticVariableΔT, all of which will be detailed below. We emphasize that only the treatment of SN feedback is different between the fiducial COLIBRE model and its simplified versions – Basic, ThermalKinetic, and ThermalKineticVariableΔT – while all other parts of the galaxy formation physics remain identical. In particular, the BH physics is kept fixed.

The Basic model

Compared the fiducial COLIBRE model for SN feedback from Section 5.2, in the Basic model we make the following simplifications:

(i) The energy of a single SN in units of 10^{51} erg, $f_{\rm E}$, is constant, as opposed to being dependent on the stellar birth pressure, $P_{\rm birth}$ (see Eq. 5.8);

(ii) All energy released by SNe is injected thermally. In other words, the fraction of SN energy injected in kinetic form, $f_{\rm kin}$, is set to 0 and the kinetic channel of SN feedback is not used;

(iii) Instead of being density-dependent (Eq. 5.14), the heating temperature in the thermal channel of SN feedback, $\Delta T_{\rm SN}$, is constant and equal to $10^{7.5}$ K, the value used in the EAGLE model.

The thermal-kinetic model

In addition to the Basic model, we consider a modification in which the prescription for SN feedback consists of both kinetic and thermal channels of energy injection (i.e. $f_{\rm kin}$ is no longer necessarily 0). We will call this model ThermalKinetic. As is the case in the COLIBRE reference model, the kinetic channel of the ThermalKinetic model uses the desired kick velocity of $\Delta v_{\rm kick} = 50$ km s^{-1}. Otherwise, ThermalKinetic is the same as Basic, including the constant energy in SN feedback and the constant heating temperature of $\Delta T_{\rm SN} = 10^{7.5}$ K in the thermal channel of energy injection.

Compared to Basic, the ThermalKinetic model has one extra free parameter: the fraction of SN energy injected in kinetic form, $f_{\rm kin}$. For $f_{\rm E} = 2$ and $f_{\rm kin} = 0.1$, the CC SN feedback in the ThermalKinetic model becomes identical to that in the fiducial model used in the simulations of isolated disc galaxies by Chaikin et al. (2023).

The thermal-kinetic model with a variable heating temperature

Our final simplified model is ThermalKineticVariableΔT. As the name suggests, compared to ThermalKinetic, ThermalKineticVariableΔT adopts the density-dependent heating temperature for thermal SN feedback (for both CC and type-Ia SNe) detailed in §5.2.8, while Basic and ThermalKinetic use a constant value of $\Delta T_{\rm SN} = 10^{7.5}$ K.

Of the simplified models, ThermalKineticVariableΔT is closest to the COLIBRE reference model. The only difference between the two models is that in the latter one the SN energy $f_{\rm E}$ depends on the stellar birth gas pressure, following equation (5.8), while ThermalKineticVariableΔT uses a constant $f_{\rm E}$.

5.4.2 Selection of subgrid parameters for calibration with emulators

We next describe the selection of the subgrid parameters that will be calibrated using the emulators. These parameters, which we denote by the vector $\boldsymbol{\theta}$, enter the emulators defined in §5.3.2 and are optimized with the methods of Bayesian statistics (§5.3.4), such that the simulation provides the best match to the observational data from §5.3.3.

We will only examine the subgrid parameters that are important for SN and AGN feedback. We will consider no parameters related to other parts of the galaxy formation physics (such as star formation, chemical enrichment, or radiative cooling) because those parameters will have a smaller-to-no impact on the galaxy properties that are relevant to our calibration (i.e. GSMF and SSM), and/or because their values are well constrained by the fundamental physics or inferred from other observations.

AGN feedback parameters

The notable parameters of the COLIBRE model related AGN feedback are (i) the AGN heating temperature, $\Delta T_{\rm AGN}$; (ii) the seed mass of SMBHs, $M_{\rm BH,seed}$; and (iii) the minimum mass of a halo in which BH particles can be seeded, $M_{\rm FOF,seed}$.

Table 5.1: Latin hypercubes used to train the emulators. Column (1): the name of the model for which the Latin hypercube is created (see §5.4.1); column (2): the number of simulations in the Latin hypercube; column (3): the energy per single CC SN in units of 10^{51} erg; column (4): the fraction of SN energy that is injected in kinetic form; column (5): the pivot density in the relation between the SN heating temperature and the gas density (Eq. 5.14); column (6) the pivot birth pressure in the relation between the energy in CC SN feedback and the stellar birth gas pressure (Eq. 5.8); column (7) the seed mass of SMBHs. The two numbers in each cell of columns 3 – 7 specify the interval over which each parameter is varied in the Latin hypercube. For a given model, an empty cell means that the corresponding parameter does not exist in the model.

Model name	$N_{\rm runs}$	Model parameters that are varied in the Latin hypercube				
		$f_{\rm E}$	$f_{\rm kin}$	$n_{\rm SN,0}$ [cm^{-3}]	$\log_{10} P_{\rm birth,0}/k_{\rm B}$ [K cm^{-3}]	$\log_{10} M_{\rm BH,seed}$ [M$_\odot$]
Basic	24	$[0.1, 5]$	–	–	–	$[3, 5]$
ThermalKinetic	32	$[0.5, 2.5]$	$[0, 1]$	–	–	$[3, 5]$
ThermalKineticVariableΔT	40	$[0.5, 2.5]$	$[0, 0.5]$	$[0.1, 3]$	–	$[3.8, 4.8]$
Reference (COLIBRE)	40	–	$[0, 0.5]$	$[0.05, 2]$	$[3.2, 4.4]$	$[3.7, 4.7]$

- The AGN heating temperature, $\Delta T_{\rm AGN}$, determines the thermal energy received by a gas particle in a single AGN energy injection event. In other words, $\Delta T_{\rm AGN}$ is a measure of the 'burstiness' of AGN feedback. In principle, higher (lower) values of $\Delta T_{\rm AGN}$ are expected to yield stronger (weaker) AGN feedback. However, owing to the ability of SMBHs to self-regulate, the exact value of $\Delta T_{\rm AGN}$ may have only a small impact on the $z = 0$ GSMF and SSM[12] (e.g. McCarthy et al., 2017), as long as the temperature of the heated gas remains high enough that the injected thermal energy is not quickly radiated away (e.g. Booth & Schaye, 2009). We will therefore not consider $\Delta T_{\rm AGN}$ as one of the subgrid parameters for calibration. In §5.5.4 we will confirm that (modest) variations in $\Delta T_{\rm AGN}$ indeed have only a minor effect on the $z = 0$ GSMF and SSM, and that these small differences can be compensated for by adjusting other model parameters.

- The SMBH seed mass, $M_{\rm BH,seed}$, determines how quickly SMBHs can grow over time. Higher $M_{\rm BH,seed}$ will cause faster SMBH growth, leading to more energetic AGN feedback in lower-mass galaxies and at higher redshifts (e.g. Booth & Schaye, 2009). Because both the GSMF and SSM depend sensitively on the strength of AGN feedback, we will include $M_{\rm BH,seed}$ in the set of subgrid parameters for optimization with emulators, $\boldsymbol{\theta}$.

- The minimum halo (friends-of-friends) mass in which SMBHs are seeded, $M_{\rm FOF,seed}$, has a prominent effect on the calibrated relations too, because, similarly to $M_{\rm BH,seed}$, $M_{\rm FOF,seed}$ determines how fast SMBHs can start growing in mass (e.g. Booth & Schaye, 2009). However, for the same reasons, $M_{\rm FOF,seed}$ is strongly degenerate with $M_{\rm BH,seed}$. For example, increasing $M_{\rm FOF,seed}$ will delay the growth of SMBHs, but a similar effect can be achieved by decreasing $M_{\rm BH,seed}$. Owing to this degeneracy, which will be demonstrated in §5.5.4, we will not include $M_{\rm FOF,seed}$ in our set of parameters for optimization. As already explained in §5.2.10, we set $M_{\rm FOF,seed} = 10^{10}\,{\rm M}_\odot$.

Supernova feedback parameters

Because we consider four different prescriptions for SN feedback, we will, for clarity, describe the SN feedback parameters for each model separately.

- In the Basic model, the only free parameter is the energy per single SN in units of 10^{51} erg, $f_{\rm E}$. Therefore, the full parameter vector $\boldsymbol{\theta}$ for the

[12] In contrast, $\Delta T_{\rm AGN}$ will have a strong effect on gas fractions within galaxy clusters (e.g. Le Brun et al., 2014; Kugel et al., 2023), which we, however, cannot incorporate as an additional constraint in the emulation because of the limited size of the simulations in our training set (50^3 comoving Mpc^3). By re-running a single simulation with the fiducial COLIBRE model in the 100^3 comoving Mpc^3 volume, we verified that the value of $\Delta T_{\rm AGN} = 10^9$ K gives reasonable cluster gas fractions.

Basic model, including the AGN feedback parameters from §5.4.2, is $\boldsymbol{\theta} = (M_{\rm BH,seed}, f_{\rm E})$.

- The ThermalKinetic model contains an additional free parameter: the fraction of SN energy injected in kinetic form, $f_{\rm kin}$. This makes the total number of parameters entering the parameter vector $\boldsymbol{\theta}$ equal to three: $\boldsymbol{\theta} = (M_{\rm BH,seed}, f_{\rm E}, f_{\rm kin})$.

- The ThermalKineticVariableΔT model employs a density-dependent heating temperature for SN feedback (Eq. 5.14), which is described by four parameters: the pivot gas density $n_{\rm SN,0}$, the heating temperature at the pivot density $\Delta T_{\rm SN,0}$, and the minimum and maximum heating temperatures, $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$. As documented in §5.2.8, we set $\Delta T_{\rm SN,0}$, $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$ to $10^{6.5}$, $10^{6.5}$, and $10^{7.5}$ K, respectively, which leaves us with only one free parameter: $n_{\rm SN,0}$. Therefore, the final form of the parameter vector $\boldsymbol{\theta}$ for the ThermalKineticVariableΔT model is $\boldsymbol{\theta} = (M_{\rm BH,seed}, f_{\rm E}, f_{\rm kin}, n_{\rm SN,0})$.

- Finally, the reference COLIBRE model (henceforth, Reference) uses a stellar birth pressure dependent energy for CC SN feedback (Eq. 5.8). This comes with another set of four parameters, $f_{\rm E,min}$, $f_{\rm E,max}$, $P_{\rm birth,0}$ and σ_P, which together replace the parameter $f_{\rm E}$ that specifies the constant energy in SN feedback in the three simplified models. As explained in §5.2.8, the values of three out of the four extra parameters are fixed: $f_{\rm E,min} = 0.1$, $f_{\rm E,max} = 3.5$, and $\sigma_P = 0.3$. Thus, in the emulation, the dependence of $f_{\rm E}$ on the stellar birth pressure will be described with a single free parameter, $P_{\rm birth,0}$, resulting in the parameter vector for the Reference model, $\boldsymbol{\theta} = (M_{\rm BH,seed}, P_{\rm birth,0}, f_{\rm kin}, n_{\rm SN,0})$.

  We note that due to the functional degeneracies between $P_{\rm birth,0}$, $f_{\rm E,min}$, and $f_{\rm E,max}$ (see Eq. 5.8), a very low (high) $P_{\rm birth,0}$ preferred by the emulator will indicate that our chosen value of $f_{\rm E,max}$ ($f_{\rm E,min}$) is too low (high). In §5.5.4 we will show that $P_{\rm birth,0}$, $f_{\rm E,min}$, and $f_{\rm E,max}$ are indeed degenerate with one another, and that the exact value of σ_P has only a weak impact on the GSMF and SSM.

5.4.3 Training data for emulators

For the fiducial COLIBRE model from Section 5.2, as well as for each of the simplified versions described in §5.4.1, we create three emulators introduced in §5.3.2. Each emulator needs to be trained before it can be exploited in the search for the best-fitting parameter values.

To build the training datasets, we run actual simulations. For a given model, each simulation will represent a unique combination of values of the subgrid

parameters $\boldsymbol{\theta}$. To optimally choose the locations of the simulations in the parameter space, we make use of the *Latin hypercube* sampling technique (McKay et al., 1979). The main advantage of Latin hypercube sampling compared to naive random sampling is that the former requires a much smaller number of sample points to achieve the desired accuracy of the emulator predictions. Therefore, the number of actual simulations needed to train the emulators for each model can be relatively modest. We choose to run $N_{\rm runs} = 24$ simulations for the Basic model, 32 for the ThermalKinetic model, and 40 for ThermalKineticVariableΔT and Reference. The number of simulations in the Latin hypercube increases with the size of the parameter vector $\boldsymbol{\theta}$, whose dimensions for the four models are, respectively, 2, 3, 4, and 4. The properties of the Latin hypercubes for the four models are summarized in Table 5.1, including the ranges over which the parameters $\boldsymbol{\theta}$ are varied. These ranges have been preselected such that the peaks of the posterior distribution of the parameters $\boldsymbol{\theta}$ fall inside the hypercube domain[13].

All simulations in the Latin hypercubes were run to $z = 0$ in (50 comoving Mpc)^3 volumes with initial numbers of gas and dark-matter particles of 376^3 and 4×376^3, respectively (see §5.2.1 for more details on the ICs). Each simulation was run on 128 cores and took approximately one week to reach $z = 0$. In total, this translates into $\approx 3 \times 10^6$ core hours. The values of the subgrid parameters that are not part of $\boldsymbol{\theta}$ do not change between different simulations from the same Latin hypercube.

As an illustration, in Fig. 5.1 we show the Latin hypercube for the ThermalKinetic model. Different axes correspond to different parameters of the hypercube: $f_{\rm E}$, $M_{\rm BH,seed}$, and $f_{\rm kin}$. The grey hatched rectangle indicates the domain of the Latin hypercube and orange triangles represent 32 individual simulations sampling the parameter space. Thanks to the Latin hypercube sampling, the range of values of each parameter is sampled nearly uniformly despite using only 32 simulations. These simulations are used to train the emulators to predict the $z = 0$ GSMF, SSM, and SMHM as functions of the hypercube's parameters. Once the GSMF and SSM emulators have been trained, we can fit the ThermalKinetic model to the observed GSMF and SSM, as described in §5.3.4, to find the best-fitting values of the model parameters.

The Latin hypercubes for the other three models – Basic, ThermalKineticVariabledT and Fiducial – look qualitatively similar but have different numbers of dimensions.

[13] This information was obtained by running small cosmological volumes (25^3 comoving Mpc^3) for much wider ranges of parameter values.

Table 5.2: The best-fitting values of the parameters identified by the emulator by matching the model to the observational data. Column (1): the name of the model; column (2): the observational data to which the model was fit; column (3): the reduced χ^2 of the fit. The remaining columns indicate the best-fitting values of the model parameters and the corresponding 1σ errors. The model parameters are arranged in the same way as in Table 5.1. For a given model, an empty cell means that the corresponding parameter does not exist in the model.

Model name	Emulator was fit to	Reduced χ^2	Best-fitting values of model parameters					
			$f_{\rm E}$	$f_{\rm kin}$	$n_{\rm SN,0}$ [cm^{-3}]	$\log_{10} P_{\rm birth,0}/k_{\rm B}$ [K cm^{-3}]	$\log_{10} M_{\rm BH,seed}$ [M$_\odot$]
Basic	GSMF and SSM	11.7	1.0 ± 0.1	–	–	–	3.4 ± 0.1
ThermalKinetic	GSMF and SSM	5.5	1.1 ± 0.1	0.37 ± 0.05	–	–	4.2 ± 0.1
ThermalKineticVariableΔT	GSMF and SSM	3.3	1.4 ± 0.1	0.1 ± 0.05	0.5 ± 0.2	–	4.1 ± 0.1
Reference (COLIBRE)	GSMF and SSM	1.1	–	0.1 ± 0.05	0.6 ± 0.1	3.8 ± 0.1	4.2 ± 0.1
ThermalKinetic	GSMF	1.6	1.4 ± 0.1	0.72 ± 0.1	–	–	4.4 ± 0.2
ThermalKinetic	SSM	0.7	0.8 ± 0.1	$0^{+0.05}_{-0}$	–	–	3.8 ± 0.1

5.4.4 Best-fitting parameters and extra simulations

The best-fitting values of the parameters of the COLIBRE reference model, as well as its three simplified analogues, are presented in Table 5.2. For each model, the table lists only the values of those parameters that were optimized by the emulators by fitting the model to the observational data. Additionally, the table provides 1σ errors on the parameter values. Finally, we show the values of reduced χ^2, which indicate the goodness of the fit to the observational data. We compute χ^2 as a sum of the differences (squared) between the predictions of the best-fitting model and the observational data to which the model was fit, normalized by the errors that are used in the likelihood (Eq. 5.24).

