Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **548,** 1–42 (2026) https://doi.org/10.1093/mnras/stag300

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

Evgenii Chaikin [1★] Joop Schaye [1] Matthieu Schaller [2,1] Sylvia Ploeckinger [3] Yannick M. Bahé [4,5] Alejandro Benítez-Llambay [6] Camila Correa [1] Victor J. Forouhar Moreno [1] Carlos S. Frenk,[7] Filip Huško [1] Roi Kugel [1] Robert McGibbon [1] Alexander J. Richings [8,9] James W. Trayford [10] Josh Borrow [11] Robert A. Crain [12] John C. Helly,[7] Cedric G. Lacey [7] Aaron Ludlow [13] and Folkert S. J. Nobels [1,14]

[1]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[2]*Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, NL-2300 RA Leiden, the Netherlands*
[3]*Department of Astrophysics, University of Vienna, Türkenschanzstrasse 17, A-1180 Vienna, Austria*
[4]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[5]*Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[6]*Dipartimento di Fisica G. Occhialini, Università degli Studi di Milano Bicocca, Piazza della Scienza, 3, I-20126 Milano MI, Italy*
[7]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[8]*Centre for Data Science, Artificial Intelligence and Modelling, University of Hull, Cottingham Road, Hull HU6 7RX, UK*
[9]*E. A. Milne Centre for Astrophysics, University of Hull, Cottingham Road, Hull HU6 7RX, UK*
[10]*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK*
[11]*Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA 19104, USA*
[12]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[13]*International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia*
[14]*Netherlands Organisation for Applied Scientific Research (TNO), Molengraaffsingel 8, NL-2629 JD Delft, the Netherlands*

Accepted 2026 February 3. Received 2026 January 14; in original form 2025 September 1

ABSTRACT

We present the calibration of stellar and active galactic nucleus (AGN) feedback in the subgrid model for the new COLIBRE hydrodynamical simulations of galaxy formation. COLIBRE directly simulates the multiphase interstellar medium and the evolution of dust grains, which is coupled to the chemistry. COLIBRE is calibrated at three resolutions: particle masses of $m_{\rm gas} \approx m_{\rm dm} \sim 10^7$ (m7), 10^6 (m6), and $10^5\,{\rm M}_\odot$ (m5). To calibrate the COLIBRE feedback at m7 resolution, we run Latin hypercubes of ≈ 200 simulations that vary up to four subgrid parameters in cosmological volumes of $(50\,{\rm cMpc})^3$. We train Gaussian process emulators on these simulations to predict the $z = 0$ galaxy stellar mass function (GSMF) and size–stellar mass relation (SSMR) 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 prescriptions for supernova and AGN feedback affect the predictions. In particular, we demonstrate that while the observed $z = 0$ GSMF and SSMR can be matched individually with a relatively simple supernova feedback model, simultaneously reproducing both necessitates a more sophisticated prescription. We show that the calibrated m7 COLIBRE model not only reproduces the calibration target observables, but also matches various other galaxy properties to which the model was not calibrated. Finally, we apply the calibrated m7 model to the m6 and m5 resolutions and, after slight manual adjustments of the subgrid parameters, achieve a similar level of agreement with the observed $z = 0$ GSMF and SSMR.

Key words: methods: numerical – galaxies: general – galaxies: formation – galaxies: evolution.

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 R. A. Crain & F. van de Voort 2023 for a recent review). The rapid growth of computational facilities has opened up the possibility of simulating large cosmological volumes ($\gtrsim 100^3$ comoving Mpc^3; hereafter cMpc^3) with self-consistent modelling of baryonic processes (e.g. J. Schaye et al. 2010; Y. Dubois et al. 2014; J. Schaye et al. 2015; I. G. McCarthy et al. 2017; M. Tremmel et al. 2017; A. Pillepich et al. 2018; R. Davé et al. 2019; S. Bird et al. 2022; R. Pakmor et al. 2023; J. Schaye et al. 2023; K. Dolag et al. 2025), as well

★ E-mail: chaikin@strw.leidenuniv.nl

as studying the properties of the multiphase interstellar medium (ISM) of galaxies in smaller volumes situated in a cosmological environment (e.g. Y. Dubois et al. 2021; R. 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. V. Springel et al. 2021; M. Schaller et al. 2024).

At the same time, the advent of advanced observational facilities, such as the Atacama Large Millimeter/submillimeter Array (ALMA; A. Wootten & A. R. Thompson 2009) and *James Webb Space Telescope* (*JWST*; J. 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-)kpc resolution either through CO rotation-line emission at $z \lesssim 2$ (e.g. R. Ikeda et al. 2022; C. Ramos Almeida et al. 2022) or via [C II] 158 μm line at higher redshifts (e.g. O. Le Fèvre et al. 2020; M. Béthermin et al. 2023). The properties of the warmer gas of high$-z$ objects can be studied at comparable resolution with *JWST*, using emission lines such as [O III] or H β (e.g. Z. Chen et al. 2023; C. Giménez-Arteaga et al. 2023). Clearly, for a fair comparison between theory and observations, these cutting-edge observational 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 yr have achieved remarkable success in matching observational data and producing galaxies with realistic properties (see e.g. R. A. Crain & F. van de Voort 2023, and references therein). Various observed relations are broadly reproduced by the simulations, including the observed galaxy stellar mass functions (GSMFs) and luminosity functions at different redshifts, the galaxy size–stellar mass relation (SSMR), the galaxy star-forming main sequence, the galaxy stellar mass–metallicity relation and many others (e.g. J. Schaye et al. 2015; A. Pillepich et al. 2018; R. Davé et al. 2019). However, those successful simulations largely neglected the modelling of the cold neutral gas, which is believed to play a key role in the cosmic baryon cycle (e.g. C. Péroux & J. C. Howk 2020).

In fact, the large, high-resolution cosmological-volume simulations from the past decade such as HORIZONAGN (Y. Dubois et al. 2014), EAGLE (J. Schaye et al. 2015), ILLUSTRISTNG (A. Pillepich et al. 2018), and SIMBA (R. 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 two-fold. First, simulating the cold phase is computationally expensive because of the small time-steps and the small Jeans lengths and masses that are readily reached in the dense, cold phase. Secondly, modelling the cold phase cannot be accomplished without accounting for physical processes that are important in this regime. These processes include (self-)shielding of gas from the extragalactic UV background and radiation from local sources, the formation and dissociation of molecules and cooling emission therefrom, and the formation and evolution of dust grains, including their interactions with the cold gas phase (e.g. L. J. Tacconi, R. Genzel & A. Sternberg 2020). It has also been argued that, unless an effective pressure floor is used, including cold gas in simulations that do not formally resolve the thermal Jeans mass in the cold ISM ($\lesssim 10^3\,\mathrm{M}_\odot$) is problematic (e.g. B. E. Robertson & A. V. Kravtsov 2008; J. Schaye & C. Dalla Vecchia 2008), as it can lead to numerical artefacts such as artificial fragmentation (M. R. Bate & A. Burkert 1997; J. K. Truelove et al. 1997). However, S. Ploeckinger et al. (2024) recently showed that imposing an effective pressure floor is unnecessary, as in galaxy formation simulations with softened gravity the thermal Jeans mass criterion based on Newtonian gravity is inappropriate: in unresolved regimes, the Jeans mass scale is set by the softened Jeans mass, which substantially exceeds the Newtonian value.

Among the most recent simulations of cosmological volumes that allow the gas to enter the cold phase are NEWHORIZON (Y. Dubois et al. 2021) and FIREBOX (R. Feldmann et al. 2023). The FIREBOX simulation used the FIRE2 galaxy formation model (P. F. Hopkins et al. 2018) and was run in a cosmological volume of $(22.1\ \mathrm{cMpc})^3$ down to redshift $z = 0$ with gas and dark-matter (DM) particle masses of, respectively, $m_{\rm gas} = 6.3 \times 10^4\,\mathrm{M}_\odot$ and $m_{\rm dm} = 3.3 \times 10^5\,\mathrm{M}_\odot$. As with the FIRE2 simulations, FIREBOX was run with the GIZMO mesh-less finite-mass hydrodynamic and gravity solver (P. F. Hopkins 2015). The element ionization states were calculated based on tabulated equilibrium results from the simulations with the photoionization code CLOUDY (G. J. 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 M. R. Krumholz & N. Y. Gnedin (2011). FIREBOX does not model the self-consistent evolution of dust grains but accounts for dust collisional heating/cooling and photo-electric heating using analytic expressions, assuming a constant dust-to-metal ratio (for further details, see P. F. Hopkins et al. 2018). The FIREBOX model does not include a prescription for active galactic nucleus (AGN) feedback. As shown by R. Feldmann et al. (2023), FIREBOX reproduces the mass–metallicity relations for both the stellar and gas-phase components, as well as the star-forming main sequence and the relations between galaxy H I and H_2 masses and stellar mass, although the $z = 0$ GSMF in FIREBOX is systematically higher than the observed GSMF.

The domain of the NEWHORIZON simulation is a zoom-in region of $\sim(16\ \mathrm{cMpc})^3$ taken from the larger, $(142\ \mathrm{cMpc})^3$ volume of the HORIZONAGN simulations (Y. 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 (R. Teyssier 2002). Inside the zoom-in region, the DM particle mass is equal to $1.2 \times 10^6\,\mathrm{M}_\odot$ and the cell size can reach ≈ 34 pc in the densest environments. The cooling of metal-enriched gas in NEWHORIZON is based on tabulated equilibrium rates from R. S. Sutherland & M. A. Dopita (1993) at temperatures above $\approx 10^4$ K and A. Dalgarno & R. A. 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 under a homogeneous, redshift-dependent UV background, whose intensity is exponentially suppressed at densities $n_{\rm H} \gtrsim 0.01\ \mathrm{cm}^{-3}$ due to self-shielding. No dust evolution model is included. As shown by Y. Dubois et al. (2021), NEWHORIZON agrees with the observed galaxy-averaged Kennicutt–Schmidt (KS) star formation law (C. J. Kennicutt 1998), the black hole mass–stellar mass relation (BSMR), and the relations between stellar mass and the gas-phase and stellar metallicities, though, similarly to FIREBOX, galaxies in NEWHORIZON tend to be overmassive, resulting in a discrepancy between the predicted stellar-to-halo mass ratios and observational data.

The results from NEWHORIZON and FIREBOX demonstrate that simulations of galaxy formation that include a cold ISM are challenging but possible. In this companion paper to J. Schaye et al. (2025), we present the calibration of the strengths of stellar and AGN feedback in the new galaxy formation model COLIBRE[1] (J. Schaye et al. 2025), which – like NEWHORIZON and FIREBOX – captures the multiphase nature of the ISM. The COLIBRE model builds upon the OWLS (J. Schaye et al. 2010) and EAGLE (J. Schaye et al. 2015) galaxy formation models with significant improvements on various fronts: in addition to the presence of the multiphase ISM, COLIBRE includes non-equilibrium gas cooling, a live dust model coupled to the chemistry, and more sophisticated prescriptions for star formation and feedback from stellar evolution and AGN. Furthermore, in the initial conditions (ICs) of the COLIBRE simulations, we use four dark matter particles per gas particle to minimize spurious collisional heating of galaxies, which can negatively impact the sizes, kinematics, and morphologies of their stellar components (A. D. Ludlow, J. Schaye & R. Bower 2019; A. D. Ludlow et al. 2021, 2023; M. J. Wilkinson et al. 2023).

The need for calibration of galaxy formation simulations has been discussed extensively in the literature (see e.g. J. Schaye et al. 2010; M. Vogelsberger et al. 2013; R. A. Crain et al. 2015; I. G. McCarthy et al. 2017, and in particular section 2.1 of J. Schaye et al. 2015). Briefly, due to the finite resolution and large dynamic range of cosmological simulations, many astrophysical processes that occur on smaller scales – such as stellar and AGN feedback – are unresolved or only partially resolved. As a result, these processes must be implemented via *subgrid* models, which typically involve free (also referred to as *subgrid*) parameters that cannot be determined from first principles. The purpose of a subgrid model applied to an unresolved process is to reproduce the *effective* impact of that process on the larger, resolved scales. The role of the free parameters is to control this effective impact while compensating for numerically induced effects, such as excessive radiative cooling of feedback-heated gas due to limited numerical resolution (e.g. C. Dalla Vecchia & J. Schaye 2012). In the case of supernova (SN) feedback, free parameters may include the (initial) mass loading and velocity of SN-driven winds (e.g. V. Springel & L. Hernquist 2003; J. Schaye et al. 2010; M. Vogelsberger et al. 2013; R. Davé, R. Thompson & P. F. Hopkins 2016; M. C. Smith et al. 2024), while for AGN feedback, they may include a boost to the accretion rate of supermassive black holes (SMBHs) or the energy released in a single AGN injection event (e.g. C. M. Booth & J. Schaye 2009; I. G. McCarthy et al. 2017; N. A. Henden et al. 2018; R. Kugel et al. 2023). By comparing predictions from simulations with different subgrid parameters to observational data (such as the observed $z = 0$ GSMF), one can identify the parameter values that yield the best agreement with observations. This process is termed *calibration*. A calibrated simulation loses its predictive power for the specific relations used in the calibration, but can still make genuine predictions at other redshifts and for quantities that were not used as calibration targets.

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 be infeasible for COLIBRE because of the computational cost of running a new simulation at each step in the parameter space. Instead, a far more efficient approach is to use an emulator (R. G. Bower et al. 2010; R. Kugel et al. 2023).

In this work, we calibrate the SN and AGN feedback in the COLIBRE model at three resolutions: particle masses of $m_{\rm gas} \approx m_{\rm dm} \sim 10^7$ (m7), 10^6 (m6), and $10^5\,{\rm M}_\odot$ (m5). The calibration at m7 resolution is carried out by exploiting machine-learning techniques, following the approach taken by R. Kugel et al. (2023) for the FLAMINGO simulations (J. Schaye et al. 2023). Based on a modest number of simulations at m7 resolution in $(50\,{\rm cMpc})^3$ volumes, we train Gaussian process emulators that reconstruct the GSMF and SSMR from the COLIBRE simulations as smooth functions of a small number of subgrid parameters. We then fit these emulators to the observed GSMF and SSMR at $z = 0$ in the stellar mass range $10^9 < M_*/{\rm M}_\odot < 10^{11.3}$ and obtain the best-fitting values for up to four subgrid parameters. Having calibrated the m7 COLIBRE model, we use it as a starting point to calibrate the m6 and m5 COLIBRE models. The calibration at m6 and m5 resolutions is performed through small, manual adjustments of the subgrid parameters of SN and AGN feedback relative to their best-fitting values at m7 resolution.