We round the best-fitting values of model parameters, as we found no significant gain in the accuracy of the fits by using more precise numbers. The ThermalKinetic model appears three times in Table 5.2, as we fit it to the observed GSMF and SSM separately and together (see §5.5.1).

For each model and each combination of best-fitting parameter values (different rows in Table 5.2), we run a separate numerical simulation, which in total gives us 6 new simulations[14]. In addition, we run 20 variations of the Reference model with the best-fitting parameters where one of the following 10 subgrid parameters is varied: $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,min}$, $n_{\rm SN,0}$, $f_{\rm E,min}$, $f_{\rm E,max}$, $P_{\rm birth,0}$, σ_P, $M_{\rm BH,seed}$, $M_{\rm FOF,seed}$, and $\Delta T_{\rm AGN}$ (note that 7 out of these 10 parameters were not considered in the hypercubes). We will discuss these runs in §5.5.4. Finally, we run one more variation of the Reference model that uses the best-fitting parameters, but in which the energy feedback from type-Ia SNe is switched off. This simulation is discussed in Appendix 5.A. All simulations use the same ICs from §5.2.1.

5.5 Results

We begin this section by assessing the performance of the two simplest best-fitting models: Basic and ThermalKinetic, focussing on their predicted GSMF, SSM, and SMHM relations and testing the accuracy of the emulators relative to the actual simulations (§5.5.1). We then investigate the $z = 0$ GSMF and SSM in the more complex models: ThermalKineticVariabledT and Reference (§5.5.2). Next, we show how the simulations with the four best-fitting models compare to observations for various galaxy properties that were not considered during emulation and to which the models have not been tuned (§5.5.3). We then demonstrate how the calibrated galaxy properties in the Reference model are impacted

[14] Because slight changes in the parameter values do not have a significant effect on the accuracy of the fit to the observational data, we choose to run the simulation with the Reference model for rounder values of the best-fitting parameters when expressed in scientific format. We take $P_{\rm birth,0}/k_{\rm B} = 7\times10^3$ K cm^{-3} instead of $10^{3.8}$ K cm^{-3} and $M_{\rm BH,seed} = 2\times10^4\,{\rm M}_\odot$ instead of $10^{4.2}\,{\rm M}_\odot$. The values of the parameters $n_{\rm SN,0}$ and $f_{\rm kin}$ remain the same as in Table 5.2.

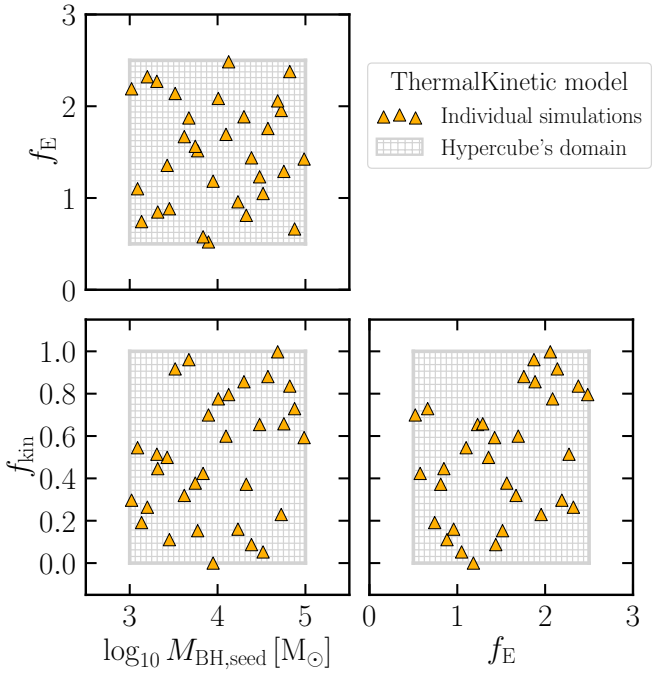

Figure 5.1: The Latin hypercube for the ThermalKinetic model. The panels' axes correspond to different parameters of the model: $f_{\rm E}$, $M_{\rm BH,seed}$, and $f_{\rm kin}$. The grey hatched rectangle indicates the hypercube's domain and orange triangles represent its sampling with 32 individual simulations, which constitute the training dataset of the ThermalKinetic model. These simulations are used to train the emulators of the $z = 0$ GSMF, SSM, and SMHM.

by changes in the subgrid parameters that were not optimized by the emulators. Additionally, we point out degeneracies between those parameters that were optimized and those that were not (§5.5.4).

5.5.1 Basic and ThermalKinetic models

GSMF and SSM with the best-fitting parameters

Fig. 5.2 shows the $z = 0$ GSMF and median SSM for the Basic (green) and ThermalKinetic (orange) models. The dashed curves are the best-fitting predictions of the emulators that were trained on the Latin hypercubes and fit to the observed GSMF from Driver et al. (2022) and the observed SSM from Hardwick et al. (2022). The solid curves are the GSMF and SSM from the simulations that use the best-fitting parameters found by the emulators. The shaded orange region designates the scatter in the simulation with the ThermalKinetic model:

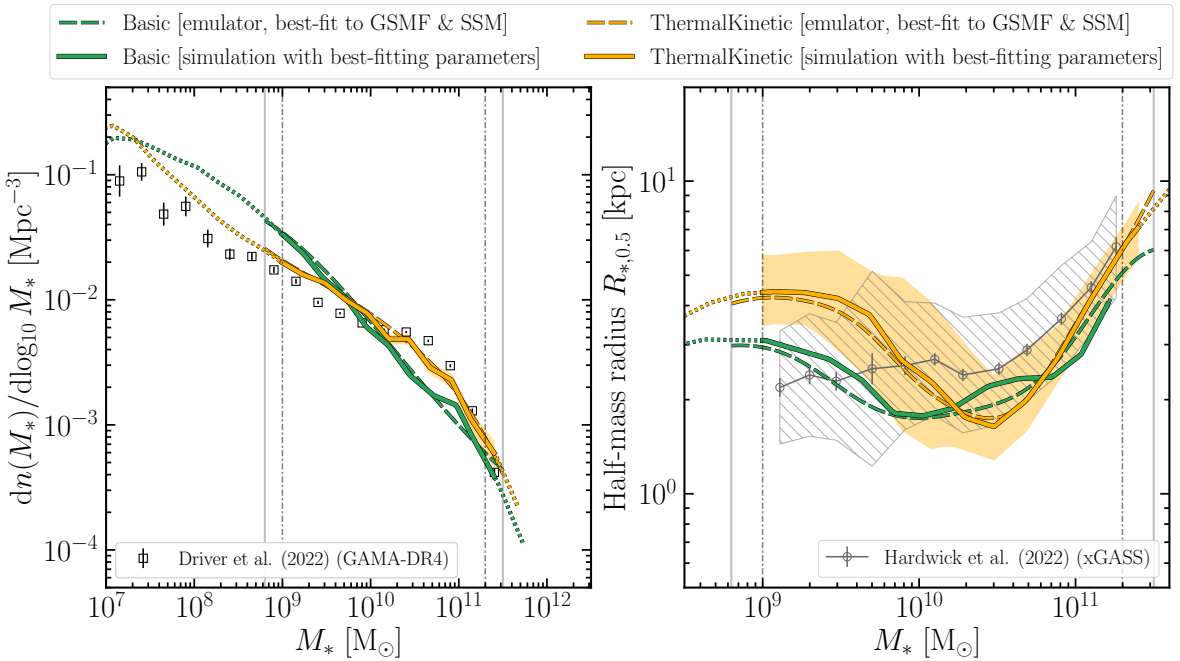

Figure 5.2: The galaxy stellar mass function (GSMF; *left*) and median size-stellar mass relation (SSM; *right*) at $z = 0$, for the Basic (green) and ThermalKinetic (orange) models fit to the observed GSMF and SSM. The dashed and solid curves indicate, respectively, the best-fitting predictions of the emulator and the corresponding simulations with the best-fitting parameters. The shaded orange region shows the scatter in the GSMF and SSM for the simulation with the ThermalKinetic model. We convert the green and orange solid curves into dotted curves where galaxies are unresolved ($M_* < 10^9\,\mathrm{M}_\odot$) and where the number of galaxies is strongly limited by the finite simulated volume ($M_* > 10^{11.4}\,\mathrm{M}_\odot$). The vertical solid (dash-dotted) lines show the mass range within which the emulators were trained on the simulations (fit to observational data). The target observational data from Driver et al. (2022) and Hardwick et al. (2022) are shown in grey-scale colours. Although the ThermalKinetic model produces a combined fit to the observed GSMF and SSM that is better than the fit with the Basic model, neither model is particularly satisfactory.

the Poisson error for the GSMF and the 16$^{\text{th}}$ to 84$^{\text{th}}$ percentile scatter for the SSM. We change the style of the solid curves to 'dotted' in the stellar mass range where galaxies become poorly resolved ($M_* < 10^9\,\mathrm{M}_\odot$) and where the number of galaxies becomes too small due to the finite simulated volume ($M_* > 10^{11.4}\,\mathrm{M}_\odot$). The vertical solid lines indicate the edges of the stellar mass interval used in the training of the emulators. The vertical dash-dotted lines show the mass range within which the trained emulators were fit to the observational data. The observed GSMF from Driver et al. (2022) is shown in the left panel as black squares and the observed SSM from Hardwick et al. (2022) is shown in the right panel as dark-grey circles connected by a line. The error bars in the observed GSMF indicate Poisson scatter, while in the observed SSM, they show the 1σ error on the median. The grey hatched region in the right panel additionally shows the galaxy population-wide scatter in the SSM from Hardwick et al. (2022).

By comparing the solid curves to the dashed curves of the same colour, we find that the differences between the GSMF and SSM predicted by the emulators and resulting from the actual simulations are negligibly small. Specifically, there are no systematic differences between the simulations and emulators, and the emulator errors in different stellar-mass bins of the SSM and GSMF range between 0 and ~ 0.1 dex, which is comparable to the intrinsic scatter of numerical simulations due to their stochastic nature (e.g. Borrow et al., 2023). In other words, our trained emulators can predict the GSMF and SSM from actual simulations both accurately and robustly.

By comparing the solid curves to the black squares in the left panel and the dark-grey circles in the right panel, we find that the ThermalKinetic model is closer to the observational data than the Basic model is. In particular, the Basic model severely underpredicts the number of galaxies with stellar masses around $M_* \sim 10^{10.5}\,\mathrm{M}_\odot$ and overpredicts it at $M_* \lesssim 10^{9.5}\,\mathrm{M}_\odot$. In fact, the shape of the Basic model's GSMF resembles a power-law, which disagrees with the shape of the observed GSMF, which is known to be described by single- or double-component Schechter (1976) functions, featuring an exponential down-turn at high stellar mass. Although the ThermalKinetic model matches the observed GSMF significantly better than Basic, the discrepancy between its GSMF and the observed data is still quite large. Moreover, both models perform poorly in matching the observed galaxy sizes: the SSM in ThermalKinetic features a prominent dip around $M_* \sim 10^{10.5}\,\mathrm{M}_\odot$, while in the Basic model, the sizes of galaxies with stellar mass $M_* \gtrsim 10^{9.5}\,\mathrm{M}_\odot$ are consistently lower than the observed relation by ≈ 0.15 dex.

Overall, the combined fit to the observed GSMF and SSM is better in the ThermalKinetic model than in Basic, but still not satisfactory.

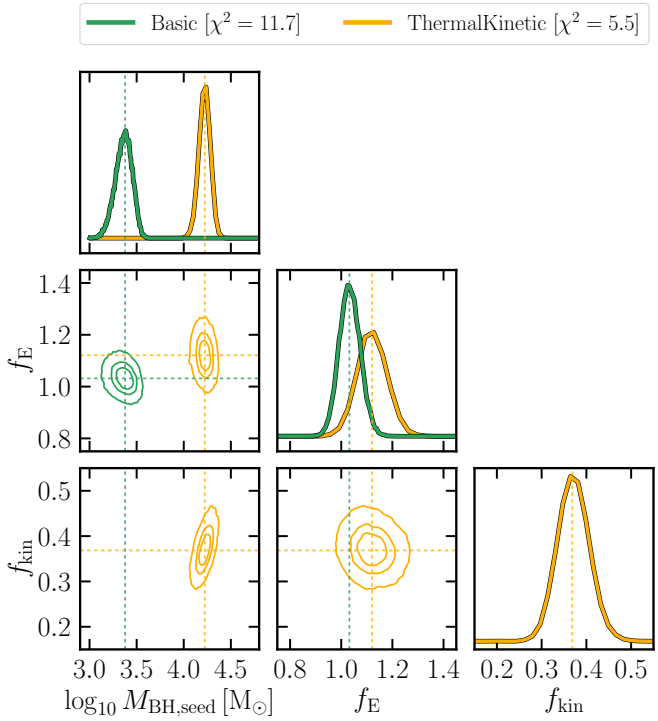

Figure 5.3: Posterior distribution of the parameters for the Basic (green) and ThermalKinetic (orange) models fit to the observed $z = 0$ GSMF and SSM. The values of reduced χ^2 of the fits are shown in the legend. The three contours of each colour indicate 34, 68, 95 per cent credibility levels. The vertical and horizontal dotted lines indicate the best-fitting values of the model parameters, corresponding to the maximum of the posterior.

Posterior distributions of the model parameters

Fig. 5.3 shows the posterior distributions of the parameters of the Basic (green) and ThermalKinetic (orange) models resulting from fitting the emulators to the observed GSMF and SSM, as explained in §5.3.4. The three contours of the same colour signify the 34, 68, 95 per cent credibility levels of the posterior. Additionally, we show one-dimensional projections of the posterior distribution for each subgrid parameter. Because the Basic model does not include the kinetic channel of SN feedback, this model is not displayed in the bottom row where the values of the kinetic feedback-related parameter, $f_{\rm kin}$, are plotted.

First, we notice that the regions of the parameter space inspected with the Latin hypercubes contain the models' best-fitting parameters within 2σ (the outer contours). This is reassuring in that it implies our results are not driven

by the sharp edges of the chosen prior. Second, both models prefer the value of the dimensionless SN energy parameter $f_{\rm E}$ to be around 1, implying that a single CC SN releases $\sim 10^{51}$ erg of energy, which is in agreement with the standard theoretical expectations. Third, the best-fitting ThermalKinetic model has a BH seed mass of $M_{\rm BH,seed} \approx 10^{4.2}\,{\rm M}_\odot$, whereas for the Basic model, $M_{\rm BH,seed}$ is almost one order of magnitude lower, $M_{\rm BH,seed} \approx 10^{3.4}\,{\rm M}_\odot$. This is likely because the prescription for SN feedback in the Basic model is too simple, such that the model's only way to find a better agreement with the observed GSMF at the massive end, without making the agreement at the low-mass end even worse, is to reduce the importance of AGN feedback. The emulator achieves this via lowering $M_{\rm BH,seed}$, as there is no other AGN feedback-related parameter to tune. Fourth, the posterior of the ThermalKinetic model peaks at the kinetic energy fraction in SN feedback of $f_{\rm kin} \approx 0.37$, indicating the importance of SN kinetic feedback in bringing the model closer to the observational data.

Finally, in the legend next to the models' names, we show the values of the reduced χ^2 of their fits to the observed data, which is 11.7 for the Basic model and 5.5 for ThermalKinetic. These values are in line with our conclusions from Fig. 5.2: that the ThermalKinetic model outperforms the Basic model but neither model is a particularly good fit to the data.

5

The effect of changing the model parameters

Fig. 5.4 shows the SMHM relation in the ThermalKinetic model. As previously explained, we do not try to fit the emulator to the SMHM relation because it is not directly observed and ignores satellite galaxies. Here we only use the SMHM to predict how varying the model parameters affects the galaxy stellar mass. Displaying the SMHM as opposed to the GSMF makes it easier to visually distinguish the impact of different subgrid parameters, because, compared to the GSMF, the SMHM relation varies over a smaller dynamical range and includes a characteristic change in the sign of the slope.

In each panel of Fig. 5.4, differently coloured solid curves correspond to different SMHM relations predicted by the emulator in which two of the three model parameters are fixed to their best-fitting values and the remaining parameter is varied. The SMBH seed mass is varied in the left panel, the SN energy in the middle panel, and the fraction of SN energy injected in kinetic form in the right panel. In addition, in the middle panel, we display the SMHM relation predicted by the emulator of the Basic model and how it changes with $f_{\rm E}$ (dashed curves in different colours). The only other parameter of the Basic model, $M_{\rm BH,seed}$ is equal to its best-fitting value, $10^{3.4}\,{\rm M}_\odot$. The SMHM relations from the semi-empirical models of Moster et al. (2018) and Behroozi et al. (2019) are shown for reference only (grey-scale curves with error bars).

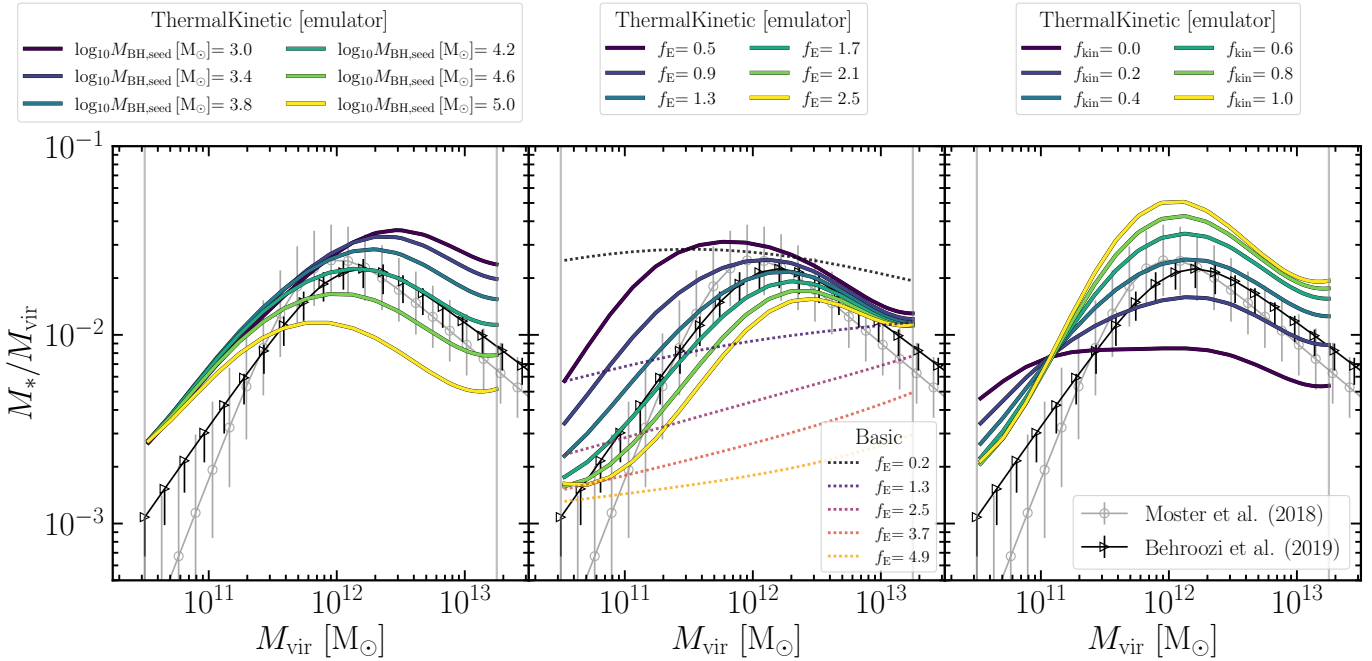

Figure 5.4: Median stellar to halo mass ratios (SMHM) versus halo mass at $z = 0$ predicted by the trained emulators. The results are shown for the ThermalKinetic model fit to the observed GSMF and SSM. The individual panels show how the emulated SMHM varies with the SMBH seed mass $M_{\rm BH,seed}$ (*left*), the energy in SN feedback in units of 10^{51} erg $f_{\rm E}$ (*middle*), and the fraction of SN energy injected in kinetic form $f_{\rm kin}$ (*right*). Different colours correspond to different values of each parameter. Only one parameter is varied at a time, while the other parameters have their best-fitting values. For reference, each panel shows the data from the semi-empirical models of Moster et al. (2018) and Behroozi et al. (2019), displayed in grey-scale colours. Additionally, the middle panel shows the SMHM relation in the Basic model, also for different values of $f_{\rm E}$ (dashed curves). Regardless of the value of $f_{\rm E}$, the SMHM in the Basic model is always too flat compared to the data. This problem is resolved in the ThermalKinetic model for high enough $f_{\rm kin}$.

First, by examining the left panel, we find that, as expected, $M_{\rm BH,seed}$ predominantly affects the stellar mass of massive haloes ($M_{\rm vir} \gtrsim 10^{11.5}\,{\rm M}_\odot$). In haloes of lower mass ($M_{\rm vir} \lesssim 10^{11.5}\,{\rm M}_\odot$), SMBHs do not grow as efficiently, which leads to a lack of AGN feedback therein and hence essentially no dependence of the SMHM relation on $M_{\rm BH,seed}$ at those halo masses. Furthermore, we observe that the value of the SMHM seed mass determines the halo mass at which the SMHM relation reaches its peak. The best-fitting value is $M_{\rm BH,seed} \approx 10^{4.2}\,{\rm M}_\odot$, which results from fitting the ThermalKinetic model's emulator to the GSMF and SSM, yields an SMHM relation that is broadly consistent with the SMHMs inferred from the data by the semi-empirical models of Moster et al. (2018) and Behroozi et al. (2019). This is expected since constraints on the GSMF and SMHM are correlated: fitting the model to either relation should improve the agreement with the other one.

We next move to the middle panel of Fig. 5.4, which shows the effect of varying $f_{\rm E}$. Unlike the left panel, here we display the results for both the ThermalKinetic and Basic models[15]. In essence, increasing (decreasing) $f_{\rm E}$ moves the bulk of the SMHM relation down (up) in both models, as the SN feedback becomes stronger (weaker) leading to less (more) stellar mass formed by $z = 0$. Crucially, there is very little dependence of the shape of the SMHM relation on $f_{\rm E}$, which renders it impossible for the Basic model to match the SMHM of the semi-empirical models solely by adjusting $f_{\rm E}$. Clearly, the change in $M_{\rm BH,seed}$ cannot help either because SMBHs have a small-to-no effect on galaxies in low-mass haloes whose SMHM ratios the Basic model also gets very wrong.

The agreement with the semi-empirical models' SMHMs is strongly improved in the ThermalKinetic model, which exploits the kinetic channel of SN feedback with low-energy kicks, corresponding to the kick velocity of 50 km s^{-1}. The right panel of the figure shows that increasing $f_{\rm kin}$ reduces the galaxy stellar mass at low $M_{\rm vir}$ and increases it at high $M_{\rm vir}$, thereby steepening the slope of the SMHM relation. This helps the ThermalKinetic model obtain a better fit to the observed GSMF, as we have seen in Fig. 5.2, and correspondingly, to the SMHM, as is seen in the current figure. Such a behaviour of the SMHM with $f_{\rm kin}$ can be expected: higher $f_{\rm kin}$ implies that more SN energy is injected kinetically through numerous 50-km s^{-1} kicks and that less energy is distributed thermally via large, rare energy injections corresponding to a gas temperature increase of $\Delta T_{\rm SN} = 10^{7.5}$ K. The kinetic channel is especially efficient in low-mass galaxies, in which the escape velocity is comparable to or lower than the kick velocity used by the kinetic channel. At the same time, the kinetic channel is too weak to push the gas out of more massive objects because of their deeper gravitational potential wells. Conversely, the thermal channel can drive vigorous outflows in galaxies as massive as the Milky Way (Chaikin et al., 2023) but is hindered by poor sampling

[15] Note that the ranges over which $f_{\rm E}$ is varied are different for the two models.

in low-mass objects (see §5.2.8).

Fitting to the observed GSMF and SSM separately and simultaneously

The best-fitting models that we have discussed so far were all fit simultaneously to the observed GSMF and SSM. We now investigate the effect of fitting the models separately to either GSMF or SSM. Mathematically, this means setting the log likelihood function in equation (5.23), $\ln \mathcal{L}(\boldsymbol{\theta})$, to either $\ln \mathcal{L}_{\rm GSMF}(\boldsymbol{\theta})$ or $\ln \mathcal{L}_{\rm SSM}(\boldsymbol{\theta})$, instead of the sum of the two.

Fig. 5.5 compares the Thermalkinetic model with three different sets of the best-fitting parameters, obtained from fitting the emulator to three different sets of observational data: the GSMF (purple), the SSM (brown), or both the GSMF and SSM (orange). The dashed curves indicate the best-fitting predictions of the emulators and the solid curves correspond to simulations with the best-fitting parameters. The shaded orange region shows the 1σ scatter in the simulation whose model was fit to both the GSMF and the SSM relation.

The left panel shows the $z = 0$ GSMF, the middle panel shows the $z = 0$ SSM, and the right panel shows the $z = 0$ SMHM. As in Fig. 5.2, we indicate the range of stellar mass where galaxies become poorly resolved ($M_* < 10^9\,\mathrm{M}_\odot$) or where the number of galaxies becomes strongly limited due to the finite size of the simulated volume ($M_* > 10^{11.4}\,\mathrm{M}_\odot$). In the right panel, which plots the halo mass instead of stellar mass, we use $M_{\rm vir} = 10^{11.15}$ and $10^{13.4}\,\mathrm{M}_\odot$. In each panel, the vertical solid lines indicate the mass range within which the emulators were trained on the Latin hypercubes, while the vertical dash-dotted lines, if present, specify the mass range where the trained emulators were fit to the observational data. As in the previous figures, the comparison data points from Moster et al. (2018), Behroozi et al. (2019), Hardwick et al. (2022), and Driver et al. (2022) are displayed in grey-scale colours.

Examining the left and middle panels of Fig. 5.5, we see that the ThermalKinetic model with the best-fitting parameters matches the observed GSMF (SSM) well if it is fit to the GSMF (SSM) alone. This, however, comes at the expense of a very poor fit to the other observed relation, which was left out of the fitting. In contrast, fitting the model to both the GSMF and SSM at the same time forces the emulator to find a compromise solution. In this case, the best-fitting model provides a mediocre match to the observed GSMF while also being less far off from (but still not close to) the observed SSM (orange). We conclude that although the ThermalKinetic model can match either of the two observed relations well, the model is too limited to be able to reproduce both relations at the same time. To succeed in doing so, a more complex model is required.

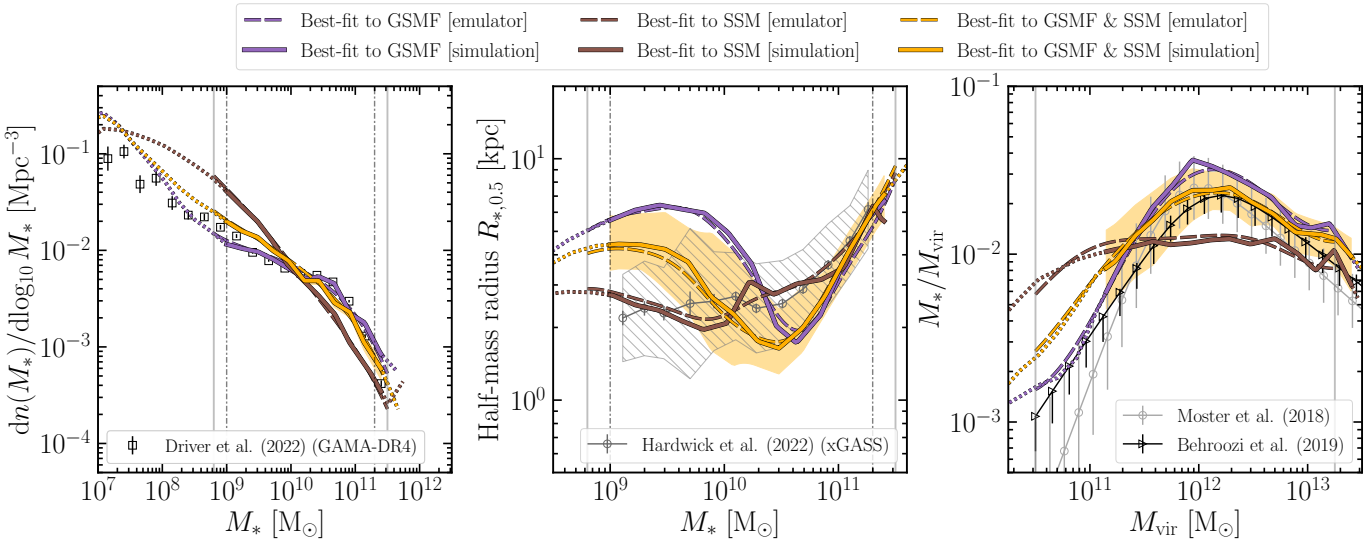

Figure 5.5: Predictions of the best-fitting ThermalKinetic model fit to the observed galaxy stellar mass function (GSMF, purple), galaxy size-stellar mass relation (SSM, brown), or to both the GSMF and SSM (orange). We show the $z = 0$ GSMF (*left*), $z = 0$ SSM relation (*middle*), and $z = 0$ stellar to halo mass ratios versus halo mass (SMHM, *right*). The emulator predictions are shown as dashed curves, and the results from simulations using the best-fitting parameters are shown as solid curves. The solid curves become dotted at stellar (or halo) masses where galaxies are poorly resolved or where the number of galaxies is small due to the finite simulation volume. The vertical solid and dash-dotted lines carry the same meaning as in Fig. 5.2. There are no vertical dash-dotted lines in the right panel because we do not fit the model to the SMHM. Fitting to the observed GSMF (SSM) separately from the SSM (GSMF) results in a good match to the observed GSMF (SSM) but a poor match to the SSM (GSMF). Fitting to both observed relations at the same time produces only a reasonable match to the two constraints.

The right panel of Fig. 5.5, which shows the SMHM relation, confirms what we have seen in the left panel, but the differences between the different cases appear more striking. The galaxies in the ThermalKinetic model fit to the SSM follow a nearly flat SMHM relation, which is clearly wrong. The two other cases, in which the observed GSMF was used as a constraint to the model, have SMHM relations whose shape resembles that in Moster et al. (2018) and Behroozi et al. (2019). Interestingly, the model that is constrained only by the GSMF produces an SMHM relation whose peak is ≈ 0.1 dex higher than in the two semi-empirical models. This discrepancy is gone in the model fit to both the GSMF and SSM.

Similar to Fig. 5.2, for all cases of observational constraints and for all emulated relations shown in Fig. 5.5, the actual simulations closely follow the predictions of the emulators, thereby validating our emulator-based approach.

Fig. 5.6 shows the posterior distributions of the subgrid parameters for the ThermalKinetic model after fitting the emulator to the observed GSMF (purple), to the observed SSM (brown), or to both (orange). The three contours of each colour indicate the 34, 68, and 95 per cent credibility regions of the posterior distributions. In the legend, we quote the values of the reduced χ^2 for each fit. Fitting to either the GSMF or SSM yields a reduced χ^2 of order unity, indicating that the model is a good fit to the data, as we have seen in Fig. 5.5. In contrast, fitting simultaneously to the GSMF and SSM yields a reduced χ^2 of 5.5, indicating the model lacks the necessary complexity to match both observed relations.

Examining the positions of the peak of the posterior distribution reveals that in each case the model fitting the GSMF and the model fitting the SSM belong to very different regions of the parameter space. In particular, the best-fitting model to the SSM (brown) prefers an SMBH seed mass of $M_{\rm BH,seed} \approx 10^{3.8}\,{\rm M}_\odot$, an energy per SN in units of 10^{51} erg of $f_{\rm E} \approx 0.8$, and no SN kinetic feedback ($f_{\rm kin} \approx 0$), whereas the best-fitting model to the GSMF (purple) has $M_{\rm BH,seed} \approx 10^{4.4}\,{\rm M}_\odot$, $f_{\rm E} \approx 1.4$, and $f_{\rm kin} \approx 0.72$. The best-fitting parameters of the model fit to both the GSMF and SSM are located roughly in between the best-fitting parameters of the models fit to only one observed relation.

5.5.2 ThermalKineticVariableΔT and Reference models

Having learned that neither the Basic model nor the ThermalKinetic model can simultaneously fit the observed GSMF and SSM, we turn our attention to the more complex models: ThermalKineticVariableΔT and Reference.

Galaxy stellar mass function and galaxy sizes

Fig. 5.7 shows the GSMF and SSM relation at $z = 0$ for the best-fitting ThermalKineticVariableΔT (navy-blue) and Reference (light-blue) models. The different symbols and line styles have the same meaning as in Fig. 5.2.