This work is structured as follows. In Section 2, we describe the most relevant aspects of the COLIBRE galaxy formation model. In Section 3, we present the details of the emulation. In Section 4, we outline our strategy for calibrating the COLIBRE model at m7 resolution using emulators. In Section 5, we present the results of the m7 calibration and its extensions to the higher resolutions, and we explore the effects of varying individual model parameters. In Section 6, we summarize our conclusions.

2 SIMULATIONS

The simulation methods are described in detail in J. Schaye et al. (2025). Here, we will provide a summary with an emphasis on the ingredients that are varied during the calibration of the COLIBRE model.

All simulations presented in this work were run with the astrophysical code SWIFT[2] (M. Schaller et al. 2024). SWIFT uses hybrid task-based parallelism to enable scalability to tens of thousands of cores. The equations of hydrodynamics are solved using the smoothed particle hydrodynamics (SPH) method with the density-energy scheme SPHENIX (J. Borrow et al. 2022), adopting its fiducial values for artificial viscosity and energy conduction. We use the quartic spline kernel with a resolution parameter $\eta = 1.2348$, which is the same value employed in the EAGLE simulations (see M. Schaller et al. 2015, for details). For the quartic spline, $\eta = 1.2348$ yields an effective number of neighbours within the kernel support of $N_{\rm ngb} = 64.9$, which satisfies the requirement for accurate density reconstruction from W. Dehnen & H. Aly (2012), who showed that reconstruction errors remain low when η is close to 1.2, while much higher values can trigger the pairing instability. We also use $\eta = 1.2348$ for black hole (BH) particles,[3] while for stellar particles η is reduced to 1.1642 to lower the computational cost. The gravity is solved with the use of the

[1] https://colibre-simulations.org

[2] www.swiftsim.com

[3] 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 modelling AGN and stellar feedback processes.

Fast Multiple Method (L. Greengard & V. Rokhlin 1987) for short-range forces and a particle-mesh method solved in Fourier space for long-range forces.

In this work, we use the simulation output at redshifts $z = 0$ and 2. To identify subhaloes in the simulation snapshots, we employ the publicly available subhalo finder HBT-HERONS (V. J. Forouhar Moreno et al. 2025), which is an updated version of the Hierarchical Bound Tracing algorithm (HBT+; J. Han et al. 2018). HBT-HERONS employs a history-based approach to identify subhaloes. The algorithm begins at the earliest simulation snapshot, using an iterative unbinding procedure to find self-bound subhaloes within spatial Friends-of-Friends (FoF) groups, and then processes each subsequent snapshot. Once a self-bound subhalo is identified, HBT-HERONS tracks its associated particles forward in time. Among these, the 10 most gravitationally bound tracer particles (dark matter or stars) are used to link subhaloes to a host FoF group, facilitating the identification of when subhaloes become satellites of a more massive host subhalo. The algorithm keeps track of the particles that were associated to satellite subhaloes before they became satellites. These tracked particles are used at later times to separate the satellite subhalo from the background of its larger host subhalo. At each simulation output time, all subhaloes are checked for self-boundness and phase-space overlap with neighbouring subhaloes, to determine whether they remain self-bound, merge with neighbouring subhaloes, or become disrupted.

We further process the HBT-HERONS output using the Spherical Overdensity and Aperture Processor (SOAP; R. McGibbon et al. 2025) to compute a comprehensive set of subhalo properties. The properties that are used in this work include subhalo stellar and halo masses, star formation rates (SFRs), projected stellar half-mass radii, gas and stellar metallicities, H I and H_2 gas masses, and masses of the most massive BH particles in the subhaloes. Halo masses (for central subhaloes) are computed using the spherical overdensity definition from G. L. Bryan & M. L. Norman (1998, their equation 6, which is a fitting formula to the results from V. R. Eke, S. Cole & C. S. Frenk 1996), whereas all other galaxy properties are measured within 3D spherical apertures of radius 50 proper kpc (pkpc), considering only gravitationally bound particles associated with each subhalo.

2.1 Initial conditions

The ICs of our simulations are produced by the MONOFONIC code (O. Hahn et al. 2020; M. Michaux et al. 2021) using second-order Lagrangian perturbation theory. We follow the 2-field prescription from O. Hahn, C. Rampf & C. Uhlemann (2021) to generate ICs for baryons and DM and take $z = 63$ as the starting redshift of the simulations. We use the '3x2pt + all external constraints' cosmology from T. M. C. 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 majority of the analysis in this work is based on simulations in a cosmological volume of $(50\,\mathrm{cMpc})^3$, with a particle mass of $m_{\rm gas} = 1.47 \times 10^7\,\mathrm{M}_\odot$ for baryons and $m_{\rm dm} = 1.94 \times 10^7\,\mathrm{M}_\odot$ for DM. Unless otherwise stated, this volume and resolution (henceforth, m7 resolution) are assumed throughout. The associated Plummer-equivalent gravitational softening length $\varepsilon_{\rm soft}$, which we set to be the same for gas, stellar, BH, and DM particles, is equal to the minimum of 3.6 ckpc and 1.4 pkpc. The mass of a DM particle is comparable to that of a gas particle because in the ICs, there are four corresponding DM particles for each gas particle. In total, the ICs of a $(50\,\mathrm{cMpc})^3$ volume simulation at m7 resolution include 376^3 gas particles and 4×376^3 DM particles. As shown by A. D. Ludlow et al. (2019, 2021, 2023), the finite number of particles in galaxy simulations makes galaxies prone to spurious energy transfer from dynamically hot DM to dynamically cold stars, which tend towards energy equipartition through gravitational interactions. Increasing the DM resolution (or, equivalently, the number of DM particles; e.g. by a factor of 4 relative to the number of gas particles in the ICs) reduces this spurious energy transfer. As a result, stellar particles experience less artificial heating from the DM, leading to more realistic structural and kinematic properties of the stellar components of galaxies.

The only exceptions where we consider volumes different from $(50\,\mathrm{cMpc})^3$ and resolutions other than m7 are in Section 5.4 and Appendix A. In Section 5.4, we compare the fiducial, calibrated COLIBRE models at three resolutions – m7 ($m_{\rm gas} = 1.47 \times 10^7\,\mathrm{M}_\odot$, $m_{\rm dm} = 1.94 \times 10^7\,\mathrm{M}_\odot$), m6 ($m_{\rm gas} = 1.8 \times 10^6\,\mathrm{M}_\odot$, $m_{\rm dm} = 2.4 \times 10^6\,\mathrm{M}_\odot$), and m5 ($m_{\rm gas} = 2.3 \times 10^5\,\mathrm{M}_\odot$, $m_{\rm dm} = 3.0 \times 10^5\,\mathrm{M}_\odot$) – all in a $(25\,\mathrm{cMpc})^3$ volume; while in Appendix A, we investigate the effect of box size by comparing the fiducial COLIBRE model at m7 resolution in cosmological volumes of 25^3, 50^3, 100^3, and 200^3 cMpc^3. Simulations at m6 (m5) resolution contain 8 (64) times more gas and DM particles in the ICs than their m7 counterparts in the same cosmological volume. The gravitational softening length-scales with resolution as $\varepsilon_{\rm soft} \propto m_{\rm gas}^{1/3}$. At a fixed resolution, increasing the cosmological volume by a factor of 2^3 results in 8 times more gas and DM particles in the ICs.

2.2 The COLIBRE model

In the following sections, we summarize the COLIBRE subgrid model at m7 resolution, which serves as the basis for building the emulators (Section 3) and performing the emulator-based calibration (Section 4). The COLIBRE models at m6 and m5 resolutions adopt slightly different subgrid parameter values for SN and AGN feedback compared to m7. The adjustments relative to the values reported in this section are detailed in Section 5.4 and summarized in table 1 of J. Schaye et al. (2025).

2.2.1 *Radiative cooling, chemistry, and dust*

The non-equilibrium abundances of hydrogen and helium species and associated free electrons, along with their radiative cooling and heating rates, are computed with the time-dependent thermochemistry solver CHIMES (A. J. Richings, J. Schaye & B. D. Oppenheimer 2014a, b). Additionally, we explicitly track nine heavy elements that contribute most to the cooling rates: C, N, O, Ne, Mg, Si, S, Ca, and Fe. Their contributions to the cooling are provided by HYBRID-CHIMES (S. Ploeckinger et al. 2025). HYBRID-CHIMES uses pre-computed cooling tables generated by CHIMES under the assumption of ionization equilibrium, but the rates are rescaled to account for the difference between the non-equilibrium and equilibrium free electron number densities. 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 grain model coupled to CHIMES, and account for cosmic ray heating, as well as Compton cooling and heating from energy exchange between the

gas and photons from the cosmic microwave background and other radiation fields included in COLIBRE (see below).

The cooling rates and the abundances of ions and molecules are evolved assuming the presence of a modified version of the uniform, redshift-dependent UV and X-ray background from C.-A. Faucher-Giguère (2020) (see appendix B of S. Ploeckinger & J. Schaye (2020) for details), a cosmic ray ionization background, and an interstellar radiation field (ISRF). The shape of the ISRF is constrained by that at the position of the Sun (J. H. Black 1987), combining the local interstellar radiation field (J. S. Mathis, P. G. Mezger & N. Panagia 1983) with the Galactic soft X-ray background (J. N. Bregman & J. P. Harrington 1986). For gas with temperatures below 10^4 K, the intensities of the cosmic ray background and ISRF scale as[4] $N_{\rm Jeans}^{1.4}$, where $N_{\rm Jeans}$ is the Jeans column density of gas with a 1D turbulent velocity dispersion of 6 km s^{-1}, and saturate at high column densities (see S. Ploeckinger et al. 2025 for details). Shielding by gas and dust is accounted for, assuming a shielding length equal to the Jeans length.

COLIBRE tracks the abundances of 12 individual elements: H, He, C, N, O, Ne, Mg, Si, Fe, Sr, Ba, and Eu.[5] All of the 12 elements are diffused among the neighbouring SPH gas particles following a velocity shear-based subgrid model for turbulent mixing described in Correa et al. (in preparation).

The COLIBRE simulations incorporate a subgrid model for the formation and evolution of interstellar dust grains, which is described in detail by J. W. Trayford et al. (2026). Briefly, dust is treated as a scalar field, with gas particles tracking the fraction of their mass that resides in dust grains. The dust model distinguishes between three chemical species 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 core collapse supernovae (CC SNe). Dust grains grow by accreting mass from the gas phase, while they are destroyed in SN feedback and lose mass in hot gas due to thermal sputtering. Additionally, the model includes two processes that alter grain sizes without changing 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 abundances are coupled to the CHIMES solver, accounting for the distribution of dust mass between the two grain-size bins. The dust is used by CHIMES to calculate the formation rate of molecular hydrogen on dust grains and other reactions facilitated by dust, as well as to compute dust shielding and dust-associated heating and cooling processes, including dust radiative cooling and photoelectric heating. Additionally, the gas-phase metal abundances (and therefore the metal cooling and heating rates) account for depletion onto dust grains.

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

[5]This set of elements is not identical to those used in the prescription for gas radiative cooling. In particular, the contributions of Sr, Ba and Eu to the cooling rates are neglected, while Ca and S – the elements used in the radiative cooling calculations – are not tracked in the COLIBRE chemistry. Instead, to reduce the memory footprint of the simulations, the abundances of Ca and S are assumed to have solar mass ratios relative to Si (M. Asplund et al. 2009). This is a reasonable approximation because Ca, S, and Si are all α-elements that track each other relatively well (R. P. C. Wiersma et al. 2009).

2.2.2 *Star formation, stellar evolution, and chemical enrichment*

The COLIBRE prescription for star formation is detailed in F. S. J. Nobels et al. (2024). A gas element is labelled as 'star-forming' if the gas is locally unstable against gravitational collapse. The instability condition is expressed by requiring that the (absolute) gravitational binding energy of a gas cloud represented by the gas element – with mass equal to the mass of the gas element multiplied by the effective number of neighbours within the SPH kernel – exceeds its kinetic energy from thermal and turbulent motions.

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 M. Schmidt (1959) law with a star formation efficiency per free-fall time of $\varepsilon = 0.01$.

A stellar particle physically represents a population of many stars that formed simultaneously from a single gas cloud with uniform chemical composition. We assume that all stellar particles are characterized by a G. Chabrier (2003) stellar initial mass function (IMF) with minimum and maximum masses of 0.1 and 100 $\mathrm{M}_\odot$, respectively. Besides standard properties such as position, velocity, and metallicity, which are inherited from the parent gas particle, a stellar particle is characterized by its age and initial mass.

Once formed, stellar particles enrich their surrounding gas with metals produced in six chemical enrichment channels: AGB stars, type-Ia SNe, CC SNe, neutron star mergers, common envelope jet SNe, and collapsars [see Correa et al. (in preparation) for further details].

The stellar feedback model includes three early stellar feedback processes from massive stars: stellar winds, direct radiation pressure, and H II regions. To determine the energies, momenta, and ionizing flux injected into the surrounding gas by these feedback processes, we use the Binary Population and Spectral Synthesis (BPASS) tables (J. J. Eldridge et al. 2017; E. R. Stanway & J. J. Eldridge 2018) version 2.2.1. These early feedback processes are not calibrated in this work; their numerical implementation and effects are presented in A. Benítez-Llambay et al. (2025).

2.2.3 *Core collapse supernova feedback*

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

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

$$\Delta E_{\rm CCSN} = 10^{51}\,\mathrm{erg}\, f_{\rm E}\, m_* \int_{m_{\rm d}(t+\Delta t)}^{m_{\rm d}(t)} \Phi(m)\,\mathrm{d}m\,, \tag{1}$$

in which $\Phi(m)$ is the G. 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 L. Portinari, C. Chiosi & A. Bressan (1998) to compute $m_{\rm d}(t)$. The function $m_{\rm d}(t)$ is non-zero only for zero-age main sequence masses between $m_{\rm min,CCSN} = 8$ and $m_{\rm max,CCSN} = 100\,\mathrm{M}_\odot$, which roughly correspond to stellar ages of ≈ 40 and 3 Myr, respectively.