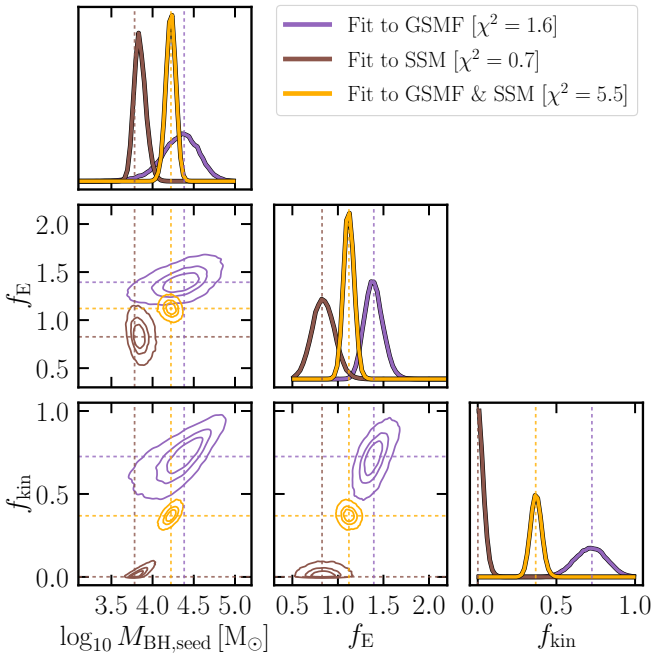

Figure 5.6: The posterior distributions of the parameters of the ThermalKinetic model resulting from fitting the emulator to the observed GSMF (purple), the observed SSM (brown), or both the GSMF and SSM (orange). The three contours of each colour indicate 34, 68, 95 credibility levels. The vertical and horizontal lines indicate the values of the best-fitting parameters for each case. The best-fitting parameter values of the model fit to the GSMF and the model fit to the SSM belong to very different regions of the parameter space. The model fit to both the GSMF and SSM is located in between the models fit to the GSMF and SSM separately.

It is evident that both the ThermalKineticVariableΔT and Reference models outperform the Basic and ThermalKinetic models, whose GSMF and SSM were presented in Fig. 5.2. The GSMFs for both ThermalKineticVariableΔT and Reference closely follow the observed relation from $M_* \approx 10^{8.5}$ to $10^{11.5}\,\rm M_\odot$, with the errors relative to the observed GSMF scattered within ≈ 0.1 dex. The SSM relation in the ThermalKineticVariableΔT model matches the observed relation only broadly: similar to the ThermalKinetic model shown in Fig. 5.2, the ThermalKineticVariableΔT model features a dip in the SSM at $M_* \approx 2 \times 10^{10}\,\rm M_\odot$ (albeit not as pronounced), which is not present in the observed relation. This dip is absent in the SSM of the Reference model, which closely follows the observed SSM relation for nearly all stellar masses, reproducing both the median galaxy half-mass size and its scatter at a fixed M_*. As expected, the model's SSM begins to diverge

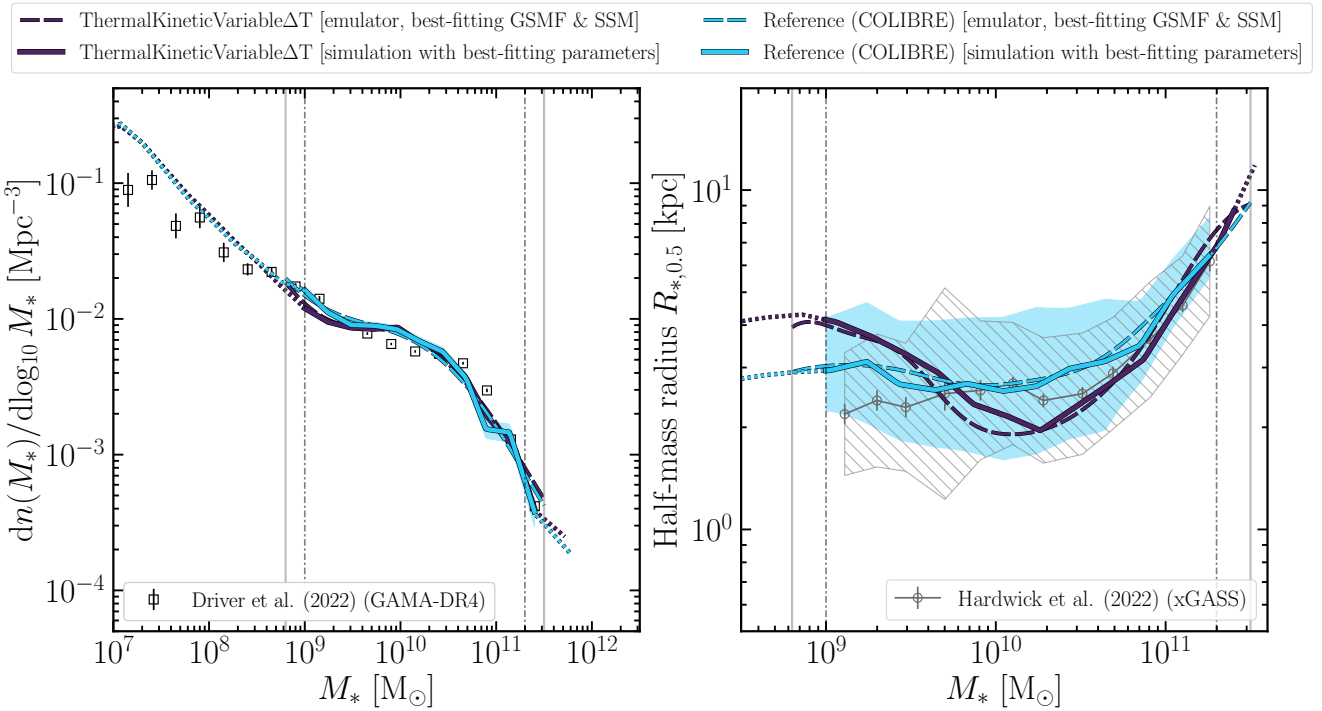

Figure 5.7: As Fig. 5.2 but showing the $z = 0$ galaxy stellar mass function (GSMF; *left*) and median size-stellar mass relation (SSM; *right*) for the ThermalKineticVariableΔT (navy-blue) and COLIBRE reference (light-blue) models. While the ThermalKineticVariableΔT model shows a good agreement with the observed GSMF but only broad agreement with the observed SSM, the COLIBRE reference model successfully reproduces both observational constraints.

from the observed relation below $M_* \sim 10^9\,\mathrm{M}_\odot$ where the finite resolution of the simulation starts to play a role (at our resolution, a galaxy with a stellar mass of $10^9\,\mathrm{M}_\odot$ is sampled with only ≈ 68 stellar particles).

To sum up, among the four considered models – Basic, ThermalKinetic, ThermalKineticVariableΔT, and Reference – the colibre reference model is the only model that is consistent with both the observed GSMF and SSM. Furthermore, as in previous figures that showed results from emulators and simulations for the same model parameters, Fig. 5.7 confirms that the emulator errors are negligibly small.

Posterior distributions of the model parameters

Fig. 5.8 displays the posterior distribution of the parameters in ThermalKineticVariableΔT (navy-blue) and Reference (light-blue) after each model was fit to both the observed GSMF and SSM. Because one of the four parameters from ThermalKineticVariableΔT does not exist in Reference, and vice versa, in the bottom row of the figure, we plot the unique parameters of the two models at the same time: f_E for ThermalKineticVariableΔT and $P_\mathrm{birth,0}$ (in log) for Reference. We show the same range of values for both parameters but attach to the panels' axes two different labels, which for clarity are shown in the colours of the corresponding models (light-blue and navy-blue).

First, both models prefer an SMBH seed mass close to $10^{4.1}\,\mathrm{M}_\odot$, with the Reference (ThermalKineticVariableΔT) model leaning towards slightly higher (lower) values. The fraction of SN energy injected in kinetic form is approximately 0.1 in both cases, which is lower than $f_\mathrm{kin} \approx 0.37$ in the ThermalKinetic model that was fit to the same observational data. This is likely because, unlike in ThermalKinetic, the heating temperature ΔT_SN in ThermalKineticVariableΔT and Reference can vary between $10^{6.5}$ and $10^{7.5}$ K. The SN thermal feedback with low ΔT_SN can reproduce some of the effects resulting from SN kinetic feedback with low Δv_kick, so the value of f_kin no longer needs to be as high as ≈ 0.37. However, the thermal feedback cannot replace kinetic feedback completely because for low ΔT_SN and/or in high-density gas, radiative energy losses will inevitably become high, rendering the thermal feedback inefficient.

Second, the best-fitting value of the pivot density in the SN thermal feedback with a variable heating temperature, $n_\mathrm{SN,0}$, is approximately 0.5 cm^{-3} in ThermalKineticVariableΔT and 0.6 cm^{-3} in the Reference model, which both correspond to the Dalla Vecchia & Schaye (2012) critical density for $f_t \approx 2$ (see Eq. 5.12). The best-fitting value of f_E in the ThermalKineticVariableΔT model is ≈ 1.4, which is slightly higher than in the ThermalKinetic model ($f_\mathrm{E} \approx 1.1$). The increase in f_E likely originates from the fact that the average ΔT_SN in ThermalKineticVariableΔT is lower than $10^{7.5}$ K, making the SN thermal feedback in this model weaker than in ThermalKinetic. To compensate for the weaker SN feedback, the

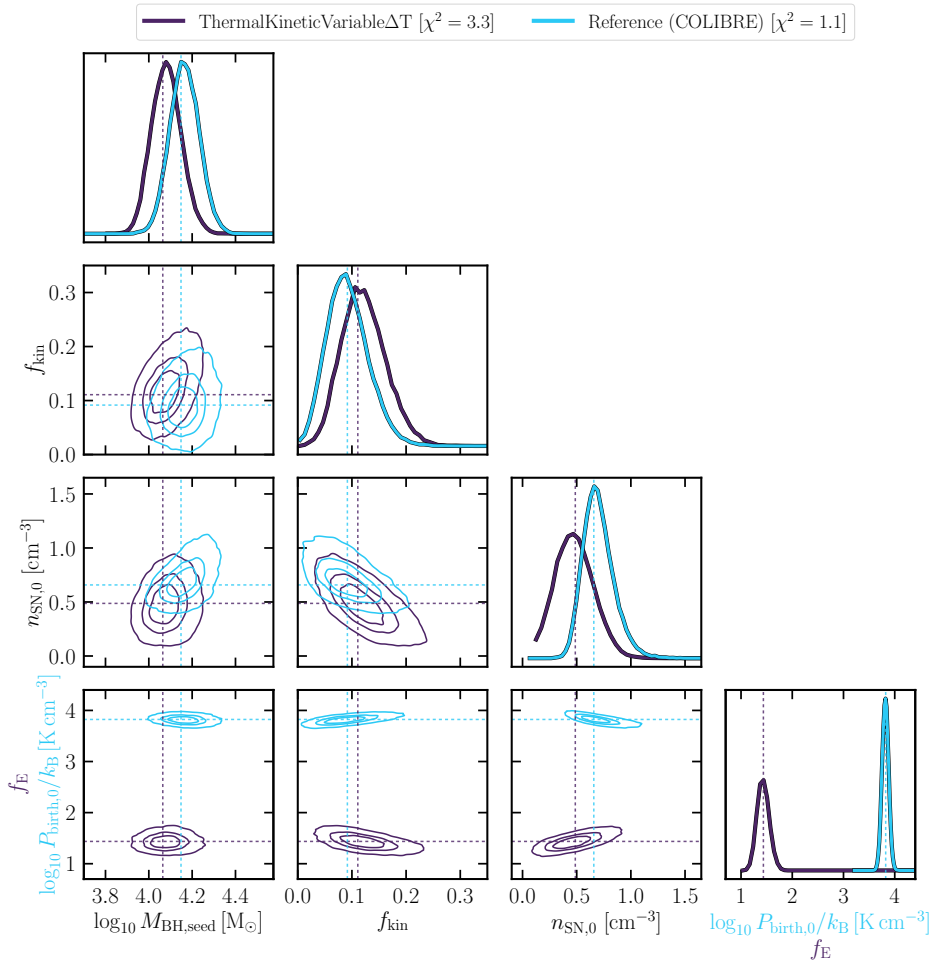

Figure 5.8: The posterior distributions of the parameters of the ThermalKineticVariableΔT model (navy-blue) and the COLIBRE reference model (light-blue) resulting from fitting the emulator to the $z = 0$ observed GSMF and observed SSM. The contours of the same colour indicate the 34, 68, and 95 per cent credibility regions of the posterior distributions. The vertical and horizontal dotted lines indicate the values of the best-fitting parameters for each model, corresponding to the maximum of the posterior distribution. Both models prefer an SMBH seed mass around $10^{4.1}\,\mathrm{M}_\odot$, a fraction of SN energy injected in kinetic form $f_{\rm kin} \approx 0.1$, and a pivot density in SN thermal feedback, $n_{\rm SN,0}$, around $0.55\,\mathrm{cm}^{-3}$. In the bottom row, we plot the parameters that are unique to each model: $f_{\rm E}$ for ThermalKineticVariableΔT, and $P_{\rm birth,0}/k_{\rm B}$ (in log) for the Reference model; their best-fitting values are approximately 1.4 and $10^{3.8}\,\mathrm{K\,cm}^{-3}$, respectively.

energy per SN in units of 10^{51} erg is increased from ≈ 1.1 to 1.4.

Third, the best-fitting value of the pivot birth pressure in the Reference model, $P_{\rm birth,0}/k_{\rm B}$, is $\approx 10^{3.8}$ K cm^{-3}. Coincidentally, the median stellar birth pressure in the Reference model with the best-fitting parameters is also $\approx 10^{3.8}/k_{\rm B}$ K cm^{-3}. Plugging this median value into equation (5.8) along with the other model parameters that we did not consider in the calibration ($f_{\rm E,min} = 0.1$, $f_{\rm E,max} = 3.5$ and $\sigma_{\rm P} = 0.3$), we obtain for the SN energy at the median stellar birth pressure $f_{\rm E}(P_{\rm birth}/k_{\rm B} = 10^{3.8}\,{\rm K\,cm^{-3}}) = 1.8$ The value of $f_{\rm E}$ averaged over all stellar particles formed in the simulation is also ≈ 1.8, which is comparable to the best-fitting value of (constant) $f_{\rm E}$ in ThermalKineticVariableΔT, $f_{\rm E} = 1.4$, indicating that both models prefer the (average) energy in SN feedback to be slightly higher than the theoretical expectation ($\sim 10^{51}$ erg).

Fourth, the legend of Fig. 5.8 gives the reduced χ^2 of the fits to the observational data, which is 3.3 for the ThermalKineticVariableΔT model and 1.1 for Reference. This confirms that both models perform better than the Basic and ThermalKinetic, and that the Reference model provides the best match to the GSMF and SSM.

Lastly, we note that based on isolated galaxy simulations at much higher resolution ($m_{\rm gas} = 10^5\,{\rm M_\odot}$), Chaikin et al. (2023) found that $f_{\rm kin} \approx 0.1$ (together with the kick velocity of $\Delta v_{\rm kick} = 50$ km s^{-1}) allows reproducing the relation between spatially resolved HI velocity dispersion and the galaxy SFR surface density, as well as the observed KS star-formation law. The latter was confirmed in Nobels et al. (2023), who showed that the observed KS relation is reproduced for the range of mass resolutions from $m_{\rm gas} = 1.25 \times 10^4$ to 5.12×10^7 M$_\odot$. These findings are reassuring given that both the ThermalKineticVariableΔT and Reference best-fitting models prefer[16] $f_{\rm kin} \approx 0.1$.

5.5.3 Galaxy properties to which the models were not explicitly calibrated

In this section, we explore the simulation predictions for galaxy properties that have not been discussed before, and, consequently, have not been considered during the search for the best-fitting parameter values. In the following figures, we will present the results for the four best-fitting models, Basic, ThermalKinetic, ThermalKineticVariableΔT, and Reference, which have all been fit to the observed GSMF and SSM. We will show only the results from the simulations, as our emulators exist only for the $z = 0$ GSMF, SSM, and SMHM.

[16] For comparison, fitting the ThermalKineticVariableΔT and Reference models to the $z = 0$ observed GSMF and SSM with a constraint $f_{\rm kin} = 0$ yields a reduced χ^2 for both models that is a factor of ≈ 1.4 higher compared to the case with $f_{\rm kin} \approx 0.1$.

Star formation rates and quenched fractions

Fig. 5.9 displays the specific star formation rates (sSFR) of active galaxies, the fraction of passive galaxies, and the masses of SMBHs. All relations are shown at $z = 0$ and plotted versus galaxy stellar mass. All quantities are computed in 3D apertures of 50 kpc. If a halo contains multiple BH particles, then the mass of the most massive one is plotted. In a given stellar mass bin, we show the median sSFR and the median mass of the SMBHs. We define a galaxy as 'active' if its sSFR is greater than 10^{-2} Gyr^{-1}. The results from the simulations with the four best-fitting models are given by the solid curves. Namely, we display the Basic model (green), ThermalKinetic (orange), ThermalKineticVariableΔT (navy-blue), and Reference (light-blue). The shaded light-blue region indicates the 16^{th} to 84^{th} percentiles scatter around the median values in the Reference model.

As comparison data, we take the sSFR-stellar mass relation for star-forming galaxies from Bauer et al. (2013) and Chang et al. (2015). The former includes $\sim 10^5$ galaxies from the GAMA survey (Driver et al., 2011) with redshifts between $0.05 < z < 0.32$, while the latter is based on $\sim 10^6$ SDSS galaxies with redshifts $z < 0.2$ including four-band photometry from WISE (Wright et al., 2010). For the quenched fractions, we take the $z \sim 0.1$ data from Moustakas et al. (2013), which is based on a sample of 1.7×10^5 SDSS galaxies; and the $z = 0$ quenched fractions from the semi-empirical model UNIVERSEMACHINE (Behroozi et al., 2019). The observations for the SMBH mass-stellar mass relation are taken from Graham & Sahu (2023) whose sample is subdivided into galaxies with different morphological types: E, ES/S0, and S. Where needed, we correct for the differences in the assumed stellar IMF by converting the data into the Chabrier (2003) IMF, and account for the differences in the assumed cosmology.

Overall, the agreement between the simulations and comparison data improves with the complexity of the galaxy formation model. First, the Basic model predicts a $z = 0$ sSFR-stellar mass relation that has an unrealistic, flat shape and is offset by more than 0.5 dex towards lower values at low and intermediate stellar masses ($\lesssim 10^{10}\,M_\odot$), compared to the observed trends. In consequence, the fraction of passive galaxies with similar stellar masses is much too high relative to the data. At the same time, the passive fraction of the most objects with $M_* > 10^{11}\,M_\odot$ is too low in the Basic model. The agreement with the data is slightly better for the ThermalKinetic model. At the high-mass end, the passive fractions are no longer suppressed compared to the data, which is a result of increasing the SMBH seed mass by more than a factor of 6 relative to that in the Basic model. However, the sSFRs (passive fractions) of low- and intermediate-mass galaxies still remain too low (high). This is a consequence of using a constant heating temperature in the SN thermal feedback, $\Delta T_{\text{SN}} = 10^{7.5}$ K, leading to an energy output by multiple SNe that is too powerful. By $z = 0$, this powerful SN feedback has likely disrupted, heated, and/or ejected most of the cold gas in

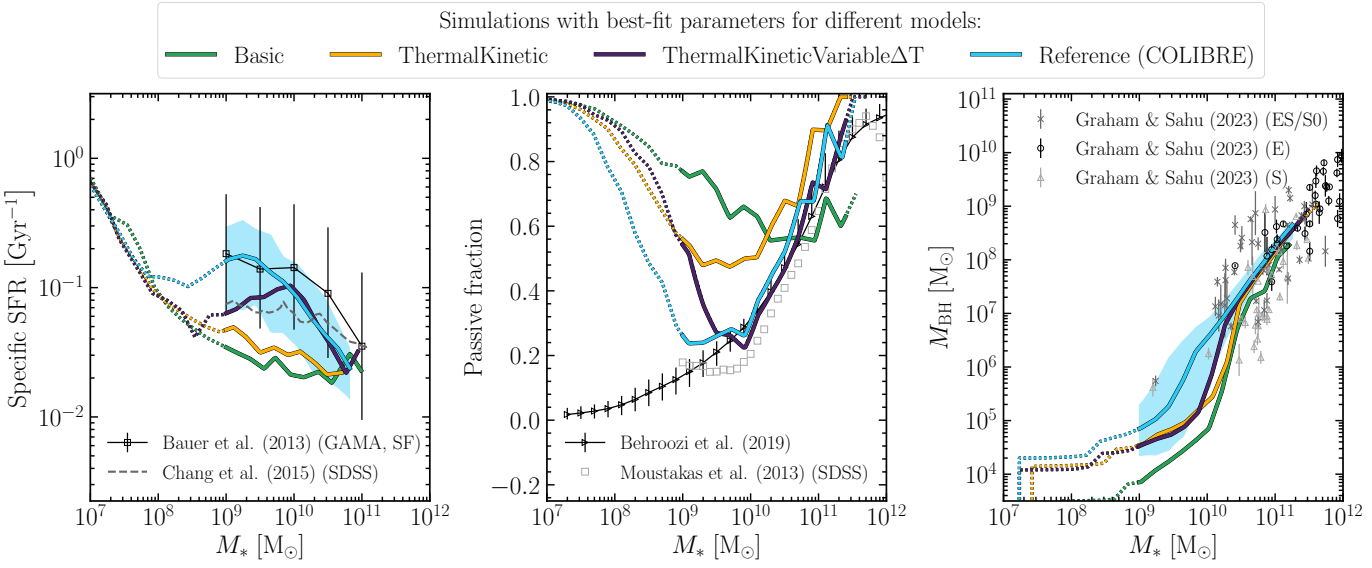

Figure 5.9: The median specific star formation rate (sSFR) of active galaxies (sSFR $> 10^{-2}$ Gyr^{-1}) versus stellar mass (*left*), the fraction of passive galaxies versus stellar mass (*middle*), and the median mass of supermassive black holes versus stellar mass (*right*), all shown at $z = 0$. Differently coloured solid curves show the results from the simulations with the best-fitting parameters for the Basic (green), ThermalKinetic (orange), ThermalKineticVariableΔT (navy-blue), and Reference models (light-blue). All models were fit to the $z = 0$ observed GSMF and SSM. The solid curves turn into dotted curves at stellar mass below $10^9\,\mathrm{M}_\odot$ indicating that those galaxies are poorly resolved, and above $10^{11.4}\,\mathrm{M}_\odot$ indicating the limit due to the simulated volume. The shaded blue region shows the 1σ scatter in the Reference model. A compilation of data from observations and semi-empirical models is shown in grey-scale colours. Both the ThermalKineticVariableΔT and Reference models show a good agreement with the comparison data for all three relations, with the Reference model exhibiting marginally better sSFR and passive fractions at low stellar mass.

low- and intermediate-mass galaxies, which would otherwise be eligible to form stars, resulting in higher sSFRs.

Indeed, switching to the density-dependent heating temperature in the ThermalKineticVariableΔT and Reference models strikingly improves the agreement with the data. ThermalKineticVariableΔT successfully reproduces both the passive fractions and sSFR for galaxies with stellar masses $M_* \gtrsim 10^{10}\,\mathrm{M}_\odot$, while in Reference the agreement with the data is extended down to $\sim 10^9\,\mathrm{M}_\odot$. At lower stellar masses, $M_* \lesssim 10^9\,\mathrm{M}_\odot$, the fractions of passive galaxies in the four models all begin to increase with decreasing stellar mass, which is a resolution effect: galaxies sampled with too few star-forming gas particles will on average appear more passive, leading to a lower median sSFR and higher passive fractions (e.g. Schaye et al., 2015).

Focussing on the right panel, we discover that in all four models, galaxies with stellar masses $M_* \gtrsim 10^{10.5}\,\mathrm{M}_\odot$ harbour SMBHs that grew to masses far exceeding $10^7\,\mathrm{M}_\odot$. The masses of these SMBMs follow a tight relation with the host galaxy's stellar mass, which has roughly the same slope and normalization as the observed scaling from Graham & Sahu (2023). We do not find any large differences between the models, which is expected since all models employ the same numerical prescription for SMBH growth and AGN feedback. The only AGN-related parameter that is different between the models is their best-fitting value of the seed mass, $M_{\mathrm{BH,seed}}$. Among the four models, the Basic model adopts the lowest seed mass, and the Reference model the highest. The higher (lower) $M_{\mathrm{BH,seed}}$ leads to a somewhat faster (slower) SMBH growth in the latter (former) model, which lasts until the SMBH has reached the self-regulating regime where AGN feedback dominates over stellar feedback. As a result, at $M_* \lesssim 10^{10.5}\,\mathrm{M}_\odot$, the median SMBH mass in the Reference model is higher than in the Basic model. The SMBH masses in the other two models, ThermalKinetic and ThermalKineticVariableΔT, have values that are intermediate between those in Basic and Reference.

Cosmic star formation history

Fig. 5.10 displays the redshift evolution of the cosmic star formation rate density (SFRD). We show the same four models as in Fig. 5.9. For comparison, we consider the observational estimates of the SFRD derived from the LOFAR Deep Fields' radio data at $0 < z < 4$ (Cochrane et al., 2023), the rest-frame far-ultraviolet data collected with ALMA at $z \approx 4.5$ to 5.5 (Khusanova et al., 2021), the sources detected in ALMA and combined with their multi-wavelength counterparts in the COSMOS and ECDFS surveys at $0.5 < z < 6$ (Gruppioni et al., 2020), the deep VLA COSMOS radio observations over $0.3 < z < 5$ (Novak et al., 2017), and the VLA radio observations from a subsample of the GOODS-N survey at $0.1 < z < 3$ (Enia et al., 2022). Additionally, we include the prediction for the

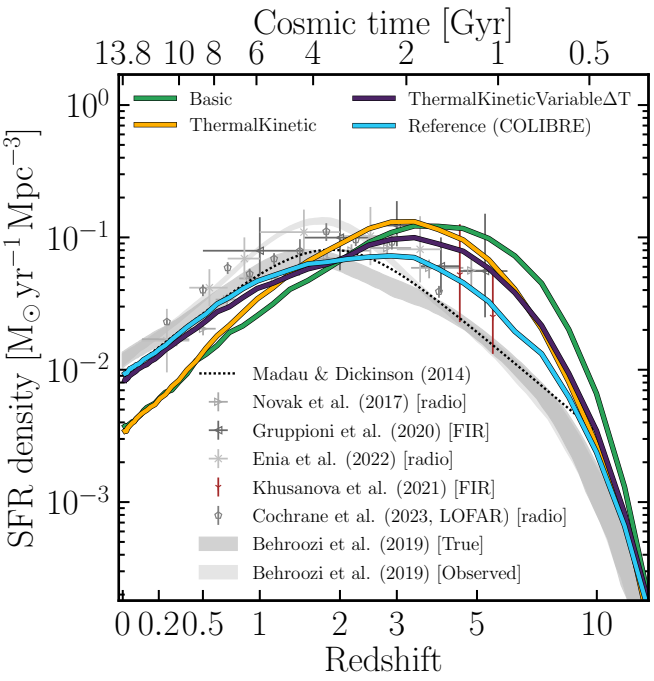

Figure 5.10: Cosmic average star formation rate density (SFRD) versus redshift from the simulations with the best-fitting models to the $z = 0$ GSMF and SSM: Basic, ThermalKinetic, ThermalKineticVariableΔT, and Reference (differently coloured solid curves). A compilation of comparison data is shown in grey-scale colours. The use of a variable heating temperature in the SN feedback of the ThermalKineticVariableΔT and Reference models greatly improves the agreement with the comparison data at low z. The inclusion of a stellar birth pressure-dependent SN energy in the Reference model results in a lower SFRD at $z > 2$, thereby further improving the agreement with the comparison data.

evolution of the SFRD using the best-fitting function from Madau & Dickinson (2014) (black dashed curve), which was derived based on a collection of IR and UV data at $0 < z < 8$. Finally, we display the SFRD from the semi-empirical model UNIVERSEMACHINE (Behroozi et al., 2019), for which we show the 'true' and 'observed' values (grey shaded regions), the latter of which accounts for systematic effects while the former reports the intrinsic values predicted by the model.

The Basic and ThermalKinetic models predict significantly steeper slopes than observed for $z < 2$ and underpredict the observed SFRD at low redshift by ≈ 0.5 dex, which is consistent with both models having too low sSFR at $z = 0$ in Fig. 5.9. As is the case with the sSFR, the suppression in the SFRD at low redshifts is related to the high (constant) heating temperature used

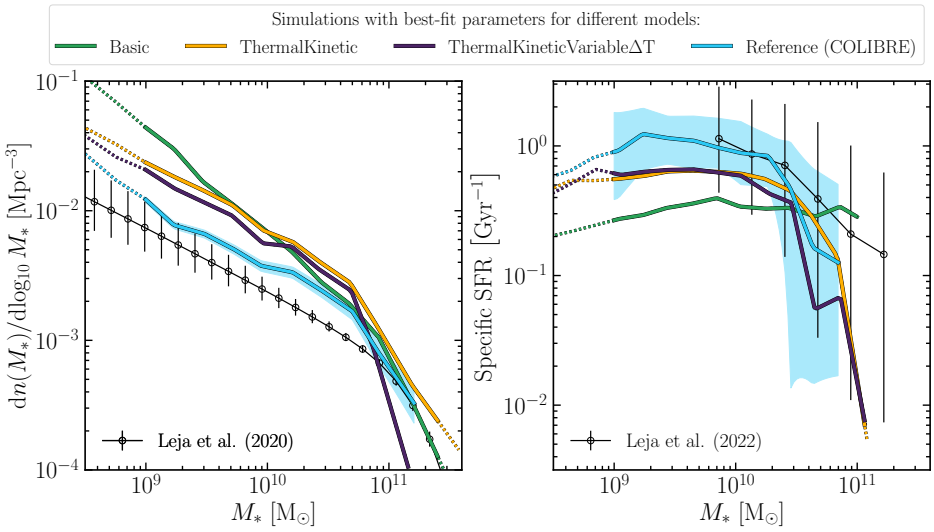

Figure 5.11: The $z = 2$ galaxy stellar mass function (GSMF; *left*) and the $z = 2$ median specific star formation rate (sSFR, *right*) versus stellar mass. The solid curves show the results from the simulations with the best-fitting parameters for the Basic, ThermalKinetic, ThermalKineticVariableΔT, and Reference models. For comparison, we show the $z = 2$ GSMF from Leja et al. (2020) and the $z = 2$ sSFR from Leja et al. (2022), which were derived from the COSMOS2015 and 3D-HST galaxy catalogues, using non-parametric methods. The shaded blue region shows the 1σ scatter in the Reference model. Among the four models, the Reference model (in light-blue) produces the best match to the $z = 2$ observed GSMF and sSFR though the simulated galaxies appear slightly overmassive, by about 0.2 dex compared to Leja et al. (2020).

in SN thermal feedback: $\Delta T_{\rm SN} = 10^{7.5}$ K. Conversely, the SFRD in the other two models, ThermalKineticVariableΔT and Reference, which incorporate the variable heating temperature, matches the $z \lesssim 1$ observed SFRD much better, following the fit from Madau & Dickinson (2014) nearly perfectly.

Moving to high redshifts ($z \gtrsim 2$), we observe that the SFRD in the Basic, ThermalKinetic, and ThermalKineticVariableΔT models rises very steeply with cosmic time and reaches its maximum before or at $z \approx 3$. This is an indication that the star formation at high redshifts might be overly efficient. Note that because all four models were calibrated to reproduce the observed GSMF (and SSM) at $z = 0$, a too-high SFRD at early cosmic times necessitates a too-low SFRD at late times (and vice versa), such that the right amount of stellar mass is formed by $z = 0$.

The situation is noticeably improved in the Reference model, in which the amount of stellar mass formed before $z = 3$ is significantly reduced compared

to the other models. The SFRD in the Reference model exhibits a broad peak between $z = 4$ and 1 and starts declining rapidly with cosmic time only thereafter. This improvement was made possible by the dependence of the SN energy on the stellar birth pressure (see Eq. 5.8), which is incorporated only into the Reference model. Specifically, the SN feedback in the Reference is more energetic at higher redshifts, as the star formation at high z proceeds on average in higher gas-pressure environments. Releasing more SN energy at high z not only reduces the cosmic SFRD but also helps avoid runaway star formation in the centres of massive galaxies, which may lead to the formation of a pronounced stellar bulge component. The presence of a dominant bulge can be traced by the dip in the $z = 0$ SSM relation at $M_* \sim 10^{10.5}\,\mathrm{M}_\odot$, which is absent only in the Reference model (see Figs. 5.2 and 5.7).