Unlike E. 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 in the time-step it turned into the stellar particle under consideration. The

relation between $f_{\rm E}$ and $P_{\rm birth}$ has the following form

$$f_{\rm E}(P_{\rm birth}) = f_{\rm E,min} + \frac{f_{\rm E,max} - f_{\rm E,min}}{1 + \exp\left(-\frac{\log_{10} P_{\rm birth} - \log_{10} P_{\rm E,pivot}}{\sigma_{\rm P}}\right)}\,, \quad (2)$$

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, $P_{\rm E,pivot}$ is a normalization constant, which we will call the pivot birth pressure, and the parameter $\sigma_{\rm P}$ defines the width of the transition from $f_{\rm E,min}$ to $f_{\rm E,max}$. The functional form of equation (2) implies that the SN feedback of stellar particles born in higher gas pressure environments will be more energetic. In our fiducial setting for m7 resolution, we take $f_{\rm E,min} = 0.1$, $f_{\rm E,max} = 4$, and $\sigma_{\rm P} = 0.3$, while the best value of $P_{\rm E,pivot}$ will be predicted by emulators (see below). In Section 5.5, we will show how variations in $f_{\rm E,min}$, $f_{\rm E,max}$, and $\sigma_{\rm P}$ 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. J. Thomas et al. 2011; M. Cappellari et al. 2012; I. Martín-Navarro et al. 2015; H. Li et al. 2017) and a pressure-dependent IMF has been employed in numerical simulations to successfully reproduce the observational trends (e.g. C. Barber, R. A. Crain & J. Schaye 2018; C. Barber, J. Schaye & R. A. Crain 2019). Values of $f_{\rm E}$ greater than one can be regarded as accounting for hypernovae (e.g. S. E. Woosley, R. G. Eastman & B. P. Schmidt 1999), and/or as compensation for some degree of numerical overcooling in high-density (and typically also high-pressure) gas (e.g. G. Stinson et al. 2006; C. Dalla Vecchia & J. Schaye 2012).

The energy $\Delta E_{\rm CCSN}$ is injected stochastically into the gas within the SPH kernel of the stellar particle. As in E. 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. The value of $f_{\rm kin}$ will be determined using emulators.

As shown in E. Chaikin et al. (2023), the thermal and kinetic channels of energy injection operate side by side with distinct roles: the former generates a hot phase of the ISM and launches strong galactic winds, while the latter drives turbulence in the ISM. This provides galaxies with two complementary means to regulate star formation: by maintaining turbulent support of the gas within the ISM and by ejecting gas from the galaxy.

2.2.4 Thermal channel of energy injection

The thermal channel of CC SN feedback utilizes the stochastic model of C. Dalla Vecchia & J. Schaye (2012), which was employed in the EAGLE simulations (J. Schaye et al. 2015). In the C. Dalla Vecchia & J. 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, $\Delta E_{\rm heat}$, 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 $\Delta E_{\rm heat}$ and $\Delta T_{\rm SN}$ is expressed as

$$\Delta E_{\rm heat}(m_{\rm gas}, \Delta T_{\rm SN}) = \frac{k_{\rm B}\Delta T_{\rm SN}}{(\gamma - 1)}\frac{m_{\rm gas}}{\mu m_{\rm p}}\,, \quad (3)$$

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 heats 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, $\Delta E_{\rm heat}(m_{\rm ngb}, \Delta T_{\rm SN})$,

$$p_{\rm SN,heat} = (1 - f_{\rm kin})\frac{\Delta E_{\rm CCSN}(t, \Delta t, m_*, f_{\rm E})}{\Delta E_{\rm heat}(m_{\rm ngb}, \Delta T_{\rm SN})}\,, \quad (4)$$

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 uniform 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 E. 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 an effectively mass-weighted neighbour selection scheme, as opposed to the isotropic method from E. 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. A second significant change is the heating temperature increment $\Delta T_{\rm SN}$, which was constant in EAGLE but depends on the gas density in COLIBRE.

2.2.5 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 C. Dalla Vecchia & J. Schaye (2012). In short, values greater than $\sim 10^{7.5}$ K would lead to undersampling of SN feedback because the average 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

$$\begin{aligned}\langle N_{\rm heat,tot}\rangle &= \frac{(1 - f_{\rm kin})\,E_{\rm CCSN,tot}(m_*, f_{\rm E})}{\Delta E_{\rm heat}(\langle m_{\rm gas}\rangle, \Delta T_{\rm SN})}\,,\\ &= 0.91\,(1 - f_{\rm kin})\,f_{\rm E}\left(\frac{m_*}{\langle m_{\rm gas}\rangle}\right)\left(\frac{\Delta T_{\rm SN}}{10^{7.5}\,{\rm K}}\right)^{-1}\,, \quad (5)\end{aligned}$$

where $E_{\rm CCSN,tot}(m_*, f_{\rm E}) = 10^{51}\,{\rm erg}\,f_{\rm E}\,m_*\int_{m_{\rm min,CCSN}}^{m_{\rm max,CCSN}}\Phi(m)\,{\rm d}m$ is the total CC SN energy released by the stellar particle over the course of its lifetime and $\langle m_{\rm gas}\rangle$ is the average mass of its neighbouring gas particles. Assuming that $m_* \approx \langle m_{\rm gas}\rangle$, $f_{\rm E} \sim 1$, 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. G. Stinson et al. 2006; C.

Dalla Vecchia & J. Schaye 2012). Assuming that at high temperatures radiative cooling is dominated by bremsstrahlung, C. Dalla Vecchia & J. Schaye (2012) showed that the maximum density at which the feedback can remain efficient is

$$n_{\rm H,crit} = 2\,{\rm cm}^{-3} \left(\frac{\Delta T_{\rm SN}}{10^{7.5}\,{\rm K}}\right)^{3/2} \left(\frac{f_t}{10}\right)^{-3/2} \left(\frac{\langle m_{\rm gas}\rangle}{1.5\times10^7 {\rm M}_\odot}\right)^{-1/2} \times \left(\frac{\langle N_{\rm ngb}\rangle}{65}\right)^{-1/2} \left(\frac{\mu}{0.6}\right)^{-9/4} \left(\frac{g(X_{\rm H})}{0.14}\right)^{3/2}, \quad (6)$$

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.[6]

In COLIBRE, we exploit equation (6) to 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 at m7 resolution, we set $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$ to $10^{6.5}$ and $10^{7.5}$ K, respectively. The use of values lower than $10^{7.5}$ K greatly increases 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 (Y. M. Bahé et al. 2016).

More precisely, 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

$$\rho_{\rm SN} = \sum_{i=1}^{N_{\rm ngb}} m_{{\rm gas},i}\, W(|\boldsymbol{r}_* - \boldsymbol{r}_{{\rm gas},i}|, h_*)\,, \quad (7)$$

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

$$\Delta T_{\rm SN}(n_{\rm H,SN}) = \Delta T_{\rm SN,pivot} \left(\frac{n_{\rm H,SN}}{n_{\rm H,pivot}}\right)^{2/3}, \quad (8)$$

in which $\Delta T_{\rm SN,pivot}$ and $n_{\rm H,pivot}$ are free parameters and the slope of 2/3 is motivated by the cooling argument from C. Dalla Vecchia & J. Schaye (2012), our equation (6). Because of the degeneracy between $n_{\rm SN,pivot}$ and $\Delta T_{\rm SN,pivot}$ ($\Delta T_{\rm SN,pivot} \propto n_{\rm H,pivot}^{2/3}$), we fix $\Delta T_{\rm SN,pivot}$, setting it to $10^{6.5}$ K at m7 resolution and only consider $n_{\rm H,pivot}$ in the following.

[6]From separate tests, we found that at m7 resolution, the requirement of $f_t = 10$ proposed by C. Dalla Vecchia & J. 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 $\approx 20\,{\rm cm}^{-3}$. For comparison, the median density in our simulations at which CC SNe take place is $\sim 1\,{\rm cm}^{-3}$.

2.2.6 *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 C. Dalla Vecchia & J. Schaye (2008). The full details of our algorithm for SN kinetic feedback are presented in E. 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. E. 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 H I 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 E. 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[7] and the injected energy is distributed statistically isotropically. For further details, we refer the reader to E. Chaikin et al. (2023).

2.2.7 *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 E. 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 (8), where the values of $\Delta T_{\rm SN,min}$, $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,pivot}$, $n_{\rm H,pivot}$, 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 delay time distribution (DTD),

$${\rm DTD}(t) = \frac{\nu}{\tau}\exp\left(-\frac{t - t_{\rm delay}}{\tau}\right)\Theta(t - t_{\rm delay})\,, \quad (9)$$

in which $\nu = 1.54\times10^{-3}\,{\rm M}_\odot^{-1}$ 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) 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 SNe 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,

[7]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 $\Delta v_{\rm kick}$.

individual type-Ia 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 (9) from t to $t + \Delta t$ and assuming an energy per individual type-Ia SN of 10^{51} erg.

Energetically, type-Ia SN feedback is subdominant to that from CC SNe, and its presence has only a minor impact on the galaxy properties relevant for the calibration of the COLIBRE simulations (see Nobels et al., in preparation, for more details). Unless stated otherwise, all discussions about SN feedback in the following text will refer entirely to CC SN feedback.

2.2.8 Supermassive black holes

In galaxy simulations, 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. V. Springel, T. Di Matteo & L. Hernquist 2005; C. M. Booth & J. Schaye 2009).

We employ an on-the-fly FoF group finder to seed BH particles in the simulation (e.g. T. Di Matteo et al. 2008). The FoF algorithm uses a linking length of 0.2 times the mean dark matter inter-particle separation and is executed every $\Delta a = 0.00751\, a$, starting at $a = 0.05$, where a is the cosmic scale factor. At m7 resolution, BHs are seeded in haloes whose FoF mass is greater than $5 \times 10^{10}\,\mathrm{M}_\odot$ and that do not already harbour a BH particle.

During seeding, we identify 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 BHs 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. V. Springel et al. 2005; C. M. Booth & J. Schaye 2009). The subgrid mass is initially equal to the seed mass, $m_{\mathrm{BH,seed}}$. The value of $m_{\mathrm{BH,seed}}$ will be calibrated using emulators.

The (instantaneous) mass accretion rate onto a BH particle is computed using a modified Bondi–Hoyle–Lyttleton formula (M. R. Krumholz, C. F. McKee & R. I. Klein 2006),

$$\dot{m}_{\mathrm{accr}} = 4\pi\, G^2\, \frac{m_{\mathrm{BH}}^2\, \rho_{\mathrm{gas}}}{c_{\mathrm{sound}}^3} \left[\frac{(1+\mathscr{M}^2)^4}{1.1^2+\mathscr{M}^2} + \frac{1}{(0.34\, f_\star)^2} \right]^{-1/2}, \qquad (10)$$

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 $r_{\mathrm{Bondi}} = Gm_{\mathrm{BH}}/c_{\mathrm{sound}}^2$ is the Bondi radius with m_{BH} being the subgrid mass of the BH particle. The magnitude of the gas bulk velocity, $v_{\mathrm{gas}} \equiv |\boldsymbol{v}_{\mathrm{gas}}|$, 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. We calculate the gas mass density, ρ_{gas}, in the standard SPH way by applying equation (7) to the gas neighbours within the BH kernel. Finally, in COLIBRE the mass accretion rate, $\dot{m}_{\mathrm{accr}}$, is capped at 100 times the mass accretion rate at the Eddington luminosity.

Following Y. M. Bahé et al. (2022), two BH particles will merge if the distance between them, Δr_{BH}, is less than three gravitational softening lengths, $\Delta r_{\mathrm{BH}} < 3\varepsilon_{\mathrm{soft}}$; the less massive BH is within the kernel of the more massive BH; and if their relative velocity Δv_{BH} satisfies $\Delta v_{\mathrm{BH}} < \sqrt{2G(M+m)/\Delta r_{\mathrm{BH}}}$ 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.

SMBHs are thought to be 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. E. C. 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. T. Di Matteo et al. 2008; Y. M. Bahé et al. 2022). In COLIBRE we follow the method of Y. M. Bahé et al. (2022) where at every time-step Δt, each BH particle searches for the gas particle within its SPH kernel that has the lowest gravitational potential. If this gas particle is also within three gravitational softening lengths of the BH and has a lower potential than at the BH's current position, the BH is immediately 're-positioned' to the location of that gas particle. The velocity of the BH remains unchanged during the re-positioning. As recommended by Y. M. 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.

2.2.9 Feedback from AGN

The COLIBRE suite includes simulations with purely thermal AGN feedback as well as simulations with a hybrid AGN feedback mode that combines BH-spin dependent kinetic jets and thermal energy injections, with the largest COLIBRE volumes available only for the thermal models. This work focuses entirely on calibrating the COLIBRE simulations with purely thermal AGN feedback. The calibration of simulations with hybrid AGN feedback, which builds upon the results of this study, is described in F. Huško et al. (2026).

The purely thermal AGN feedback from SMBHs is implemented following C. M. Booth & J. Schaye (2009) and is similar to that used in EAGLE (J. Schaye et al. 2015). Out of the total gas mass accreted by a BH particle over a given time-step from t to $t + \Delta t$, $\dot{m}_{\mathrm{accr}}\Delta t$, the BH receives[8] a fraction

$$\Delta m_{\mathrm{BH}} = (1 - \varepsilon_{\mathrm{r}})\dot{m}_{\mathrm{accr}}\Delta t\,, \qquad (11)$$

where ε_{r} is the radiative efficiency. The remaining mass, $\varepsilon_{\mathrm{r}}\dot{m}_{\mathrm{accr}}\Delta t$, is assumed to have been converted into energy that escapes the BH as radiation, a fraction of which is coupled to the gas surrounding the BH. The energy received by the gas in the time-step Δt, ΔE_{AGN}, is

$$\Delta E_{\mathrm{AGN}} = \varepsilon_{\mathrm{f}}\varepsilon_{\mathrm{r}}\, \dot{m}_{\mathrm{accr}}\, c^2\, \Delta t\,, \qquad (12)$$

where ε_{f} is the coupling efficiency. For both ε_{r} and ε_{f}, we adopt a value of 0.1 at m7 resolution. The former is motivated by theoretical considerations (N. I. Shakura & R. A. Sunyaev 1973), while the latter was chosen to yield realistic $z = 0$ SMBH masses in high-mass $z \approx 0$ galaxies.

[8] 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 the 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 Y. M. Bahé et al. (2022). Conversely, if $m_{\mathrm{BH}}^{\mathrm{new}}$ is less than $m_{\mathrm{BH}}^{\mathrm{dyn}}$, then we assume that the difference $m_{\mathrm{BH}}^{\mathrm{dyn}} - m_{\mathrm{BH}}^{\mathrm{new}} > 0$ represents a subgrid gas reservoir around the BH and all accreted mass comes therefrom. We then only reduce $m_{\mathrm{BH}}^{\mathrm{dyn}}$ by $\varepsilon_{\mathrm{r}}\Delta m_{\mathrm{BH}}$ to account for the energy that has been converted into radiation. No gas particle's mass is nibbled in this case.

As is the case with stellar feedback (see discussion in Section 2.2.5), injecting the energy $\Delta E_{\rm AGN}$ into surrounding gas may result in numerical overcooling if $\Delta E_{\rm AGN}$ is insufficient to increase the temperature of the gas in which the energy is injected to values high enough for the cooling time to be long. Following C. M. Booth & J. Schaye (2009), we wait until a sufficiently large amount of energy has been accumulated by the accreting 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 at every time-step by the value of $\Delta E_{\rm AGN}$ for that time-step. Once the energy in the reservoir exceeds a threshold energy $\Delta E_{\rm AGN,thr}$, we inject the energy $\Delta E_{\rm AGN,thr}$ into one of the gas particles within the SPH kernel of the BH particle 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 (3). 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 the maximum number of particles that can be heated (see below). 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 $N_{\rm AGN}\,\Delta E_{\rm AGN,thr}$.

For all simulations used in the calibration of the COLIBRE model at m7 resolution, we employ $\Delta T_{\rm AGN} = 10^9$ K. This value, which is the same as that used in the EAGLE-RECAL model (J. Schaye et al. 2015), ensures that AGN feedback is efficient and well-sampled in massive haloes. However, after completing the calibration simulations and identifying the best-fitting m7 model with $\Delta T_{\rm AGN} = 10^9$ K, we found that a fixed value of $\Delta T_{\rm AGN} = 10^9$ K was not the optimal choice for higher resolutions. Instead, a BH mass-dependent $\Delta T_{\rm AGN}$ (equation 20) proved to be a better option. Therefore, while the calibration at m7 resolution (described in Sections 3 and 4) is based on $\Delta T_{\rm AGN} = 10^9$ K, the fiducial COLIBRE m7 model adopts a variable $\Delta T_{\rm AGN}$, to be consistent with the fiducial models at higher resolutions. This choice is justified in more detail in Section 5.4, where we also present comparisons between the best-fitting m7 model with $\Delta T_{\rm AGN} = 10^9$ K and the fiducial COLIBRE model with variable $\Delta T_{\rm AGN}$ in Figs 14 and 15. To avoid confusion, the COLIBRE model with fixed $\Delta T_{\rm AGN} = 10^9$ K will consistently be referred to as such throughout the text. The final version, which employs a variable $\Delta T_{\rm AGN}$, will be referred to as the fiducial model, the COLIBRE model with variable $\Delta T_{\rm AGN}$, or simply the COLIBRE model.

To select the gas neighbours that will receive the energy $\Delta E_{\rm AGN,thr}$, we employ the 'Minimum Distance' algorithm from E. Chaikin et al. (2022). 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 Y. M. Bahé et al. (2022), the maximum number of gas neighbours a BH particle can heat in a single time-step, $N_{\rm BH,max}$, is set to 50.[9]

[9]In rare events where $N_{\rm AGN}$ is greater than $N_{\rm BH,max}$, we increase $\Delta T_{\rm AGN}$ by $N_{\rm AGN}/N_{\rm BH,max}$ and heat the $N_{\rm BH,max}$ closest neighbours using the updated value of $\Delta T_{\rm AGN}$. Additionally, if the number of gas particles within the kernel of the BH, $N_{\rm ngb}$, is smaller than $\min(N_{\rm AGN}, N_{\rm BH,max})$, then $\Delta T_{\rm AGN}$ is raised by $\min(N_{\rm AGN}, N_{\rm BH,max})/N_{\rm ngb}$ and all $N_{\rm ngb}$ particles receive the energy corresponding to the updated $\Delta T_{\rm AGN}$.

3 EMULATORS

We use Gaussian process emulators to determine the optimal values of the subgrid parameters for SN and AGN feedback in the COLIBRE model at m7 resolution and to demonstrate that simplified prescriptions for SN feedback with reduced numbers of free parameters cannot provide an equally good fit to the target observational data. We construct Gaussian process emulators using the python package SWIFT-EMULATOR (R. Kugel & J. Borrow 2022) and follow the method[10] from R. Kugel et al. (2023), who employed Gaussian process emulators to calibrate the large (up to 2.8^3 cGpc^3) cosmological simulations FLAMINGO (J. Schaye et al. 2023), but which have lower resolution and use a simpler galaxy formation model than COLIBRE.

We set up ≈ 200 simulations that sample the part of the COLIBRE parameter space of interest here at unique points, utilizing the Latin hypercube sampling technique (see Section 4.3). These simulations are used 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.

We apply this emulation method exclusively to the m7 resolution, where we run all simulations for the emulator training set in a $(50\,{\rm cMpc})^3$ volume. Emulators are not used for calibrating the m6 and m5 COLIBRE models, as running simulations at m6 and m5 resolutions in the same cosmological volume would be prohibitively computationally expensive. Reducing the cosmological volume to compensate for the increased computational cost is not a viable solution either, as the absence of relatively massive galaxies ($M_{\rm halo} \gtrsim 10^{13}\,{\rm M}_\odot$) in smaller volumes would prevent the emulators from properly calibrating the strength of AGN feedback. Instead, at m6 and m5 resolutions, we chose to calibrate the model manually, using the subgrid parameter values of the calibrated m7 model as an initial guess. The ability to start from the best-fitting parameter values determined at m7 resolution – combined with the relatively good convergence of the COLIBRE model with resolution and insights into the model's response to subgrid parameter variations gained through the emulators – makes manual calibration at m6 and m5 resolutions feasible.

We note that an alternative approach to calibrating higher-resolution models is to perform the emulator-based calibration at m7, m6, and m5 resolutions *simultaneously*, treating the gas particle mass as an additional emulator parameter. In this set-up, if the model exhibits reasonable convergence with resolution, the higher-resolution simulations can be run in progressively smaller volumes to reduce the computational cost of generating training data. The emulator then learns the properties of massive haloes from the lower-resolution simulations and, by using the (lower-mass) haloes present at all three resolutions, can potentially still infer how the properties of those massive haloes – absent from the higher-resolution training sets – vary with resolution. Compared to the manual approach adopted in this work, this alternative method has the advantage of incorporating resolution effects

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

directly into the emulator's dependence on a single parameter, $m_{\rm gas}$, making it straightforward to study resolution effects and enabling the simultaneous determination of best-fitting parameter values across all resolutions. Its main drawback (and the reason we did not adopt it) is that it requires significantly more prior knowledge of the model, as each iteration of narrowing the parameter space in search of the best-fitting model becomes substantially more computationally expensive due to the inclusion of higher-resolution simulations, even when these are run in smaller cosmological volumes.

3.1 Set-up

Consider a smooth mapping $y = f(x, \boldsymbol{\theta})$, where an output y depends on an input scalar variable x and a parameter vector $\boldsymbol{\theta}$.[11] Assuming that we know the true relation $y = f(x, \boldsymbol{\theta})$ only at a *finite* number of points N, denoted $\{x_n, \boldsymbol{\theta}_n, y_n = f(x_n, \boldsymbol{\theta}_n)\}_{n=1}^{N}$, our goal is to use this limited information to approximate $y = f(x, \boldsymbol{\theta})$ throughout the joint input space of x and $\boldsymbol{\theta}$. When f represents a complex, computationally expensive model (such as the COLIBRE galaxy formation model, where the data are generated by running numerical simulations), this approximation process is termed *emulation*.

We will write a hat above 'f' to distinguish an emulator, $y = \hat{f}(x, \boldsymbol{\theta})$, from the true relation, $y = f(x, \boldsymbol{\theta})$. 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. C. E. Rasmussen & C. K. I. Williams 2006). As in R. Kugel et al. (2023), we construct the covariance function using the squared exponential kernel,

$$k(\boldsymbol{X}, \boldsymbol{X}') = \exp\left[-\frac{(\boldsymbol{X} - \boldsymbol{X}')^{\rm T}\mathbf{C}^{-1}(\boldsymbol{X} - \boldsymbol{X}')}{2}\right], \tag{13}$$

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 opt for more sophisticated kernels because the relations we emulate vary smoothly with $\boldsymbol{X}$ across the entire parameter space. The entries of the matrix $\mathbf{C}$ are optimised during training of the Gaussian process emulators by maximizing the log marginal likelihood of the Gaussian process (see, e.g. C. E. Rasmussen & C. K. I. Williams 2006).

3.2 The emulated relations

We emulate two relations that are used to calibrate the COLIBRE SN and AGN feedback:

(i) GSMF at $z = 0$. Here the input variable x is the galaxy stellar mass, M_*, 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, we perform the emulation in log space, adopting $x = \log_{10} M_*$ as opposed to $x = M_*$. Likewise, for the output y, we take $y \equiv f(x) = \log_{10}(\mathrm{d}n/\mathrm{d}\log_{10} M_*)$, as opposed to $\mathrm{d}n/\mathrm{d}\log_{10} M_*$.

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

For each subhalo, the stellar mass M_* is computed as the sum of the masses of gravitationally bound stellar particles within a 3D spherical aperture of 50 pkpc, centred on the position of the most bound particle (of any type). By conducting mock observations of galaxies from the EAGLE simulations, A. 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 frequently used in 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 SSMR, we then compute the median projected galaxy half-mass radius 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. Projected stellar half-mass radii, like M_*, are calculated in 50 pkpc 3D apertures, considering only stellar particles that are gravitationally bound to the subhalo.

Before binning the simulated data, we shift all stellar masses by ΔM_* where ΔM_* is drawn from a lognormal distribution with zero mean and a standard deviation of 0.1 dex. This adjustment accounts for the A. S. Eddington (1913) bias, which affects observational data and thus must also be applied in the simulation. The choice of 0.1 dex reflects a typical uncertainty in stellar mass measurements (e.g. A. S. G. Robotham et al. 2020). Throughout this work, we apply this correction not only to the GSMF and SSMR, but also to all other simulation-predicted relations where the independent variable is galaxy stellar mass.

The uncertainties in the y values of the simulated data are accounted for as follows. For the SSMR, the error Δy is defined as half the difference between the 84$^{\rm th}$ and 16$^{\rm th}$ percentiles of the distribution of y values within each stellar mass bin. For the GSMF, the error Δy is given by the Poisson uncertainty. We use the square of Δy to define the diagonal entries of a noise covariance matrix (which is otherwise zero). This noise matrix is added to the Gaussian process covariance matrix, constructed using the kernel function (equation 13) and the training data, to form the total covariance for the Gaussian process. Including this noise term prevents overfitting to the training data, yielding smoother (i.e. less oscillatory) emulated relations. To optimise the smoothness, we introduce a hyperparameter that rescales the noise matrix. The value of this hyperparameter is set during training of the Gaussian process emulators, alongside the optimisation of the diagonal elements of the matrix $\mathbf{C}$ from equation (13).

Additionally, we emulate *the stellar-to-halo mass relation* (SHMR) for central subhaloes. This relation is not used to calibrate the COLIBRE model due to its weak observational constraints, but serves as a diagnostic tool. Our halo mass, $M_{\rm halo}$, follows the spherical overdensity definition of G. L. Bryan & M. L. Norman (1998). The emulator input consists of $x = \log_{10} M_{\rm halo}$ and $y = \log_{10}(M_*/M_{\rm halo})$. The y values represent the median stellar-to-halo mass ratios computed in halo mass bins of 0.2 dex width, with uncertainties given by half the difference between the 84$^{\rm th}$ and 16$^{\rm th}$ percentiles of the distribution of individual subhalo values in each mass bin.

3.3 Target observational data

We now describe the observational data to which the GSMF and SSMR emulators from Section 3.2 will be fit. The fitting will be

[11] As we will show later in Section 4.2, these parameters will be combinations of various subgrid parameters of the COLIBRE SN and AGN feedback.

performed over the galaxy stellar mass range $10^9 < M_*/\mathrm{M}_\odot < 10^{11.3}$. The lower bound is set by resolution constraints: at m7 resolution, $M_* = 10^9\,\mathrm{M}_\odot$ corresponds to only ~ 100 stellar particles. The upper bound is set by the relatively small number of galaxies with $M_* > 10^{11.3}\,\mathrm{M}_\odot$ in the $(50\,\mathrm{cMpc})^3$ volume simulations that are used to construct the emulator training data.

3.3.1 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 S. P. Driver et al. (2022) (their table 6, column 'all', including the 0.0807 dex correction accounting for the re-normalization to the Sloan Digital Sky Survey (SDSS) and evolution to precisely $z = 0$). The S. P. Driver et al. (2022) GSMF was derived from the Galaxy And Mass Assembly, Data Release 4 (GAMA DR4) survey, which provides multiwavelength observations for over 200,000 galaxies with spectroscopically confirmed redshifts. The S. P. Driver et al. (2022) GSMF spans ~ 5 dex in stellar mass, extending down to $M_* \sim 10^7\,\mathrm{M}_\odot$, and is presented precisely at $z = 0$, as the authors account for redshift evolution between the median redshift of their sample ($z \approx 0.079$) and $z = 0$. The stellar masses were computed assuming a G. Chabrier (2003) IMF.

3.3.2 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, R. A. 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 J. A. Hardwick et al. (2022) (their table C1, column 'Half-mass radius') as a secondary constraint on our simulations. The J. A. Hardwick et al. (2022) data come from the eXtended GALEX Arecibo SDSS Survey (xGASS; B. Catinella et al. 2018), which contains ~ 1200 galaxies selected from the SDSS DR7 catalogue. The sample has a flat stellar mass distribution spanning $M_* \approx 10^9$ to $10^{11.5}\,\mathrm{M}_\odot$ and redshifts in the range $0.01 < z < 0.05$. J. A. Hardwick et al. (2022) provide the median galaxy half-mass radii in stellar-mass bins of 0.2 dex width. The half-mass sizes were estimated by applying the S. Zibetti, S. Charlot & H.-W. Rix (2009) mass-to-light conversion to the stellar light profiles, which themselves are Sérsic fits performed by R. H. W. Cook et al. (2019). As shown by J. A. Hardwick et al. (2022), the half-mass sizes are smaller by ≈ 0.1 dex than the corresponding half-light sizes in the r-band, which is in agreement with A. de Graaff et al. (2022), who found a similar difference between mass- and luminosity-weighted galaxy sizes in the EAGLE simulations, using the SDSS pipeline for the analysis.

3.4 The search for the best-fitting parameter values

Each emulated relation $y = \hat{f}(x, \boldsymbol{\theta})$ changes 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 predictions and observational data. To quantify how well Gaussian process emulators can fit the observational data, we use Bayesian analysis.