Galaxy stellar mass function and star formation rate at high redshift

Fig. 5.11 shows the GSMF and sSFR versus stellar mass at $z = 2$ for the same four simulations as were shown in Figs. 5.9 and 5.10. At a given stellar mass, we show the median sSFR considering both passive and active galaxies. For comparison, we display the $z = 2$ GSMF and median sSFR derived by Leja et al. (2020, 2022) who applied the SED-fitting code Prospector to the data from COSMOS2015 and 3D-HST galaxy catalogues and used non-parametric methods to model the evolution of stellar populations. The error bars in Leja et al. (2020, 2022) correspond to the 16^{th} and 84^{th} percentiles estimated by their model.

Clearly, at $z = 2$, the agreement with the data is much better for the Reference model than for its three simplified counterparts. Although the Reference model was only calibrated to the $z = 0$ data, the model broadly reproduces the observed median sSFR, with the largest discrepancies with the data being smaller than ≈ 0.1 dex, and is systematically offset by only ≈ 0.2 dex from the observed GSMF. The other three models appear less satisfactory: they all systematically underpredict the observed sSFR and their GSMF is on average higher by more than 0.3 dex relative to the observed data.

Cold gas properties

We next investigate the $z = 0$ properties of cold gas predicted by the simulations. The left and right panels of Fig. 5.12 show, respectively, the $z = 0$ ratios of galaxy HI mass to stellar mass and H_2 mass to stellar mass as functions of galaxy stellar mass. The solid curves correspond to the median mass fractions in the simulations. As is conventional for observational studies, all H_2 mass fractions include helium, which makes the H_2 mass larger by a factor of ≈ 1.3. For reference, we show the $z \approx 0$ observational data of HI to stellar mass fractions from the xGASS survey (Catinella et al., 2018) and H_2 to stellar mass fractions from the xCOLD GASS survey (Saintonge et al., 2017). Furthermore, we include the data from

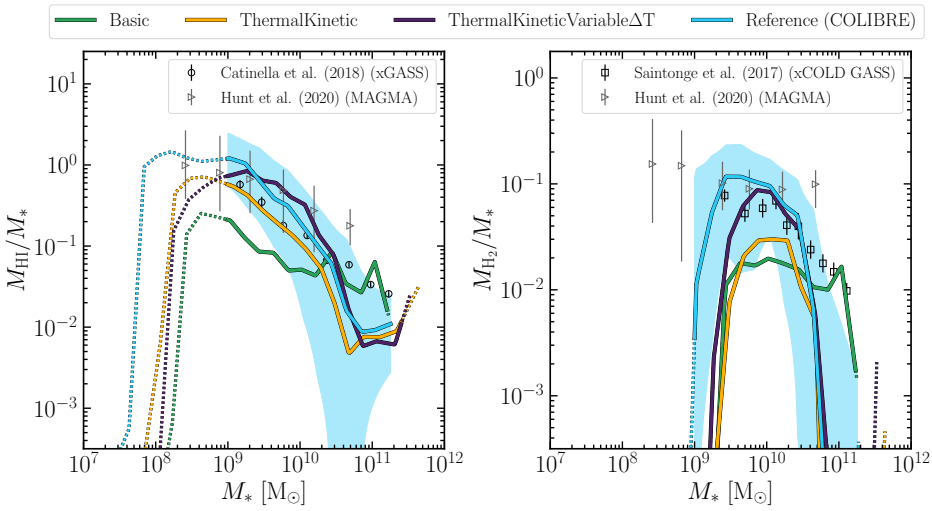

Figure 5.12: As Fig. 5.11, but showing the $z = 0$ HI mass to stellar mass ratio (*left*) and the $z = 0$ H_2 mass to stellar mass ratio (*right*) versus galaxy stellar mass. For comparison, we show the observed H_2 mass fractions from Saintonge et al. (2017), HI mass fractions from Catinella et al. (2018), and the measurements of both HI and H_2 content from Hunt et al. (2020). Only the ThermalKineticVariableΔT and Reference models reproduce both the HI and H_2 observed mass fractions for $M_* < 10^{10.5}\,M_\odot$, while at $M_* > 10^{10.5}\,M_\odot$ only the Basic model matches the observations.

the MAGMA sample of 392 galaxies at $z \approx 0$ with measurements of both HI and H_2 masses (Hunt et al., 2020). While the xGASS and xCOLD GASS samples are stellar mass selected, the selection of MAGMA galaxies requires both HI and CO detections, implying that MAGMA may be biased towards higher cold gas fractions.

We stress that the COLIBRE model does not impose an effective pressure and/or temperature floor and uses the non-equilibrium thermochemistry solver CHIMES to predict the abundances of primordial species. This allows us to take the (non-equilibrium) H_2 and HI abundances directly from the simulations, which is in contrast to many previous works, including EAGLE, IllustrisTNG, and Simba, where the atomic and molecular gas fractions need to be estimated in post-processing, using (semi-)analytic models or ionization fitting formulas calibrated by other simulations (e.g. Gnedin & Kravtsov, 2011; Rahmati et al., 2013).

We find that ThermalKineticVariableΔT and Reference are remarkably consistent with the observational data at stellar masses $10^9 < M_* < 10^{10.5}\,M_\odot$, for both the H_2 and HI mass fractions. The sharp downturn of the simulated fractions of

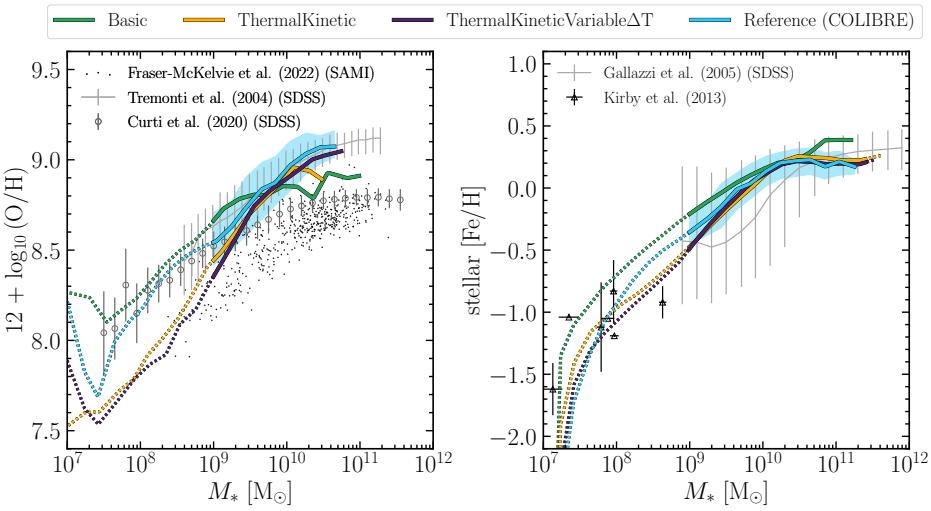

Figure 5.13: As Fig. 5.11, but showing the median gas-phase metallicity (*left*) and stellar metallicity (*right*) versus galaxy stellar mass at $z = 0$. The $z \approx 0$ comparison data (grey-scale colours) are taken from Fraser-McKelvie et al. (2022); Tremonti et al. (2004); Curti et al. (2020) for the gas-phase metallicity and from Gallazzi et al. (2005); Kirby et al. (2013) for the stellar metallicity. The gas-phase metallicity is computed in the gas that is sufficiently dense ($n_{\rm H} > 0.1$ cm^{-3}) and cool ($T < 10^{4.5}$ K) and only for star-forming galaxies, excluding metals that are present in dust. All four models are consistent with the observations for both the gas-phase and stellar metallicities.

molecular (atomic) gas below the stellar mass of $\sim 10^9\,{\rm M}_\odot$ ($\sim 10^8\,{\rm M}_\odot$) is driven by the limited numerical resolution[17]. The molecular gas fractions in the Basic and ThermalKinetic models are on average too low compared to the data, which is in line with the models' too low sSFR at $z = 0$ (Fig. 5.9).

In contrast, at the high-mass end ($M_* > 10^{10.5}\,{\rm M}_\odot$), only the Basic model matches the observed gas fractions of H_2 and HI, whereas the three more complex models underpredict the observed relations. The gas fractions in the Basic model are the largest among the four models because Basic uses a rather low SMBH seed mass, which results in weaker AGN feedback compared to the other models. In fact, there appears to be tension between the cold gas fractions and the passive fractions at $M_* > 10^{10.5}\,{\rm M}_\odot$: the Basic model reproduces the observed gas fractions but undershoots the observed galaxy passive fractions, while the opposite is true for the other three models. This discrepancy might indicate a limitation of our fairly simple treatments of gas accretion onto SMBHs and/or AGN feedback.

[17] Convergence tests show that higher resolution is required to robustly predict H_2 than to predict HI.

Alternatively, it might suggest that extra physics is needed such as cosmic ray transport or magnetic fields, which are not included in the COLIBRE model, to potentially allow massive galaxies to contain more gas that is not star-forming (see, e.g., Krumholz & Federrath, 2019; Hopkins et al., 2020, and references therein).

Metal content

The left panel of Fig. 5.13 shows the relationship between the metallicity of the gas phase, in units of $12 + \log_{10}(\mathrm{O/H})$, and galaxy stellar mass at $z = 0$. The gas metallicity in the simulations is derived directly from the oxygen abundance, as predicted by the chemistry model of COLIBRE. Each galaxy's ratio between the number of oxygen and hydrogen nuclei, O/H, is calculated as

$$\mathrm{O/H} = \frac{m_\mathrm{H}}{m_\mathrm{O}} \frac{\sum_i (X_\mathrm{O}/X_\mathrm{H})_i \, m_{\mathrm{gas},i}}{\sum_i m_{\mathrm{gas},i}}, \tag{5.27}$$

where $(X_\mathrm{O}/X_\mathrm{H})_i$ is the ratio of the oxygen and hydrogen mass fractions carried by gas particle i, $m_{\mathrm{gas},i}$ is the mass of gas particle i, and $m_\mathrm{H}/m_\mathrm{O}$ is the ratio of the masses of hydrogen and oxygen nuclei. To aid the comparison with observational data, in equation (5.27) we consider only those gas particles that are dense ($n_\mathrm{H} > 0.1$ cm^{-3}) and cool ($T < 10^{4.5}$ K). Furthermore, we apply a spatial mask by requiring the selected particles to be within 50-kpc apertures centred on the galaxies. We do not include those metals that are present in dust. In a given stellar mass bin, we show the median value of $12 + \log_{10}(\mathrm{O/H})$ considering only star-forming galaxies (sSFR $> 10^{-2}$ Gyr^{-1}).

For comparison, we display the gas-phase metallicities of 472 star-forming galaxies at $0.04 < z < 0.128$ extracted from the SAMI Galaxy Survey (Fraser-McKelvie et al., 2022), and the metallicities of $\sim 10^5$ local star-forming SDSS galaxies from Tremonti et al. (2004) and Curti et al. (2020). The observational data follow two metallicity tracks that are separated from each other by ≈ 0.3 dex. The systematic discrepancy originates from the differences in the methods used to infer gas metallicity values from the observed galaxy spectra. The lower track corresponds to the metallicity determined based on the so-called T_e-method, whereas the values of gas metallicity in the upper track are calibrated based on photoionization models (see e.g. López-Sánchez et al., 2012; Curti et al., 2020, for further details).

We find that all four models are consistent with the observational data, including the normalization of the mass-metallicity relation, its slope, and scatter. At stellar masses $M_* > 10^9\,\mathrm{M}_\odot$, ThermalKinetic and Reference closely follow the upper track of the observations, while the metallicities in Basic and ThermalKinetic saturate at slightly lower values.

The right panel of Fig. 5.13 shows the relationship between the galaxy stellar mass and the stellar metallicity, [Fe/H], at $z = 0$. The stellar metallicity in

the simulations is derived directly from the galaxies' iron abundance. For each galaxy, we first calculate the ratio of the number of iron and hydrogen nuclei, Fe/H, where we employ the same expression as for the gas-phase O/H (Eq. 5.27) but apply it to stellar particle-carried fields and replace oxygen with iron. Stellar particles that contribute to Fe/H are selected within 50-kpc apertures. The resulting ratio is subsequently normalized by the solar value of Fe/H, assuming a solar iron abundance of $12 + \log_{10}(\mathrm{Fe/H}) = 7.5$ (Asplund et al., 2009). In a given stellar mass bin, we show the median value of [Fe/H].

For reference, we display the observed stellar metallicity-mass relation for a large ($\sim 10^5$) sample of $z \approx 0.1$ SDSS galaxies from Gallazzi et al. (2005) and dwarf irregular and spheroidal satellite galaxies of the Milky Way and M31 from Kirby et al. (2013). Where needed, the solar abundances used in the observations have been converted to the solar values reported by Asplund et al. (2009).

Overall, all four models appear consistent with the observations within the stellar mass range from $\sim 10^7$ to $10^{11}\,\mathrm{M}_\odot$. The stellar mass-metallicity relation in the Basic model saturates at somewhat higher metallicity compared to the other three models. This is a consequence of the low seed mass adopted in the Basic model, which leads to weaker AGN feedback, and, hence, more late-time star formation in the massive objects, thereby elevating their present-day stellar metallicities.

5

5.5.4 Other feedback parameter variations

In the previous section, we have shown that the COLIBRE reference model with the best-fitting parameter values reproduces not only the observational data it was calibrated to (the $z = 0$ GSMF and SSM), but is also consistent with many other observed relations, both at low and high z. Because the model was calibrated by finding the best-fitting values for the set of four subgrid parameters, $\boldsymbol{\theta} = (f_{\mathrm{kin}}, n_{\mathrm{SN,0}}, P_{\mathrm{birth,0}}, M_{\mathrm{BH,seed}})$, an important question is: why were these four parameters optimized, while other subgrid parameters describing stellar and AGN feedback were not? To answer this question, we demonstrate in this section that the feedback-related subgrid parameters that were held fixed during emulation either have little effect on the calibrated galaxy properties or are degenerate with the parameters already contained in $\boldsymbol{\theta}$.

Figs. 5.14 and 5.15 show how the $z = 0$ GSMF and SSM in the Reference model respond to variations in $\Delta T_{\mathrm{SN,max}}$, $\Delta T_{\mathrm{SN,min}}$, $n_{\mathrm{SN,0}}$, $f_{\mathrm{E,min}}$, $f_{\mathrm{E,max}}$, $P_{\mathrm{birth,0}}$, σ_P, $M_{\mathrm{BH,seed}}$, $M_{\mathrm{FOF,seed}}$, and ΔT_{AGN}. Each figure contains 10 panels, where in each panel, we vary one of the 10 subgrid parameters, while the other parameters are fixed at their best-fitting values. For each variation, we run a separate simulation (that is, what is shown in the figures are the results from actual simulations, not from the emulators). We show the results for three different values of each parameter: the value from the Reference model (light-blue), a lower value (dark-

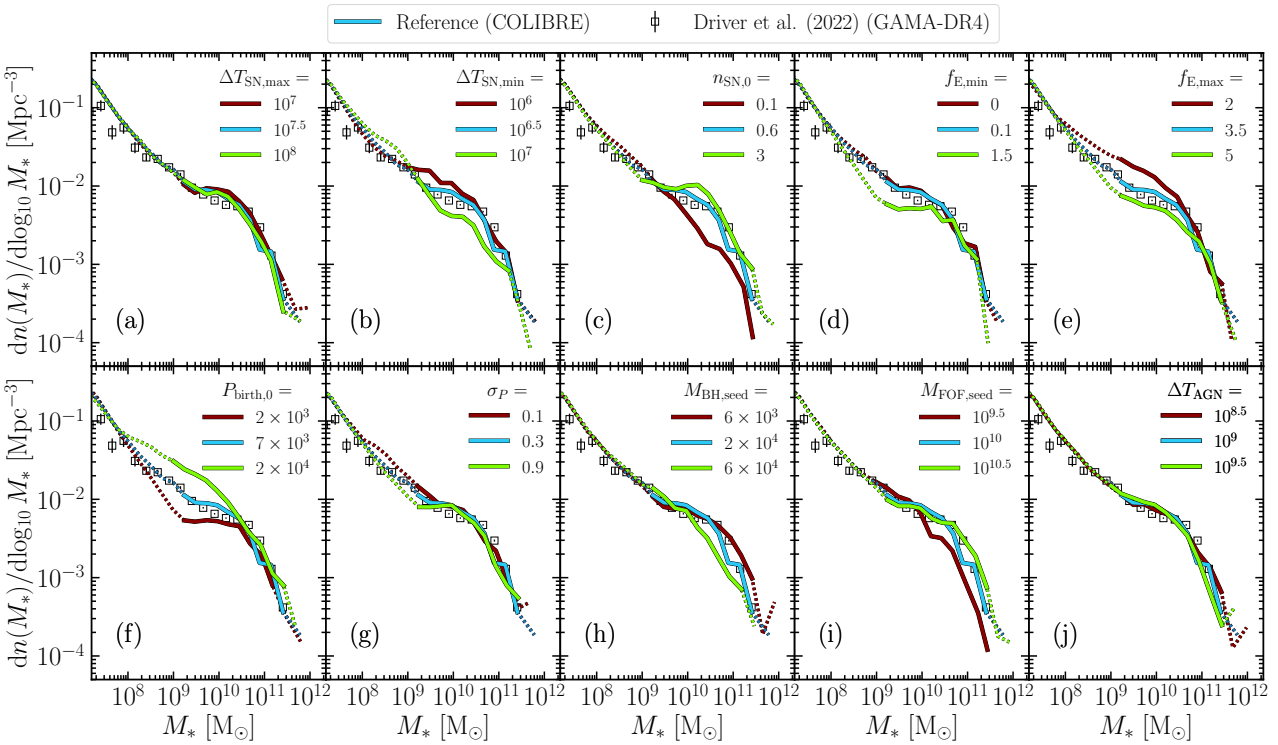

Figure 5.14: The galaxy stellar mass function (GSMF) at redshift $z = 0$. Different panels show the effect of varying different feedback-related subgrid parameters of the COLIBRE reference model. Only one parameter is varied per panel, while the others are held fixed to their best-fitting values. Starting with the top left panel and going clock-wise, the varied parameters are $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,min}$, $n_{\rm SN,0}$, $f_{\rm E,min}$, $f_{\rm E,max}$, $P_{\rm birth,0}$, σ_P, $M_{\rm BH,seed}$, $M_{\rm FOF,seed}$, and $\Delta T_{\rm AGN}$. The light-green (dark-red) curve corresponds to the higher (lower) value of the parameter, relative to that in the COLIBRE reference model, which is shown in light-blue. The comparison data from Driver et al. (2022) are displayed as black squares. Many parameters are degenerate with one another while others have little effect on the GSMF.