3.4.1 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 is otherwise zero,

$$\mathcal{P}_{i,\mathrm{prior}}(\theta_i) = \begin{cases} 1, & \text{if } \theta_{i,\mathrm{min}} \leq \theta_i \leq \theta_{i,\mathrm{max}} \\ 0, & \text{otherwise}\,. \end{cases} \tag{14}$$

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

We opt for such a prior because of our limited knowledge about the parameters $\boldsymbol{\theta}$. The vector $\boldsymbol{\theta}$ contains the subgrid parameters of the SN and AGN feedback model. Probing N random realizations of $\boldsymbol{\theta}$ requires running N independent simulations, each of which may take a long time to complete. Therefore, given N simulations that we can afford to run, our training data contain N unique values for each subgrid parameter θ_i. The values of θ_i are distributed within a certain interval, whose lower and upper bounds define, respectively, $\theta_{i,\mathrm{min}}$ and $\theta_{i,\mathrm{max}}$ in equation (14). We set the prior to zero outside the domain sampled by the simulations because the errors of a Gaussian process emulator become large when it is used for extrapolation.

3.4.2 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 SSMR,

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \ln \mathcal{L}_{\mathrm{GSMF}}(\boldsymbol{\theta}) + \ln \mathcal{L}_{\mathrm{SSMR}}(\boldsymbol{\theta})\,, \tag{15}$$

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

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

$$\ln \mathcal{L}_R(\boldsymbol{\theta}) = -\frac{\langle N_{\mathrm{obs}} \rangle}{N_{R,\mathrm{obs}}} \frac{1}{2} \sum_{n=1}^{N_{R,\mathrm{obs}}} \left[\frac{\hat{f}_R(x_{R,n}, \boldsymbol{\theta}) - y_{R,n}}{\sqrt{\sigma_{R,n}^2 + \varepsilon_{R,\mathrm{emu}}^2}} \right]^2 , \tag{16}$$

where the subscript R is a placeholder for GSMF or SSMR. Next, $x_{R,n}$, $y_{R,n}$, and $\sigma_{R,n}$ 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 Section 3.3), and $N_{R,\mathrm{obs}}$ is the number of observational data points over which the sum is computed. $\hat{f}_R(x_{R,n}, \boldsymbol{\theta})$ is the prediction of the emulator of the relation R evaluated at $x_{R,n}$ and for the parameter vector $\boldsymbol{\theta}$. The SSMR and GSMF likelihood functions are normalized by $N_{R,\mathrm{obs}}/\langle N_{\mathrm{obs}} \rangle$ where $\langle N_{\mathrm{obs}} \rangle = (N_{\mathrm{GSMF,obs}} + N_{\mathrm{SSMR,obs}})/2$ is the average number of data points contained in the observational data for the GSMF and SSMR emulators. This normalization ensures that differences in the number of data points between the GSMF and SSMR datasets do not affect their relative contributions to the total likelihood. Note that since the emulators are constructed using simulation data in log-log space (see Section 3.2), the observational data used in equation (16) – including both the x and y values, as well as the errors on the y values – are also logarithmic.

Finally, $\varepsilon_{R,\mathrm{emu}}$ is the uncertainty in the emulator predictions for the relation R. To estimate $\varepsilon_{R,\mathrm{emu}}$, we train the emulators on all but one simulation from the training data of a given model (see Table 1 and Section 4.1), and ask the emulator to predict the GSMF and SSMR 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 simulation data. For both GSMF and SSMR, this gives us N_{runs} vectors with emulator errors where entries of each vector correspond to different stellar mass bins and the number N_{runs} is the total number of simulations in the training data (see Section 4.3). We concatenate all N_{runs} vectors into a single list and compute $\varepsilon_{R,\mathrm{emu}}$ as the standard deviation of the entries in this list. The value of $\varepsilon_{R,\mathrm{emu}}$ slightly changes depending on the relation and the model for which the emulator is constructed but averages to ≈ 0.07 dex.[12]

3.4.3 *Posterior*

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

$$\ln \mathcal{P}_{\mathrm{posterior}}(\boldsymbol{\theta}) = \ln \mathcal{L}(\boldsymbol{\theta}) + \ln \mathcal{P}_{\mathrm{prior}}(\boldsymbol{\theta}) \,, \tag{17}$$

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

$$\ln \mathcal{P}_{\mathrm{posterior}}(\boldsymbol{\theta}_{\mathrm{best}}) = \max \left(\ln \mathcal{P}_{\mathrm{posterior}}(\boldsymbol{\theta}) \right) \,. \tag{18}$$

To find the maximum of the posterior distribution, $\ln \mathcal{P}_{\mathrm{posterior}}(\boldsymbol{\theta})$, we use the Markov chain Monte Carlo (MCMC) python package EMCEE (D. Foreman-Mackey et al. 2013). We run MCMC for 5,000 steps using 30 independent walkers, which is more than sufficient for convergence, as indicated by the MCMC trace plot (not shown here). The walkers are initialized at random positions within the region of parameter space where the prior is non-zero. In the analysis of the posterior distribution, we remove the first 200 steps for each walker to avoid the 'burn-in' phase. To generate proposal steps for the random walk through the parameter space, we employ the 'stretch move' algorithm (J. Goodman & J. 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 (18).

4 CALIBRATION WITH EMULATORS

This section describes the calibration strategy of the COLIBRE model with fixed $\Delta T_{\mathrm{AGN}} = 10^9$ K at m7 resolution, 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.

[12]Because the emulator error, $\varepsilon_{R,\mathrm{emu}}$, varies slightly between different emulated relations R, it affects the relative contribution of each relation to the total likelihood. To estimate the impact of this, we tested alternative choices for $\varepsilon_{R,\mathrm{emu}}$ in equation (16), such as using an average error over the two emulated relations – GSMF and SSMR – or an error further averaged over all models used in the emulation. We did not find any advantage of using these alternatives, nor any significant impact on the best-fitting parameter values of the final model.

Table 1. Latin hypercubes used to train the emulators. Column (1): the name of the model for which the Latin hypercube is created (see Section 4.1); column (2): the hypercube level in the hierarchy (level 2 is a subregion of level 1 with finer sampling; the `Basic` model only has level 1); column (3): the number of simulations included in the Latin hypercube at a given level; column (4): the energy per single CC SN in units of 10^{51} erg; column (5): the fraction of SN energy that is injected in kinetic form; column (6): the pivot density in the relation between the SN heating temperature and the gas density (equation 8); column (7) the pivot birth pressure in the relation between the energy in CC SN feedback and the stellar birth gas pressure (equation 2); column (8) the BH seed mass. The two numbers in each cell of columns 4-8 specify the interval over which each parameter is varied in the Latin hypercube. For a given model, a cell left blank indicates that the model does not include the corresponding parameter.

Model name	Hypercube level	$N_{\mathrm{runs}}^{\mathrm{L1,2}}$	Model parameters that are varied in the Latin hypercube				
			f_{E}	f_{kin}	$n_{\mathrm{H,pivot}}$ (cm^{-3})	$\log_{10} P_{\mathrm{E,pivot}}/k_{\mathrm{B}}$ ($\mathrm{K\ cm}^{-3}$)	$\log_{10} m_{\mathrm{BH,seed}}$ ($\mathrm{M}_\odot$)
`Basic`	1	24	[0.1, 5]	–	–	–	[3, 6]
	–	–	–	–	–	–	–
`ThermalKinetic`	1	32	[0.3, 2.3]	[0, 1]	–	–	[3.5, 6]
	2	40	[0.3, 2.3]	[0, 0.5]	–	–	[4.2, 5.5]
`ThermalKinetic_var`ΔT_{SN}	1	40	[0.3, 2.3]	[0, 0.5]	[0.05, 2.5]	–	[4.2, 5.5]
	2	8	[1.25, 1.55]	[0, 0.14]	[0.4, 0.65]	–	[4.4, 4.6]
`ThermalKinetic_var`ΔT_{SN}`var`f_{E}	1	40	–	[0, 0.5]	[0.05, 2]	[3.3, 4.5]	[4.2, 5.5]
	2	8	–	[0.07, 0.14]	[0.4, 0.65]	[3.8, 4.0]	[4.7, 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 prescription presented in Section 2.2.3. 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 2.

We do not commence with calibrating the fiducial COLIBRE 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 is necessary.

Besides the `Basic` model, we consider two other simplified prescriptions for SN feedback: `ThermalKinetic` and `ThermalKinetic_var`$\Delta T_{\rm SN}$, each of which is detailed below. We emphasize that the only difference between the COLIBRE model with fixed $\Delta T_{\rm AGN}$ and its simplified versions – `Basic`, `ThermalKinetic`, and `ThermalKinetic_var`$\Delta T_{\rm SN}$ – is the treatment of SN feedback, while all other parts of the galaxy formation physics remain identical, including the heating temperature $\Delta T_{\rm AGN} = 10^9$ K in AGN feedback. Unlike SN feedback, we do not consider simplified prescriptions for AGN feedback because BH particles heating gas neighbours with a constant temperature and using fixed radiative and coupling efficiencies (see Section 2.2.9) already constitutes a basic AGN algorithm.

4.1.1 The basic model

Relative to the fiducial COLIBRE prescription for SN feedback from Section 2, we make the following simplifications in the `Basic` model:

(i) The energy of a single SN, in units of 10^{51} erg, $f_{\rm E}$, is constant rather than dependent on the stellar birth pressure, $P_{\rm birth}$ (see equation 2). The value of the constant $f_{\rm E}$ will be determined using emulators.

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

(iii) The heating temperature in the thermal (CC and type-Ia) SN feedback, $\Delta T_{\rm SN}$, is set to a constant value of $10^{7.5}$ K – the value used in the EAGLE simulations – rather than being density-dependent (equation 8).

4.1.2 The thermal-kinetic model

In addition to the `Basic` model, we consider a modification in which the prescription for SN feedback includes both kinetic and thermal channels of energy injection (i.e. $f_{\rm kin}$ is no longer necessarily 0). We refer to this model as `ThermalKinetic`. As is the case in the COLIBRE model with fixed $\Delta T_{\rm AGN}$, the kinetic channel of the `ThermalKinetic` model uses the desired kick velocity parameter of $\Delta v_{\rm kick} = 50\,{\rm km\,s^{-1}}$. 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 introduces one additional free parameter: the fraction of SN energy injected in kinetic form, $f_{\rm kin}$, which will be determined using emulators. We note that 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 E. Chaikin et al. (2023).

4.1.3 The thermal-kinetic model with a variable heating temperature

Our final simplified model is `ThermalKinetic_var`$\Delta T_{\rm SN}$. As the name suggests, compared to `ThermalKinetic`, `ThermalKinetic_var`$\Delta T_{\rm SN}$ adopts the density-dependent heating temperature for thermal SN feedback (for both CC and type-Ia SNe) detailed in Section 2.2.5, while `Basic` and `ThermalKinetic` use a constant value of $\Delta T_{\rm SN} = 10^{7.5}$ K.

Of the simplified models, `ThermalKinetic_var`$\Delta T_{\rm SN}$ is the closest to the COLIBRE model with fixed $\Delta T_{\rm AGN} = 10^9$ K. The only difference is that the COLIBRE fiducial prescription for SN feedback adopts an $f_{\rm E}$ that depends on the stellar birth gas pressure, following equation (2), whereas `ThermalKinetic_var`$\Delta T_{\rm SN}$ uses a constant $f_{\rm E}$.

4.2 Selection of subgrid parameters for emulator-based calibration

We next describe the selection of the subgrid parameters that will be calibrated using emulators. These parameters, which we denote by the vector $\boldsymbol{\theta}$, enter the emulators defined in Section 3.2 and are optimized with the methods of Bayesian statistics (Section 3.4), such that the simulation provides the best match to the observational data (Section 3.3).

We will use the emulators to optimize only the subgrid parameters that govern the strengths of SN and AGN feedback. Parameters related to other aspects of galaxy formation physics (such as star formation, chemical enrichment, or radiative cooling) will not be considered, either because they have little to no impact on the galaxy properties relevant to our calibration (i.e. GSMF and SSMR), or because their values are well constrained by fundamental physics or inferred from independent observations.

4.2.1 AGN feedback parameters

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

(i) 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 $\Delta T_{\rm AGN}$ tend to yield stronger (weaker) AGN feedback. However, owing to the ability of SMBHs to self-regulate (C. M. Booth & J. Schaye 2009, 2010), we expect the exact value of $\Delta T_{\rm AGN}$ to have only a minor impact on the $z = 0$ GSMF and SSMR (e.g. I. G. McCarthy et al. 2017), provided that the temperature of the heated gas remains sufficiently high for the injected thermal energy not to be rapidly radiated away. We will therefore not consider $\Delta T_{\rm AGN}$ as one of the subgrid parameters for calibration and instead keep it fixed at 10^9 K. In Section 5.5 we will confirm that (modest) variations in $\Delta T_{\rm AGN}$ indeed have only a minor effect on the $z = 0$ GSMF and SSMR, and that these small differences can be compensated for by adjusting other model parameters.

(ii) The BH seed mass, $m_{\rm BH,seed}$, determines how quickly BHs can grow over time (see equation 10). Higher $m_{\rm BH,seed}$ will cause

faster BH growth, leading to more energetic AGN feedback in lower-mass galaxies and at higher redshifts (e.g. C. M. Booth & J. Schaye 2009). Because both the GSMF and SSMR 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, $\boldsymbol{\theta}$.

(iii) The minimum halo FoF mass in which BHs 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 early BHs can start growing in mass (e.g. C. M. Booth & J. 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 BHs, but a similar effect can be achieved by decreasing $m_{\rm BH,seed}$. Owing to this degeneracy, which will be shown in Section 5.5, we will not include $M_{\rm FoF,seed}$ in our set of parameters for optimization. As already explained in Section 2.2.8, at m7 resolution we set $M_{\rm FoF,seed} = 5 \times 10^{10}\,{\rm M}_\odot$.

4.2.2 *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.

(i) 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 `Basic` model, including the AGN feedback parameters from Section 4.2.1, is given by $\boldsymbol{\theta} = (m_{\rm BH,seed}, f_{\rm E})$.

(ii) 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: $\boldsymbol{\theta} = (m_{\rm BH,seed}, f_{\rm E}, f_{\rm kin})$.

(iii) The `ThermalKinetic_var`Δ`T`$_{\rm SN}$ model employs a density-dependent heating temperature for SN feedback (equation 8), which is described by four parameters: the pivot gas density $n_{\rm H,pivot}$, the heating temperature at the pivot density $\Delta T_{\rm SN,pivot}$, and the minimum and maximum heating temperatures, $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$. As documented in Section 2.2.5, at m7 resolution we set $\Delta T_{\rm SN,pivot}$, $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$ to $10^{6.5}$ K, $10^{6.5}$ K, and $10^{7.5}$ K, respectively, which leaves us with only one free parameter: $n_{\rm H,pivot}$. Therefore, the final form of the parameter vector $\boldsymbol{\theta}$ for the `ThermalKinetic_var`Δ`T`$_{\rm SN}$ model is $\boldsymbol{\theta} = (m_{\rm BH,seed}, f_{\rm E}, f_{\rm kin}, n_{\rm H,pivot})$.

(iv) Finally, the COLIBRE model with fixed $\Delta T_{\rm AGN} = 10^9$ K (henceforth, the `ThermalKinetic_var`Δ`T`$_{\rm SN}$`varf`$_{\rm E}$ model) uses a stellar birth pressure dependent energy for CC SN feedback (equation 2). This comes with another set of four parameters, $f_{\rm E,min}$, $f_{\rm E,max}$, $\sigma_{\rm P}$, and $P_{\rm E,pivot}$, which together replace the parameter $f_{\rm E}$ that specifies the constant energy in CC SN feedback in the three simplified models. As explained in Section 2.2.3, the values of three out of the four extra parameters are fixed: $f_{\rm E,min} = 0.1$, $f_{\rm E,max} = 4$, and $\sigma_{\rm P} = 0.3$. Thus, in the emulation, the dependence of $f_{\rm E}$ on the stellar birth pressure will be described with a single free parameter, $P_{\rm E,pivot}$, resulting in the parameter vector for the `ThermalKinetic_var`Δ`T`$_{\rm SN}$`varf`$_{\rm E}$ model, $\boldsymbol{\theta} = (m_{\rm BH,seed}, P_{\rm E,pivot}, f_{\rm kin}, n_{\rm H,pivot})$.

We note that due to the functional degeneracies between $P_{\rm E,pivot}$, $f_{\rm E,min}$, and $f_{\rm E,max}$ (see equation 2), a very low (high) $P_{\rm E,pivot}$ preferred by the emulator will suggest that our chosen value of $f_{\rm E,max}$ ($f_{\rm E,min}$) may be too low (high). In Section 5.5, we will show that $P_{\rm E,pivot}$, $f_{\rm E,min}$, and $f_{\rm E,max}$ are indeed degenerate with one another (and also with $\sigma_{\rm P}$).

4.3 Training data for emulators

For the `ThermalKinetic_var`Δ`T`$_{\rm SN}$`varf`$_{\rm E}$ model described in Section 2, as well as its three simplified counterparts introduced in Section 4.1, we construct emulators of both the GSMF and the SSMR as defined in Section 3.2, resulting in a total of eight independent emulators. Additionally, for diagnostic purposes, we construct emulators of the SHMR for the `Basic` and `ThermalKinetic` models, yielding two more independent emulators. Each emulator must be trained before it can be used for parameter inference or diagnostics.

To build the training datasets, we run a set of simulations. For a given model, each simulation represents a unique combination of values of the subgrid parameters $\boldsymbol{\theta}$ (i.e. it is a unique sampling point in the parameter space). To evenly sample the parameter space given our uniform priors, we make use of the *Latin hypercube* sampling technique (M. D. McKay, R. J. Beckman & W. J. Conover 1979). The main advantage of Latin hypercube sampling over random sampling is that it requires significantly fewer sampling points to achieve the desired accuracy in emulator predictions. Given a target of N sampling points in an $N_{\rm param}$-dimensional parameter space, this is achieved by first creating a grid that divides the space into $N^{N_{\rm param}}$ equal-volume cells. Sampling is then performed by placing N points into N cells such that, when projected onto any single parameter dimension, there is exactly one point in each of the N equally spaced intervals, with each point positioned randomly within its assigned cell. As a result, Latin hypercube sampling covers the parameter space more evenly, allowing the number of simulations needed to train the emulators for each model to remain relatively modest.

To further enhance emulator accuracy, we construct the Latin hypercubes hierarchically at two levels: a coarse level 1 with broad parameter variations and level 2, a refined subregion within level 1 with finer sampling. The boundaries of level 2 are determined by first training the emulator using only the simulations from level 1 and identifying the subregion of the parameter space that most likely contains the best-fitting model. At level 1, we run $N^{\rm L1}_{\rm runs} = 24$ simulations for the `Basic` model, 32 for the `ThermalKinetic` model, and 40 for `ThermalKinetic_var`Δ`T`$_{\rm SN}$ and `ThermalKinetic_var`Δ`T`$_{\rm SN}$`varf`$_{\rm E}$. 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. At level 2, we use $N^{\rm L2}_{\rm runs} = 40$ simulations for the `ThermalKinetic` model, 8 for `ThermalKinetic_var`Δ`T`$_{\rm SN}$, and 8 for `ThermalKinetic_var`Δ`T`$_{\rm SN}$`varf`$_{\rm E}$. We do not use level 2 for the `Basic` model because its simplicity allows the emulator to accurately determine the best-fitting parameter values using only level 1. Level 2 of the `ThermalKinetic` model contains significantly more simulations than those of `ThermalKinetic_var`Δ`T`$_{\rm SN}$ and `ThermalKinetic_var`Δ`T`$_{\rm SN}$`varf`$_{\rm E}$, as we fit `ThermalKinetic` not only to the observed GSMF and SSMR together but also separately to each observable. This results in *three* best-fitting models located in different regions of the parameter space (see Section 5.1.4). Furthermore, since the `Basic` model is equivalent to the `ThermalKinetic` model with $f_{\rm kin} = 0$, we incorporate 24 simulations from the `Basic` model's hypercube into that of the `ThermalKinetic` model. This improves the accuracy of the emulator predictions for `ThermalKinetic` near the hypercube boundary where $f_{\rm kin} = 0$.

In total, the Latin hypercubes for `Basic`, `ThermalKinetic`, `ThermalKinetic_var`Δ`T`$_{\rm SN}$, and

ThermalKinetic_varΔT_{SN}varf_E contain $N_{runs}^{total} = 24$, 96, 48, and 48 simulations, respectively. Table 1 summarizes the properties of these Latin hypercubes, including the parameter ranges $\boldsymbol{\theta}$ explored at each level. The level 1 ranges were preselected to ensure that the peaks of the posterior distributions of $\boldsymbol{\theta}$ fall within the hypercube domain.[13] In the training set, no distinction is made between the simulations from level 1 and level 2; the emulator is trained using simulations from both levels simultaneously.

All simulations in the Latin hypercubes were run to $z = 0$ in $(50\ \mathrm{cMpc})^3$ volumes at m7 resolution, with initial numbers of gas and DM particles of 376^3 and 4×376^3, respectively (see Section 2.1 for more details on the ICs). Each simulation was run on 128 cores and took on average ≈ 5 days[14] to reach $z = 0$. For the entire set of simulations, this translates to $\approx 3 \times 10^6$ 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. In Appendix A, we demonstrate that a $(50\ \mathrm{cMpc})^3$ volume is sufficient to produce a GSMF and SSMR within the stellar mass range considered in the calibration, $10^9 < M_*/\mathrm{M}_\odot < 10^{11.3}$, that are consistent with those obtained from simulations in larger volumes.

As an illustration, Fig. 1 shows the Latin hypercube for the ThermalKinetic_varΔT_{SN}varf_E model. The axes correspond to different hypercube parameters: $m_{BH,seed}$, f_{kin}, $n_{H,pivot}$, and $P_{E,pivot}$. Grey and black hatched rectangles denote levels 1 and 2 of the hypercube's domain, while light-blue triangles and circles represent 40 and 8 individual simulations, respectively, sampling the parameter space at these levels. These 48 simulations are used to train the emulators to predict the $z = 0$ GSMF and SSMR as functions of the hypercube's parameters. Once the GSMF and SSMR emulators have been trained, we can fit the ThermalKinetic_varΔT_{SN}varf_E model to the observed GSMF and SSMR, as described in Section 3.4, to find the best-fitting values of the model parameters. The Latin hypercubes for the other three models – Basic, ThermalKinetic, and ThermalKinetic_varΔT_{SN} – look qualitatively similar, except that the parameter spaces of the first two models have fewer dimensions.

4.4 Best-fitting parameter values and simulations with best-fitting models

The (rounded) best-fitting values of the parameters of the ThermalKinetic_varΔT_{SN}varf_E model, as well as its three simplified analogues, are presented in Table 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 (rounded) 1σ errors on the parameter values. Finally, we show the values of the reduced χ^2 (i.e. χ^2_ν), which quantify the goodness of fit to the observational data. We compute χ^2 as the sum of the squared differences between the predictions of the best-fitting model and the observational data to which the model was calibrated, normalized by the combined uncertainties from the observations and the emulator (see Section 3.4.2). However, since the emulator uncertainty, $\varepsilon_{R,emu}$, varies slightly between models and emulated relations, we adopt a fixed value of $\varepsilon_{R,emu} = 0.07$ – the average across the four models and the two emulated relations (GSMF and SSMR) – for computing χ^2. This normalization places all best-fitting models on equal footing, enabling a direct comparison of their χ^2_ν values.

We round the best-fitting values of the model parameters to one digit to the right of the decimal point, as we found no significant improvement in the accuracy of the fits when using more precise values. The direction of rounding – up or down – is determined not only by the proximity of the best-fitting value to its rounded counterpart, but also to ensure that the model with the rounded values of the subgrid parameters remains within the 68 per cent credibility interval of the posterior (i.e. within 1σ from the peak if the posterior has a Gaussian form).[15]

The ThermalKinetic model appears three times in Table 2, as we fit it to the observed GSMF and SSMR separately and jointly (see Section 5.1.4). For each model and each set of best-fitting parameter values (different rows in Table 2), we run a separate numerical simulation in a $(50\ \mathrm{cMpc})^3$ volume, resulting in a total of six new simulations. In these simulations, we use the rounded best-fitting parameter values as reported in Table 2, except that, for convenience, the rounded value of $P_{E,pivot}/k_B$ is specified in scientific notation as 8×10^3 K cm^{-3} ($\approx 10^{3.903}$ K cm^{-3}), rather than as $10^{3.9}$ K cm^{-3}.

5 RESULTS

We begin this section by evaluating the performance of the two simplest best-fitting models to the $z = 0$ observed GSMF and SSMR, Basic and ThermalKinetic, focusing on their predictions for the GSMF, SSMR, and SHMR, as well as assessing the accuracy of the emulators relative to the simulations (Section 5.1). We then examine the $z = 0$ GSMF and SSMR in the more advanced best-fitting models, ThermalKinetic_varΔT_{SN} and ThermalKinetic_varΔT_{SN}varf_E (Section 5.2). Next, we compare the simulations using the four best-fitting models to observations of various galaxy properties not considered during the emulation-based calibration (Section 5.3).

After showing that the best-fitting ThermalKinetic_varΔT_{SN}varf_E model outperforms its three counterparts with more simplified SN feedback, we apply it to the higher COLIBRE resolutions, m6 and m5 (Section 5.4). We discuss that, at m5 resolution, ThermalKinetic_varΔT_{SN}varf_E requires an undesirably low $m_{BH,seed}$. We show that to allow for a higher $m_{BH,seed}$ while maintaining a good fit to the observed GSMF and SSMR, the

[13] These ranges were determined by running simulations in small cosmological volumes (25^3 cMpc^3) with much broader parameter variations.

[14] From additional tests, we found that decreasing the number of DM particles in the ICs from 4×376^3 to 376^3 (i.e. matching the initial number of gas particles) reduces the total wall-clock time to reach $z = 0$ by about 30 per cent.

[15] There are, however, two exceptions to this rule: (i) the best-fitting value of f_E in the Basic model is ≈ 0.94, but it is rounded to 1 instead of 0.9, as we found no improvement in using the more precise value; (ii) the best-fitting value of $n_{H,pivot}$ in the ThermalKinetic_varΔT_{SN}varf_E model is ≈ 0.53 cm^{-3}, but it is rounded to 0.6 cm^{-3} instead of 0.5 cm^{-3} because at the time we finalized the model parameters, we were using a slightly different (older) version of the emulator training data, which preferred 0.6 over 0.5 cm^{-3}. In practice, this difference has a negligible impact, as the ThermalKinetic_varΔT_{SN}varf_E models with both $n_{H,pivot} = 0.6$ and $0.5\,\mathrm{cm}^{-3}$ lie within the 68 per cent credibility interval of the posterior and yield similar χ^2_ν values.

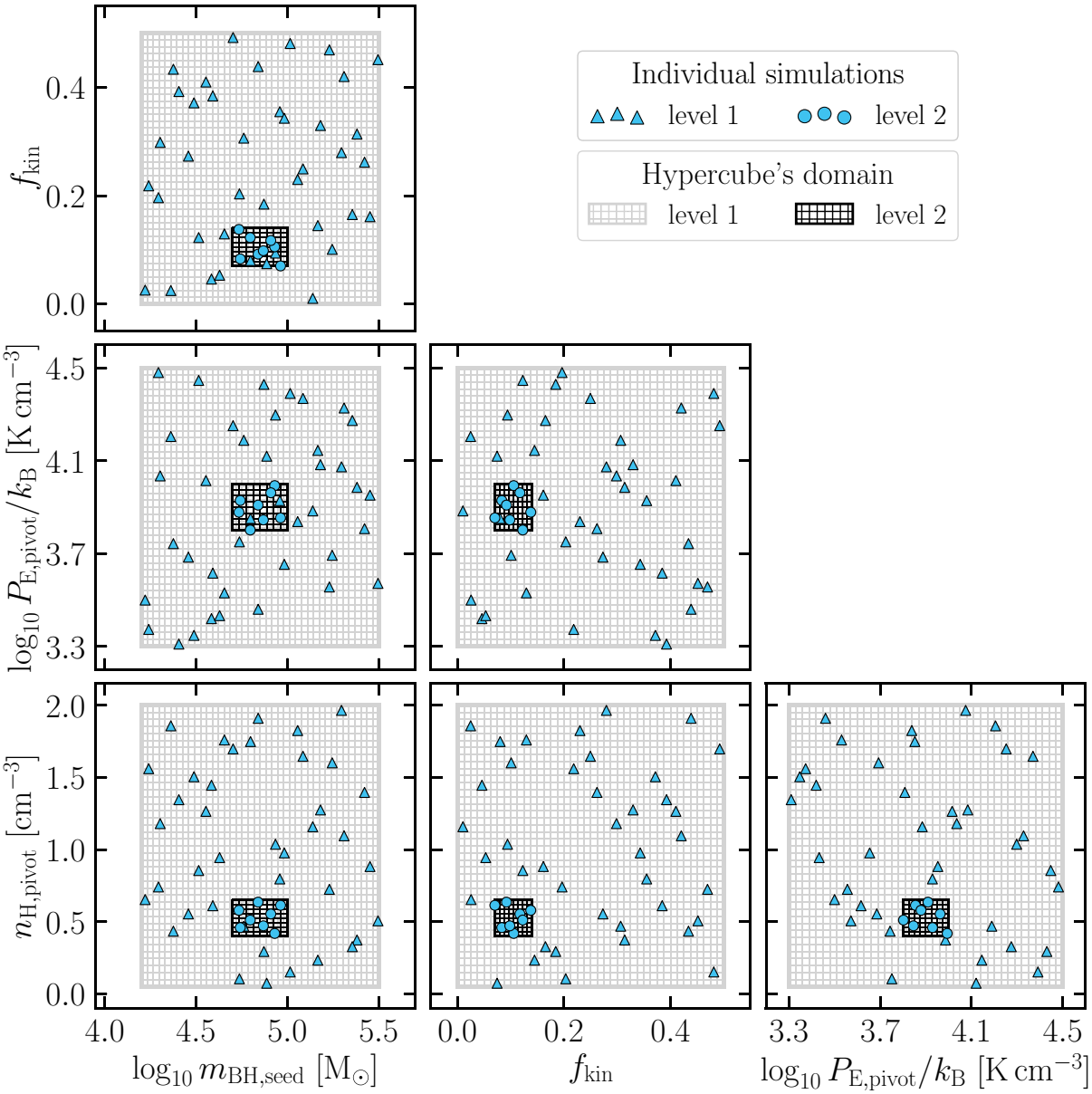

Figure 1. The Latin hypercube for the `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$_{\rm E}$ model. The axes of the panels correspond to different parameters of the model: $m_{\rm BH,seed}$, $f_{\rm kin}$, $n_{\rm H,pivot}$, and $P_{\rm E,pivot}$ (see Section 4.2 for details). The grey (black) hatched rectangle marks level 1 (level 2) of the hypercube's domain, while light-blue triangles (circles) indicate the sampling for level 1 (level 2), consisting of 40 (8) individual simulations. Together, the simulations from levels 1 and 2 form the training dataset for the `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$_{\rm E}$ model, which is used to train the emulators for the $z = 0$ GSMF and SSMR.