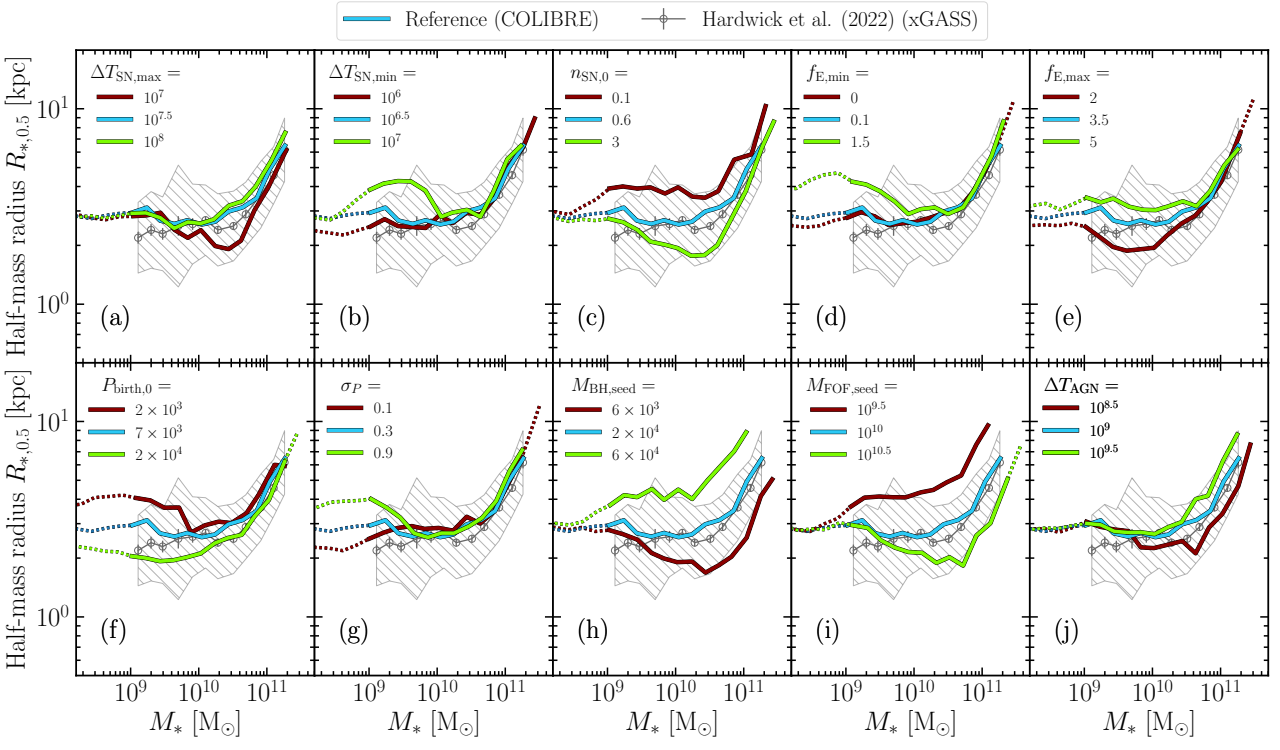

Figure 5.15: As Fig. 5.14, but showing the $z = 0$ median galaxy size-stellar mass relation. The median observed relation from Hardwick et al. (2022) is displayed as dark-grey circles. The error bars show the 1σ error on the median observed sizes, while the grey hatched region indicates the galaxy population-wide scatter.

red), and a higher value (light-green). The values of the varied parameters are shown in each panel's legend.

First, we recall that the density-dependent heating temperature in the SN thermal feedback of the Reference model depends on the model parameters $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,min}$, and $n_{\rm SN,0}$ (see Eq. 5.14). The effects of varying these parameters are shown in panels (a), (b), and (c). We observe that changing $\Delta T_{\rm SN,max}$ by ± 0.5 dex has little influence on the GSMF. The galaxy sizes do not change if we increase $\Delta T_{\rm SN,max}$ from $10^{7.5}$ to 10^{8} K, but begin to dip around $M_* = 10^{10.5}\,{\rm M}_\odot$ if $\Delta T_{\rm SN,max}$ is lowered to 10^{7} K, which is an indication of inefficient SN feedback. Changing $\Delta T_{\rm SN,min}$ by ± 0.5 dex has a pronounced effect on both the GSMF and SSM. As expected, higher (lower) $\Delta T_{\rm SN,min}$ makes SN thermal feedback more (less) efficient, resulting in less (more) stellar mass formed and less (more) centrally concentrated galaxies, moving the low mass-end of the GSMF down (up) and moving the low mass-end of the SSM up (down). By comparing panels (a) and (b) with panel (c), in which $n_{\rm SN,0}$ is varied, we observed that most of the effects on the GSMF and SSM produced by variations in $\Delta T_{\rm SN,min}$ and/or $\Delta T_{\rm SN,max}$ can be captured by solely changing $n_{\rm SN,0}$. Furthermore, we recall that $\Delta T_{\rm SN,min}$ cannot be much lower than $10^{6.5}$ K (otherwise the SN thermal feedback will suffer from catastrophic overcooling) and $\Delta T_{\rm SN,max}$ cannot be much greater than $10^{7.5}$ K (otherwise the sampling of SN thermal injection events will become too poor, see §5.2.8). Based on these arguments, we decided to refrain from optimizing $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$ and instead fix these parameters to $10^{6.5}$ and $10^{7.5}$ K, respectively.

We next move to panels (d), (e), (f), and (g), which vary the parameters of the relation between the SN energy $f_{\rm E}$ and stellar birth pressure $P_{\rm birth}$ (Eq. 5.8). The parameters are: $f_{\rm E,min}$, $f_{\rm E,max}$, $P_{\rm birth,0}$, and σ_P. We observe that varying the parameters $f_{\rm E,min}$ and $f_{\rm E,max}$, which specify the amount of energy injected in SN feedback by stellar particles formed in low- and high-pressure gas environments, respectively, can make the SN feedback weaker or stronger, resulting in more or less stellar mass formed and more or less compact galaxies. The combined effect on the GSMF and SSM of varying $f_{\rm E,min}$ and $f_{\rm E,max}$ can be well captured by solely changing $P_{\rm birth,0}$ (compare panels d and e versus panel f). Based on this finding, we decided to optimize $P_{\rm birth,0}$, while keeping $f_{\rm E,min}$ and $f_{\rm E,max}$ fixed. Because a low value of $f_{\rm E,min}$ yields slightly better galaxy sizes at low stellar mass (see panel d in Fig. 5.15), we set $f_{\rm E,min}$ to 0.1. As for $f_{\rm E,max}$, we set its value to 3.5, which gives an average value of the energy per SN within $1 - 2 \times 10^{51}$ erg. We note that we could use a somewhat different value for $f_{\rm E,max}$ by adjusting $P_{\rm birth,0}$ due to the degeneracy between these two parameters (panel e vs. panel f). The remaining parameter of the relation between $f_{\rm E}$ and $P_{\rm birth}$ is σ_P, which is shown in panel (g). We find that a nearly one-order-of-magnitude change in the value of σ_P, which determines how steeply $f_{\rm E}$ increases with $P_{\rm birth}$, has hardly any impact on the simulated GSMF and SSM of galaxies that are resolved with more than

$\sim 10^2$ particles ($M_* \gtrsim 2\times10^9\,\mathrm{M}_\odot$). Hence, we did not include this parameter in $\boldsymbol{\theta}$.

Lastly, the remaining three panels – (h), (i), and (j) – show the effect of varying AGN-related parameters: $M_{\rm BH,seed}$, $M_{\rm FOF,seed}$, and $\Delta T_{\rm AGN}$. These parameters affect the most massive galaxies in the simulation, where AGN feedback dominates. First, by comparing panels (h) and (i), we find that the effect of increasing (decreasing) $M_{\rm BH,seed}$ by ≈ 0.5 dex is almost exactly the same as if we would decrease (increase) $M_{\rm FOF,seed}$ by the same amount. In other words, it suffices to optimize only one of these two parameters. We chose to optimize $M_{\rm BH,seed}$, while $M_{\rm FOF,seed}$ was fixed to $10^{10}\,\mathrm{M}_\odot$. The impact of varying the AGN heating temperature by ±0.5 dex, which is shown in panel (j), is insignificant for the GSMF but shifts the sizes of massive objects by ≈ 0.1 dex (at a fixed stellar mass). Because this difference is rather small and because a similar effect can be achieved by varying $M_{\rm BH,seed}$, we opted to freeze $\Delta T_{\rm AGN}$ at 10^9 K. As we explained in §5.4.2, we verified that $\Delta T_{\rm AGN} = 10^9$ K yields plausible gas fractions of galaxy clusters in the Reference model, which are sensitive to $\Delta T_{\rm AGN}$ (e.g. Le Brun et al., 2014; McCarthy et al., 2017; Kugel et al., 2023).

5

5.6 Conclusions

We presented the calibration of a new subgrid model for cosmological hydrodynamical galaxy formation: COLIBRE (Schaye et al., in preparation). COLIBRE is based on the EAGLE galaxy-formation model (Schaye et al., 2015) with a large number of improvements and modifications. The most significant ones are: (i) the presence of a cold interstellar gas phase; (ii) the suppression of spurious heat transfer from dark matter to baryons by using four times more dark matter particles than baryonic particles; (iii) a model for the formation and evolution of dust, which is coupled to the chemistry; (iv) the use of a non-equilibrium network for the calculation of radiative cooling rates and ion and molecular fractions of hydrogen and helium; and (v) improved prescriptions for the modelling of subgrid physics processes, including the prescriptions for star formation and energy feedback from stars and SMBHs.

In order to calibrate the COLIBRE model, we made use of Gaussian process emulators. The emulators were trained on the actual simulations. Each simulation was run for a unique combination of values of the model's subgrid parameters, allowing the emulators to learn how galaxy properties vary as functions of a subset of the subgrid parameters. These parameters are (i) the fraction of SN energy injected in kinetic form, $f_{\rm kin}$, (ii) the pivot density in the thermal channel of SN feedback with a variable heating temperature, $n_{\rm SN,0}$; (iii) the pivot stellar birth pressure in the relation between the SN energy and stellar birth pressure, $P_{\rm birth,0}$; and (iv) the seed mass of SMBHs, $M_{\rm BH,seed}$. All simulations in the training set were run in a (50 comoving Mpc)^3 volume at the gas-mass resolution of $1.47\times10^7\,\mathrm{M}_\odot$. By fitting the trained emulators to the $z = 0$ observed galaxy stellar

mass function (GSMF) and to the $z = 0$ observed galaxy size-stellar mass relation (SSM), we found the values of the model's subgrid parameters that result in the best agreement with the target observational data.

The prescription for SN feedback in the COLIBRE model assumes that (i) stellar particles inject their SN energy into surrounding gas in both thermal and kinetic forms, (ii) the heating temperature in the thermal channel, $\Delta T_{\rm SN}$, is an increasing function of the gas density (Eq. 5.14), and the energy per SN is an increasing function of stellar birth gas pressure (Eq. 5.8). In order to demonstrate that these assumptions are all necessary to successfully reproduce the observed GSMF and SSM, we explored three variations of the COLIBRE model in which the modelling of SN feedback was significantly simplified:

(i) We first considered the Basic model, in which the energy in SN feedback is constant and is only injected thermally, stochastically heating the gas by a constant value of $\Delta T_{\rm SN} = 10^{7.5}$ K.

(ii) Our second simplified model was ThermalKinetic, which allows some fraction of the SN energy, $f_{\rm kin}$, to be injected kinetically via low-energy kicks with the kick velocity of 50 km s^{-1}, while the remainder is injected thermally as in the Basic model.

(iii) Finally, in the third simplified model, ThermalKineticVariableΔT, the heating temperature $\Delta T_{\rm SN}$ increases with the density of the gas surrounding the SNe. Compared to the Reference model, this model uses a constant energy per SN, as opposed to the SN energy that increases with the stellar birth pressure that is adopted in the Reference model.

These three simplified models were fit to the $z = 0$ observed GSMF and SSM in the same manner as the Reference model, and for each model the best-fitting subgrid parameter values were found. In total, we ran 163 simulations for various combinations of the subgrid parameters and models. Our main results with regard to the calibration are as follows:

- Owing to its naive prescription for SN feedback, the Basic model fails to produce a good fit to the $z = 0$ observed GSMF (Fig. 5.2). The GSMF exhibits a power-law shape as opposed to the observed Schechter (1976) shape. Increasing or decreasing the SN energy, which is described by the subgrid parameter $f_{\rm E}$, cannot resolve this discrepancy (middle panel of Fig. 5.4).

- The ThermalKinetic model can successfully reproduce the $z = 0$ observed GSMF or the observed SSM separately but cannot fit both relations simultaneously (Fig. 5.5). The fact that ThermalKinetic can provide a good match to the observed GSMF is a consequence of the ability to use the large energy injections of the thermal channel of SN feedback and low-energy kicks of

the kinetic channel (Fig. 5.4). The relative strengths of the two channels are optimized by the emulators via the parameter $f_{\rm kin}$: the model fit to the observed GSMF (SSM) prefers $f_{\rm kin} \approx 0.72$ ($f_{\rm kin} \approx 0$), while fitting to both constraints gives an intermediate value of $f_{\rm kin} \approx 0.37$ (Fig. 5.6).

- Adopting a density-dependent heating temperature $\Delta T_{\rm SN}$ in the ThermalKineticVariableΔT model improves the combined fit to the GSMF and SSM. Adding a stellar birth pressure dependence of the SN energy in the Reference model results in excellent agreement with the observed GSMF and SSM (Fig. 5.7).

Having calibrated each model to the observed GSMF and SSM, we proceeded to compare the best-fitting versions of each model to a number of observables that were not considered in the calibration:

- The $z = 0$ observed sSFR and the galaxy passive fractions are broadly matched by the ThermalKineticVariableΔT and Reference models, but not by Basic and ThermalKinetic (Fig. 5.9). The SN feedback with a constant $\Delta T_{\rm SN}$ of $10^{7.5}$ K, which is employed in the latter two models, is overly powerful, leading to a lack of star-forming gas by $z = 0$ in low- and intermediate-mass galaxies.

- The observed mass fractions of molecular and atomic gas are reproduced by the ThermalKineticVariableΔT and Reference models at stellar masses $10^9 < M_* < 10^{10.5}\,{\rm M}_\odot$ (Fig. 5.12). At higher masses, $M_* > 10^{10.5}\,{\rm M}_\odot$, the gas fractions appear to be too low in all but the Basic model, which, however, underpredicts the passive fraction for those stellar masses (Fig. 5.9).

- Owing to its stellar birth pressure dependence of the energy in SN feedback, the Reference model is the only model that provides a reasonably good match to the observed GSMF and sSFR at $z = 2$ (Fig. 5.11).

- Similarly, due to the pressure dependence of its SN feedback, the cosmic star formation rate density (SFRD) in the Reference model is suppressed at high z relative to the other three models (Fig. 5.10). As a result, the SFRD in the Reference model has a broad peak between $1 < z < 4$ and only begins to decline steeply below $z \approx 1$, which agrees with observations. In contrast, in the other models, the SFRD is a steeply declining function of cosmic time already after $z \approx 3$.

- The observed relations between galaxy stellar mass and stellar and gas-phase metallicities are reproduced in all four models (Fig. 5.13).

- The $z = 0$ masses of SMBHs, as well as their scaling with the stellar mass of the host galaxy, are consistent with observations for all four models (Fig. 5.9).

We then proceeded to investigate the effect of changing the model parameters that were not optimized with the use of emulators. We confirmed that these parameters are either degenerate with the parameters that were optimized and/or have little impact on the $z = 0$ GSMF and SSM (Figs. 5.14 and 5.15).

In closing, we stress that calibrating a galaxy formation model is a numerically demanding process with no guarantee of success. The fact that the Reference model fit to the $z = 0$ GSMF and SSM reproduces so many observed relations (Figs. 5.9, 5.10, 5.11, 5.12, 5.13), to which it was not tuned, is an encouraging result. In future work, we will investigate the performance of the COLIBRE model using larger cosmological volumes and different resolutions.

Acknowledgements

This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. EC is supported by the funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860744 (BiD4BESt). The research in this paper made use of the SWIFT open-source simulation code (http://www.swiftsim.com, Schaller et al. 2018) version 0.9.0. The data analysis was carried out with the use of SWIFTSIMIO (Borrow & Borrisov, 2020; Borrow & Kelly, 2021), NUMPY (Harris et al., 2020), MATPLOTLIB (Hunter, 2007), and SEABORN (Waskom, 2021).

Data Availability

The SWIFT simulation code is publicly available at http://www.swiftsim.com. The data underlying this article will be shared on reasonable request to the corresponding author.

References

Abbott T. M. C., et al., 2022, Phys. Rev. D, 105, 023520
Ali-Haïmoud Y., Bird S., 2013, MNRAS, 428, 3375
Asano R. S., Takeuchi T. T., Hirashita H., Inoue A. K., 2013, Earth, Planets, and Space, 65, 213
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, Annual Review of Astronomy and Astrophysics, 47, 481
Bahé Y. M., et al., 2016, MNRAS, 456, 1115
Bahé Y. M., et al., 2022, MNRAS, 516, 167
Barber C., Crain R. A., Schaye J., 2018, MNRAS, 479, 5448
Barber C., Schaye J., Crain R. A., 2019, MNRAS, 483, 985
Bauer A. E., et al., 2013, MNRAS, 434, 209
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, MNRAS, 512, 3703
Black J. H., 1987, Heating and Cooling of the Interstellar Gas, doi:10.1007/978-94-009-3861-8_27
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Borrow J., Borrisov A., 2020, Journal of Open Source Software, p. 2430
Borrow J., Kelly A. J., 2021, arXiv e-prints, p. arXiv:2106.05281
Borrow J., Schaller M., Bower R. G., Schaye J., 2022, MNRAS, 511, 2367
Borrow J., Schaller M., Bahé Y. M., Schaye J., Ludlow A. D., Ploeckinger S., Nobels F. S. J., Altamura E., 2023, MNRAS,
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Cañas R., Elahi P. J., Welker C., del P Lagos C., Power C., Dubois Y., Pichon C., 2019, MNRAS, 482, 2039
Cappellari M., et al., 2012, Nature, 484, 485
Catinella B., et al., 2018, MNRAS, 476, 875
Chabrier G., 2003, PASP, 115, 763
Chaikin E., Schaye J., Schaller M., Bahé Y. M., Nobels F. S. J., Ploeckinger S., 2022, MNRAS, 514, 249
Chaikin E., Schaye J., Schaller M., Benítez-Llambay A., Nobels F. S. J., Ploeckinger S., 2023, MNRAS, 523, 3709
Chang Y.-Y., van der Wel A., da Cunha E., Rix H.-W., 2015, ApJS, 219, 8
Chen Z., Stark D. P., Endsley R., Topping M., Whitler L., Charlot S., 2023, MNRAS, 518, 5607
Cinquegrana G. C., Karakas A. I., 2022, MNRAS, 510, 1557
Cochrane R. K., et al., 2023, MNRAS, 523, 6082
Crain R. A., van de Voort F., 2023, ARA&A, 61, 473
Crain R. A., et al., 2015, MNRAS, 450, 1937
Curti M., Mannucci F., Cresci G., Maiolino R., 2020, MNRAS, 491, 944
Dalgarno A., McCray R. A., 1972, ARA&A, 10, 375
Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431
Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, ApJ, 676, 33
Doherty C. L., Gil-Pons P., Lau H. H. B., Lattanzio J. C., Siess L., 2014, MNRAS, 437, 195

Driver S. P., et al., 2011, MNRAS, 413, 971
Driver S. P., et al., 2022, MNRAS, 513, 439
Dubois Y., et al., 2014, MNRAS, 444, 1453
Dubois Y., et al., 2021, A&A, 651, A109
Durier F., Dalla Vecchia C., 2012, MNRAS, 419, 465
Elahi P. J., Cañas R., Poulton R. J. J., Tobar R. J., Willis J. S., Lagos C. d. P., Power C., Robotham A. S. G., 2019, PASA, 36, e021
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, PASA, 34, e058
Enia A., et al., 2022, ApJ, 927, 204
Faucher-Giguère C.-A., 2020, MNRAS, 493, 1614
Feldmann R., et al., 2023, MNRAS, 522, 3831
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Ferland G. J., et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385
Fishlock C. K., Karakas A. I., Lugaro M., Yong D., 2014, ApJ, 797, 44
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fraser-McKelvie A., et al., 2022, MNRAS, 510, 320
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Gardner J., et al., 2006, Space Science Reviews, 123, 485
Giménez-Arteaga C., et al., 2023, ApJ, 948, 126
Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 728, 88
Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Graham A. W., Sahu N., 2023, MNRAS, 518, 2177
Greengard L., Rokhlin V., 1987, Journal of Computational Physics, 73, 325
Gruppioni C., et al., 2020, A&A, 643, A8
Gurman A., Hu C.-Y., Sternberg A., van Dishoeck E. F., 2023, arXiv e-prints, p. arXiv:2308.07338
Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, MUSIC2-monofonIC: 3LPT initial condition generator, Astrophysics Source Code Library, record ascl:2008.024 (ascl:2008.024)
Hahn O., Rampf C., Uhlemann C., 2021, MNRAS, 503, 426
Hardwick J. A., Cortese L., Obreschkow D., Catinella B., Cook R. H. W., 2022, MNRAS, 509, 3751
Harris C. R., et al., 2020, Nature, 585, 357
Hennebelle P., Chabrier G., 2011, ApJ, 743, L29
Hirashita H., 2000, PASJ, 52, 585
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., et al., 2018, MNRAS, 480, 800
Hopkins P. F., et al., 2020, MNRAS, 492, 3465
Hunt L. K., Tortora C., Ginolfi M., Schneider R., 2020, A&A, 643, A180
Hunter J. D., 2007, Computing in Science and Engineering, 9, 90
Jenkins A., 2013, MNRAS, 434, 2094
Karakas A. I., 2010, MNRAS, 403, 1413
Karakas A. I., Lugaro M., 2016, ApJ, 825, 26
Kennicutt Robert C. J., 1998a, ApJ, 498, 541
Kennicutt Robert C. J., 1998b, ApJ, 498, 541
Kennicutt Robert C. J., et al., 2007, ApJ, 671, 333
Khusanova Y., et al., 2021, A&A, 649, A152
Kimm T., Cen R., 2014, ApJ, 788, 121
Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, ApJ, 779, 102
Klessen R. S., Lin D. N., 2003, Phys. Rev. E, 67, 046311
Kobayashi C., Leung S.-C., Nomoto K., 2020, ApJ, 895, 138
Krumholz M. R., Federrath C., 2019, Frontiers in Astronomy and Space Sciences, 6
Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36

Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304
Krumholz M. R., McKee C. F., Klein R. I., 2006, ApJ, 638, 369
Kugel R., Borrow J., 2022, Journal of Open Source Software, 7, 4240
Kugel R., et al., 2023, MNRAS, 526, 6103
Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, MNRAS, 441, 1270
Lee E. J., Miville-Deschênes M.-A., Murray N. W., 2016, ApJ, 833, 229
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
Li H., et al., 2017, ApJ, 838, 77
López-Sánchez Á. R., Dopita M. A., Kewley L. J., Zahid H. J., Nicholls D. C., Scharwächter J., 2012, MNRAS, 426, 2630
Ludlow A. D., Schaye J., Bower R., 2019, MNRAS, 488, 3663
Ludlow A. D., Fall S. M., Schaye J., Obreschkow D., 2021, MNRAS, 508, 5114
Ludlow A. D., Fall S. M., Wilkinson M. J., Schaye J., Obreschkow D., 2023, MNRAS, 525, 5614
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Martín-Navarro I., La Barbera F., Vazdekis A., Falcón-Barroso J., Ferreras I., 2015, MNRAS, 447, 1033
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
McKay M. D., Beckman R. J., Conover W. J., 1979, Technometrics, 21, 239
Meijerink R., Spaans M., 2005, A&A, 436, 397
Michaux M., Hahn O., Rampf C., Angulo R. E., 2021, MNRAS, 500, 663
Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822
Moustakas J., et al., 2013, ApJ, 767, 50
Nobels F. S. J., Schaye J., Schaller M., Ploeckinger S., Chaikin E., Richings A. J., 2023, arXiv e-prints, p. arXiv:2309.13750
Nomoto K., Kobayashi C., Tominaga N., 2013, ARA&A, 51, 457
Novak M., et al., 2017, A&A, 602, A5
Ostriker E. C., 1999, ApJ, 513, 252
Padoan P., Nordlund Å., 2011, ApJ, 730, 40
Pakmor R., et al., 2023, MNRAS, 524, 2539
Péroux C., Howk J. C., 2020, ARA&A, 58, 363
Pillepich A., et al., 2018, MNRAS, 473, 4077
Ploeckinger S., Schaye J., 2020, MNRAS, 497, 4857
Ploeckinger S., Nobels F. S. J., Schaller M., Schaye J., 2023, arXiv e-prints, p. arXiv:2310.10721
Portinari L., Chiosi C., Bressan A., 1998, A&A, 334, 505
Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427
Ramos Almeida C., et al., 2022, A&A, 658, A155
Rasmussen C. E., Williams C. K. I., 2006, Gaussian processes for machine learning.. Adaptive computation and machine learning, MIT Press
Richings A. J., Schaye J., Oppenheimer B. D., 2014a, MNRAS, 440, 3349
Richings A. J., Schaye J., Oppenheimer B. D., 2014b, MNRAS, 442, 2780
Richings J., Frenk C., Jenkins A., Robertson A., Schaller M., 2021, MNRAS, 501, 4657
Saintonge A., et al., 2017, ApJS, 233, 22
Schaller M., et al., 2018, SWIFT: SPH With Inter-dependent Fine-grained Tasking, Astrophysics Source Code Library (ascl:1805.020)
Schaller M., et al., 2023, arXiv e-prints, p. arXiv:2305.13380
Schaye J., et al., 2010, MNRAS, 402, 1536
Schaye J., et al., 2015, MNRAS, 446, 521
Schaye J., et al., 2023, MNRAS, 526, 4978
Schechter P., 1976, ApJ, 203, 297
Schmidt M., 1959, ApJ, 129, 243
Shen S., Wadsley J., Stinson G., 2010, MNRAS, 407, 1581
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Springel V., Pakmor R., Zier O., Reinecke M., 2021, MNRAS, 506, 2871

Stanway E. R., Eldridge J. J., 2018, MNRAS, 479, 75
Stepnik B., et al., 2003, A&A, 398, 551
Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Tacconi L. J., Genzel R., Sternberg A., 2020, Annual Review of Astronomy and Astrophysics, 58, 157
Teyssier R., 2002, A&A, 385, 337
Thomas J., et al., 2011, MNRAS, 415, 545
Tremonti C. A., et al., 2004, ApJ, 613, 898
Wadsley J. W., Veeravalli G., Couchman H. M. P., 2008, MNRAS, 387, 427
Waskom M. L., 2021, Journal of Open Source Software, 6, 3021
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, MNRAS, 399, 574
Wolfire M. G., McKee C. F., Hollenbach D., Tielens A. G. G. M., 2003, ApJ, 587, 278
Woosley S. E., Eastman R. G., Schmidt B. P., 1999, ApJ, 516, 788
Wootten A., Thompson A. R., 2009, Proceedings of the IEEE, 97, 1463
Wright E. L., et al., 2010, AJ, 140, 1868
de Graaff A., Trayford J., Franx M., Schaller M., Schaye J., van der Wel A., 2022, MNRAS, 511, 2544

Appendix

5.A The effect of energy feedback from type-Ia SNe

In Fig. 5.16 we show the $z = 0$ GSMF and SSM relations for the Reference model with the best-fitting parameters (light-blue) and a variation in which the energy feedback from type-Ia SNe is turned off (magenta). By comparing the two simulations, we conclude that energy feedback from type-Ia SNe has only a marginal impact on the calibrated GSMF and SSM. This confirms that the differences in the GSMF and SSM that are produced by varying the model parameters $\Delta T_{\rm SN,min}$, $\Delta T_{\rm SN,max}$, and $n_{\rm SN,0}$ (see, e.g., Figs. 5.14 and 5.15), which affect the thermal feedback from both CC and type-Ia SNe, are dominated by the changes in the feedback from CC SNe.

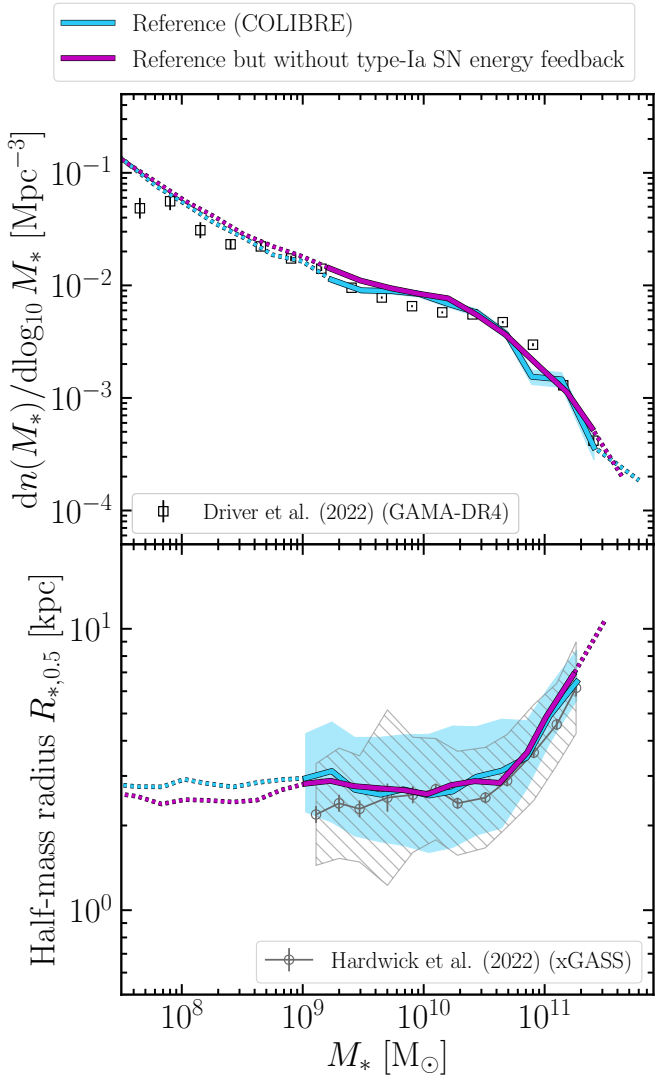

Figure 5.16: The impact of energy feedback from type-Ia SNe on the calibrated galaxy properties. Shown are the $z = 0$ galaxy stellar mass function (GSMF, *top panel*) and $z = 0$ median galaxy size-stellar mass relation (SSM, *bottom panel*). The style of the panels is the same as in Fig. 5.2. The light-blue curve corresponds to the simulation with the best-fitting Reference model. The simulation shown in magenta uses the same model parameter values as Reference but the energy feedback from type-Ia SNe is switched off. Turning off energy feedback from type-Ia SNe has only a marginal effect on the GSMF and SSM.