fixed $\Delta T_{\rm AGN} = 10^9$ K in `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$_{\rm E}$ needs to be replaced with a variable $\Delta T_{\rm AGN}$. We then compare the best-fitting `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$_{\rm E}$ model with its variable $\Delta T_{\rm AGN}$ modification, showing that both achieve similar level of agreement with the observed GSMF, SSMR, and other galaxy properties. Consequently, we establish the latter as the fiducial COLIBRE model and demonstrate that the COLIBRE simulations successfully reproduce the observed GSMF and SSMR not only at m7 resolution but also at m6 and m5. Finally, in Section 5.5, we explore how variations in individual subgrid parameters, including those not optimized by the emulators, impact the calibrated galaxy properties in the COLIBRE fiducial model at m7 resolution.

5.1 Calibration diagnostics for `Basic` and `ThermalKinetic` models

5.1.1 GSMF and SSMR with the best-fitting parameters

Fig. 2 shows the $z = 0$ GSMF and SSMR for the `Basic` (green) and `ThermalKinetic` (yellow) 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 S. P. Driver et al. (2022) and the observed SSMR from J. A. Hardwick et al. (2022). The solid curves are the GSMF and SSMR from the simulations that use the best-fitting parameter values found by the emulators (see Table 2). The shaded yellow region designates the scatter in the simulation with the

Table 2. The best-fitting values of the parameters identified by the emulator by matching the model to the $z = 0$ observational data: the GSMF from S. P. Driver et al. (2022) and SSMR from J. A. Hardwick et al. (2022). Column (1): the name of the model; column (2): the observational data to which the model was fit; column (3): the χ^2_ν value of the best-fitting model. 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 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	χ^2_ν	Best-fitting values of model parameters				
			f_E	f_kin	$n_\mathrm{H,pivot}$ [cm^{-3}]	$\log_{10} P_\mathrm{E,pivot}/k_\mathrm{B}$ [K cm^{-3}]	$\log_{10} m_\mathrm{BH,seed}$ [M$_\odot$]
`Basic`	GSMF and SSMR	9.3	$1.0^{+0.05}_{-0.1}$	–	–	–	4.0 ± 0.1
`ThermalKinetic`	GSMF and SSMR	4.7	1.0 ± 0.1	0.3 ± 0.05	–	–	$4.7^{+0.1}_{-0.05}$
`ThermalKinetic_varΔT`$_\mathrm{SN}$	GSMF and SSMR	2.9	1.3 ± 0.1	0.1 ± 0.05	$0.5^{+0.1}_{-0.2}$	–	$4.6^{+0.05}_{-0.1}$
`ThermalKinetic_varΔT`$_\mathrm{SN}$`varf`$_\mathrm{E}$	GSMF and SSMR	0.8	–	0.1 ± 0.05	$0.6^{+0.1}_{-0.2}$	3.9 ± 0.1	$4.8^{+0.05}_{-0.1}$
`ThermalKinetic`	GSMF	0.3	1.3 ± 0.1	0.6 ± 0.1	–	–	4.8 ± 0.2
`ThermalKinetic`	SSMR	0.9	$1.0^{+0.1}_{-0.05}$	$0^{+0.05}_{-0}$	–	–	4.3 ± 0.1

`ThermalKinetic` model: the Poisson uncertainty for the GSMF and the 16$^\mathrm{th}$ to 84$^\mathrm{th}$ percentile scatter for the SSMR. 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 per bin drops below 5 (roughly corresponding to $M_* \gtrsim 10^{11.5}\,\mathrm{M}_\odot$). The vertical solid lines indicate the edges of the stellar mass interval used in the training of the emulators: $10^{8.8} < M_*/\mathrm{M}_\odot < 10^{11.35}$. The vertical dash–dotted lines show the mass range within which the trained emulators were fit to the observational data: $10^9 < M_*/\mathrm{M}_\odot < 10^{11.3}$. The observed GSMF from S. P. Driver et al. (2022) is shown in the left panel as black squares, and the observed SSMR from J. A. Hardwick et al. (2022) is shown in the right panel as black circles. The error bars in the observed GSMF indicate the Poisson uncertainty, while in the observed SSMR, 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 SSMR from J. A. 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 SSMR predicted by the emulators and resulting from the 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 SSMR and GSMF range between 0 and ≈ 0.1 dex, which is comparable to the intrinsic scatter in simulations such as ours due to their stochastic nature (e.g. J. Borrow et al. 2023).

By comparing the solid curves to the black squares in the left panel and the black 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 $M_* \gtrsim 10^{10}\,\mathrm{M}_\odot$ 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 a single- or double-component P. Schechter (1976) function, featuring an exponential down-turn at high stellar mass. Although the `ThermalKinetic` model matches the observed GSMF better than `Basic`, the discrepancy between its GSMF and the observed data is still significant. Moreover, both models perform poorly in matching the observed galaxy sizes: the SSMR 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 $\approx 0.1 - 0.2$ dex.

Overall, the combined fit to the observed GSMF and SSMR is better in the `ThermalKinetic` model than in `Basic`, but still not satisfactory, motivating a more sophisticated model.

5.1.2 *Posterior distributions of the model parameters*

Fig. 3 shows the posterior distributions of the parameters of the `Basic` (green) and `ThermalKinetic` (yellow) models resulting from fitting the emulators to the observed GSMF and SSMR, as explained in Section 3.4. The three contours of the same colour signify the 34, 68, and 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_kin, are plotted.

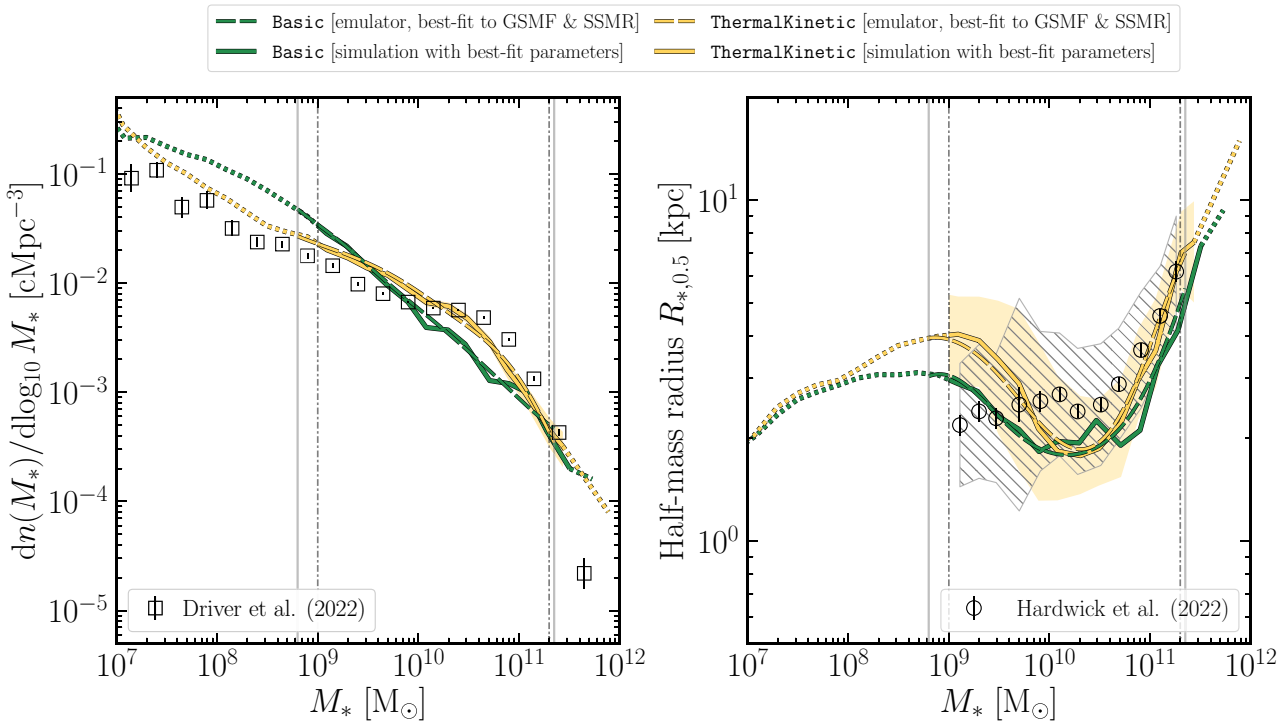

Figure 2. The GSMF (*left*) and the median SSMR (*right*) at $z = 0$, for the `Basic` (green) and `ThermalKinetic` (yellow) models fit to the observed GSMF and SSMR. The dashed and solid curves indicate, respectively, the best-fitting predictions of the emulator and the corresponding simulations with the best-fitting parameters from Table 2. The shaded yellow regions in the left and right panels indicate the Poisson uncertainty for the GSMF and the 16$^{\mathrm{th}}$ to 84$^{\mathrm{th}}$ percentile scatter for the SSMR in the simulation using the `ThermalKinetic` model, respectively. We convert the green and yellow solid curves into dotted curves where galaxies are poorly resolved ($M_* < 10^9$ M$_\odot$) and where the number of galaxies is strongly limited by the finite simulated volume (the number of objects per bin is less than 5). 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 S. P. Driver et al. (2022) and J. A. Hardwick et al. (2022) are shown as black squares and circles, respectively. Additionally, the grey hatched region in the right panel indicates the galaxy population-wide scatter in the SSMR from J. A. Hardwick et al. (2022). Although the `ThermalKinetic` model produces a combined fit to the observed GSMF and SSMR that is better than the fit with the `Basic` model, neither model is particularly satisfactory.

First, we observe that the regions of parameter space explored by the Latin hypercubes encompass the peaks of the posterior distribution in the two models, covering it by more than $\pm 2\sigma$ (the 95 per cent credibility levels). This is reassuring in that it implies our results are not driven by the boundaries of the chosen prior. Secondly, both models prefer a value of the dimensionless SN energy parameter f_{E} of order unity, implying that a single CC SN releases $\sim 10^{51}$ erg of energy, which is consistent with standard theoretical expectations. Third, the best-fitting `ThermalKinetic` model has a BH seed mass of $m_{\mathrm{BH,seed}} \approx 10^{4.7}$ M$_\odot$, whereas for the `Basic` model, $m_{\mathrm{BH,seed}}$ is nearly an order of magnitude lower, $m_{\mathrm{BH,seed}} \approx 10^{4.0}$ M$_\odot$. This is likely because the prescription for SN feedback in the `Basic` model is too simplistic, such that the model's only way to improve agreement with the observed GSMF at the massive end – without worsening the match at the low-mass end even more – is to reduce the strength of AGN feedback. The AGN feedback is suppressed by lowering $m_{\mathrm{BH,seed}}$, as there is no other AGN feedback-related parameter available for tuning in the `Basic` model. Fourth, the posterior of the `ThermalKinetic` model peaks at the kinetic energy fraction in SN feedback of $f_{\mathrm{kin}} \approx 0.3$, indicating the importance of SN kinetic feedback in bringing this model closer to the observational data.

Finally, in the legend next to the names of the models, we show the χ^2_ν value of their fits to the observational data, which is 9.3 for the `Basic` model and 4.7 for `ThermalKinetic`. These values are in line with our conclusions from Fig. 2: that the `ThermalKinetic` model outperforms the `Basic` model, but neither model is a good fit to the data.

5.1.3 *The effect of changing the model parameters*

Fig. 4 shows the $z = 0$ stellar-to-halo mass relation for central subhaloes in the `ThermalKinetic` model. As discussed in Section 3.2, we do not fit the emulator to the SHMR because it is not directly observed. Here we only use the SHMR to predict how varying the model parameters affects the galaxy stellar mass at fixed halo mass. Displaying the SHMR as opposed to the GSMF makes it easier to visually distinguish the impact of different subgrid parameters because, compared to the GSMF, the SHMR varies over a smaller dynamical range and includes a characteristic change in the sign of the slope of the relation.

In each panel of Fig. 4, differently coloured solid curves correspond to different SHMRs 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 BH 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 SHMR predicted by the emulator of the `Basic` model and how it changes with f_{E} (thin dotted curves in different colours). The only other

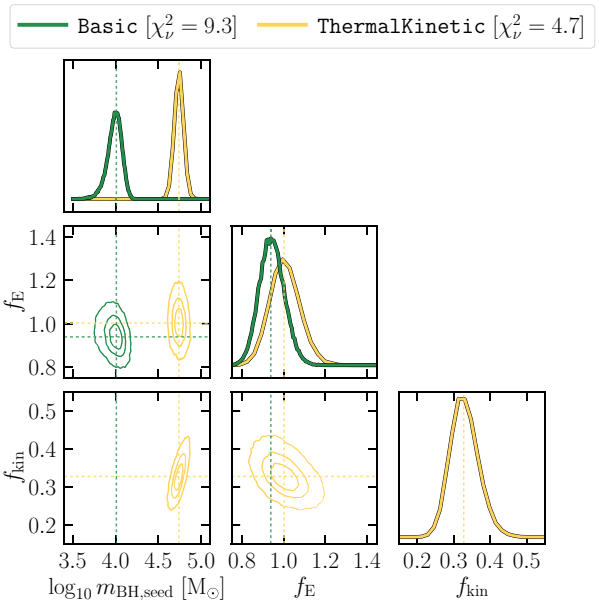

Figure 3. Posterior distribution of the parameters for the `Basic` (green) and `ThermalKinetic` (yellow) models fit to the observed $z = 0$ GSMF and SSMR. The χ^2_ν values of the fits are shown in the legend. The three contours of each colour indicate 34, 68, and 95 per cent credibility levels. The vertical and horizontal dotted lines indicate the best-fitting values of the model parameters, corresponding to the maximum of the posterior.

parameter of the `Basic` model, $m_{\rm BH,seed}$ is equal to its best-fitting value, $10^{4.0}\,{\rm M}_\odot$. The SHMR from the semi-empirical models of B. P. Moster et al. (2018) and P. Behroozi et al. (2019) are shown for reference only (black points).

First, by examining the left panel, we find that, as expected, $m_{\rm BH,seed}$ predominantly affects the stellar mass of massive haloes ($M_{\rm halo} \gtrsim 10^{11.5}\,{\rm M}_\odot$). In lower mass haloes ($M_{\rm halo} \lesssim 10^{11.5}\,{\rm M}_\odot$), BHs grow less efficiently, resulting in a lack of AGN feedback and, consequently, a much weaker dependence of the SHMR on $m_{\rm BH,seed}$, unless $m_{\rm BH,seed}$ is very high ($m_{\rm BH,seed} \gtrsim 10^{5.5}\,{\rm M}_\odot$). Furthermore, we observe that the value of the BH seed mass determines the halo mass at which the SHMR reaches its peak. The curve with $m_{\rm BH,seed} = 10^{4.8}\,{\rm M}_\odot$, which is the value nearest to $m_{\rm BH,seed} \approx 10^{4.7}\,{\rm M}_\odot$ in the best-fitting `ThermalKinetic` model to the GSMF and SSMR, yields an SHMR that is closest in shape and normalization to the SHMR inferred from the data by the semi-empirical models of B. P. Moster et al. (2018) and P. Behroozi et al. (2019). This is expected, since constraints on the GSMF and SHMR are correlated: fitting the model to either relation should improve the agreement with the other.

We next move to the middle panel of Fig. 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.[16] In essence, increasing (decreasing) $f_{\rm E}$ moves the bulk of the SHMR down (up) in both models, as the SN feedback becomes stronger (weaker), leading to less (more) stellar mass formed by $z = 0$. Crucially, in the `Basic` model, the shape of the SHMR exhibits minimal dependence on $f_{\rm E}$, which renders it impossible for this model to match the SHMR of the semi-empirical models solely by adjusting $f_{\rm E}$. Setting $f_{\rm E}$ to 2 or 3 to achieve realistic stellar-to-halo mass ratios at $M_{\rm halo} \sim 10^{11}\,{\rm M}_\odot$ (pink through violet dotted curves) while simultaneously lowering $m_{\rm BH,seed}$ from its best-fitting value of $\approx 10^{4.0}\,{\rm M}_\odot$ to better match the peak of the SHMR of the semi-empirical models, cannot help either because doing so will either result in excessively high stellar masses for the most massive haloes ($M_{\rm halo} \gtrsim 10^{13}\,{\rm M}_\odot$) or undershoot the peak of the SHMR.

The agreement with the semi-empirical models' SHMRs 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 $M_{\rm halo}$ and increases it at high $M_{\rm halo}$, thereby steepening the slope of the SHMR. This helps the `ThermalKinetic` model obtain a better fit to the observed GSMF, as we have seen in Fig. 2, and correspondingly, to the SHMR, as is seen in the current figure. Such a behaviour of the SHMR 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 (E. Chaikin et al. 2023) but is hindered by poor sampling in low-mass objects (see Section 2.2.5).

5.1.4 *Fitting to the observed GSMF and SSMR separately and simultaneously*

The best-fitting models that we have discussed so far fit simultaneously the observed GSMF and SSMR. We now investigate the effect of fitting the models separately to either the GSMF or the SSMR. Mathematically, this means setting the log likelihood function in equation (15), $\ln \mathcal{L}(\boldsymbol{\theta})$, to either $\ln \mathcal{L}_{\rm GSMF}(\boldsymbol{\theta})$ or $\ln \mathcal{L}_{\rm SSMR}(\boldsymbol{\theta})$, instead of the sum of the two.

Fig. 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 SSMR (brown), or both the GSMF and SSMR (yellow). 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 yellow region shows the 1σ scatter in the simulation whose model was simultaneously fit to the GSMF and SSMR.

The left panel shows the $z = 0$ GSMF, the middle panel shows the $z = 0$ SSMR, and the right panel shows the $z = 0$ SHMR. As in Fig. 2, the solid lines become dotted in the mass range where galaxies are poorly resolved ($M_* < 10^9\,{\rm M}_\odot$) and where the number of galaxies per bin drops below 5. In the right panel, which plots the halo mass instead of stellar mass, the mass below which the solid lines turn to dotted is $M_{\rm halo} = 10^{11}\,{\rm M}_\odot$. 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 from B. P. Moster et al. (2018), P. Behroozi et al. (2019), J. A. Hardwick et al. (2022), and S. P. Driver et al. (2022) are displayed as black points.

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

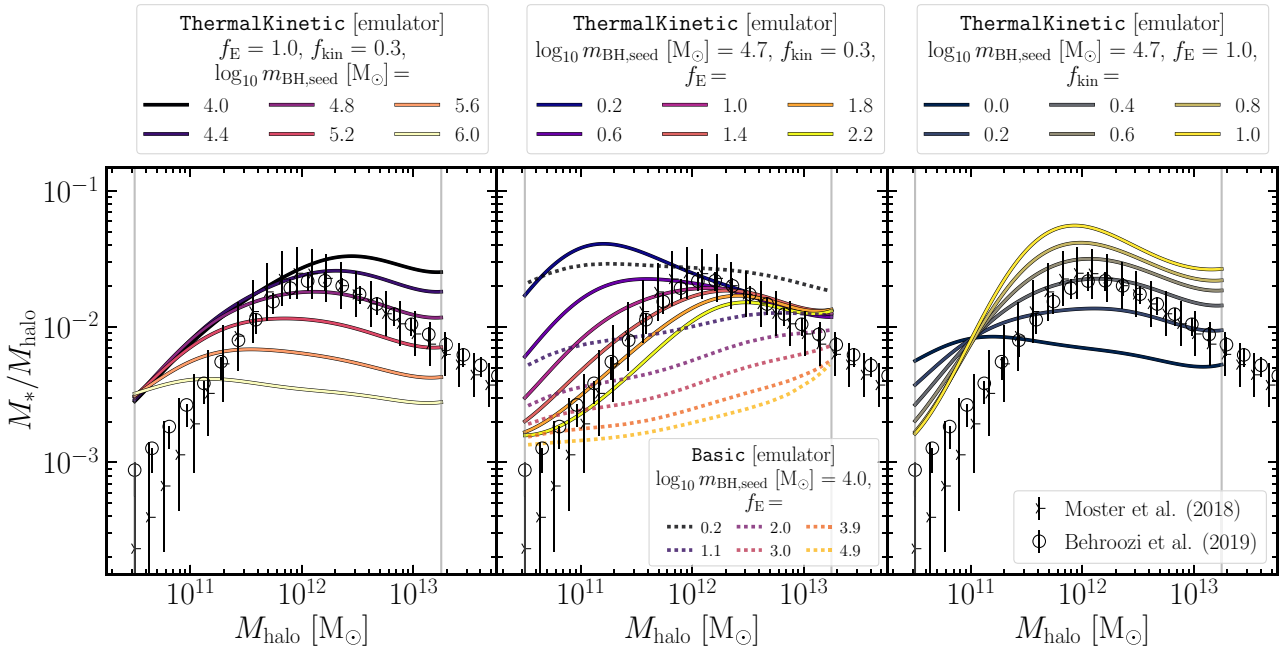

Figure 4. The stellar to halo mass relation (SHMR) at $z = 0$ predicted by the trained emulators. The results are shown for the `ThermalKinetic` model fit to the observed GSMF and SSMR. The individual panels show how the emulated SHMR varies with the BH seed mass ($m_{BH,seed}$; *left*), the energy in SN feedback in units of 10^{51} erg (f_E; *middle*), and the fraction of SN energy injected in kinetic form (f_{kin}; *right*). Different colours correspond to different values of each parameter. Only one parameter is varied at a time, while the other parameters are fixed to their best-fitting values as indicated in the legends. The vertical solid lines designate the mass range within which the emulators were trained on the simulations. For reference, each panel shows the data from the semi-empirical models of B. P. Moster, T. Naab & S. D. M. White (2018) and P. Behroozi et al. (2019), displayed as black points. Additionally, the middle panel shows the SHMR in the `Basic` model, also for different values of f_E (thin dotted curves). Regardless of the value of f_E, the SHMR in the `Basic` model is always too flat compared to the data. This problem is resolved in the `ThermalKinetic` model for high enough f_{kin}.

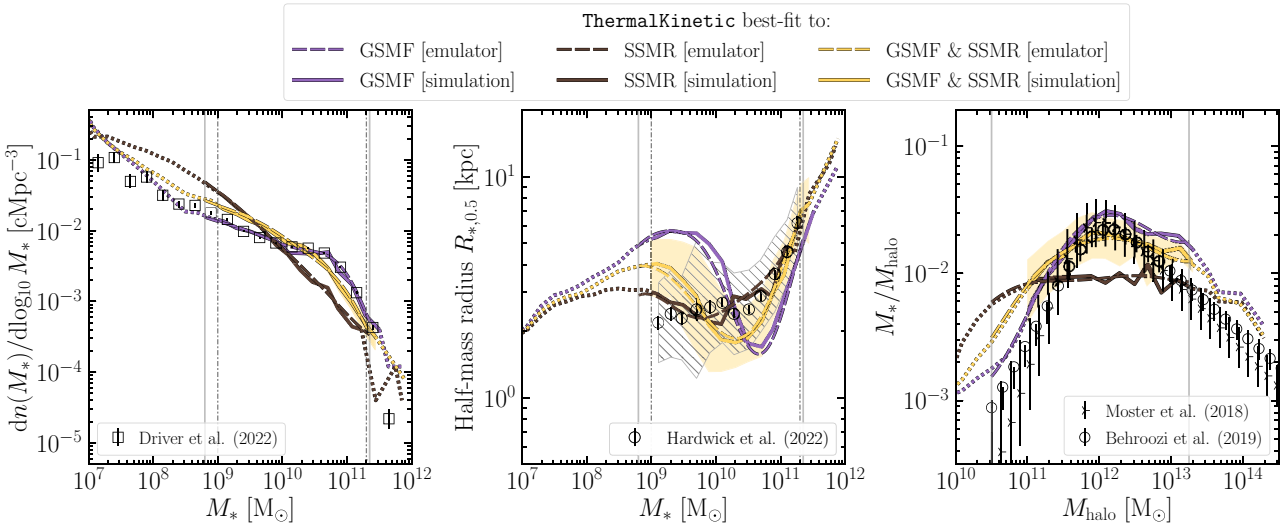

Figure 5. Predictions of the best-fitting `ThermalKinetic` model fit to the observed GSMF (purple), galaxy SSMR (brown), or to both the GSMF and SSMR (yellow). We show the $z = 0$ GSMF (*left*), the $z = 0$ SSMR (*middle*), and the $z = 0$ stellar to halo mass relation (SHMR; *right*). The emulator predictions 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. 2. There are no vertical dash–dotted lines in the right panel because we do not fit the model to the SHMR. Fitting only to the observed GSMF (SSMR) results in a good match to the observed GSMF (SSMR) but a poor match to the SSMR (GSMF). Fitting to both observed relations at the same time produces only a reasonable match to the two constraints.

Examining the left and middle panels of Fig. 5, we see that the ThermalKinetic model with the best-fitting parameters matches the observed GSMF (SSMR) well if it is fit to the GSMF (SSMR) alone. This, however, comes at the expense of a poor fit to the other observed relation, which was left out of the fitting. In contrast, fitting the model to both the GSMF and SSMR 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 SSMR. 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.

The right panel of Fig. 5, which shows the SHMR, 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 SSMR follow a nearly flat SHMR, which is clearly wrong. The two other cases, in which the observed GSMF was used as a constraint to the model, have SHMRs whose shape resembles that in B. P. Moster et al. (2018) and P. Behroozi et al. (2019). Interestingly, the model that is constrained only by the GSMF produces an SHMR whose peak is ≈ 0.1 dex higher than in the two semi-empirical models.

Similar to Fig. 2, for all cases of observational constraints and for all emulated relations shown in Fig. 5, the simulations closely follow the predictions of the emulators, thereby validating our emulator-based approach.

Fig. 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 SSMR (brown), or to both (yellow). 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 χ^2_ν for each fit. Fitting to either the GSMF or SSMR yields a χ^2_ν of order unity or less, indicating that the model is a good fit to the data or slightly overfits them, given the emulator uncertainties. In contrast, fitting simultaneously to the GSMF and SSMR yields a χ^2_ν of 4.7, indicating the model lacks the necessary complexity to match both observed relations simultaneously.

Examining the peaks of the posterior distributions reveals that, in each case, the models calibrated to only the GSMF or only the SSMR occupy different regions of parameter space. Specifically, the model best-fitting the SSMR (brown) prefers a BH seed mass of $m_{\rm BH,seed} \approx 10^{4.3}\,{\rm M}_\odot$, an SN energy per event of $f_{\rm E} \approx 1.0$, and no kinetic SN feedback ($f_{\rm kin} \approx 0$). In contrast, the model best-fitting the GSMF (purple) yields $m_{\rm BH,seed} \approx 10^{4.8}\,{\rm M}_\odot$, $f_{\rm E} \approx 1.3$, and $f_{\rm kin} \approx 0.6$. The best-fitting parameter $m_{\rm BH,seed}$ ($f_{\rm E}$) of the model calibrated to both the GSMF and SSMR is close to that of the model calibrated only to the GSMF (SSMR), while the value of $f_{\rm kin}$ lies in between the two, at $f_{\rm kin} \approx 0.3$.

5.2 Calibration diagnostics for ThermalKinetic_varΔT$_{\rm SN}$ and ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ models

Having learned that neither the Basic model nor the ThermalKinetic model can simultaneously fit the observed GSMF and SSMR, we turn our attention to the more complex models: ThermalKinetic_varΔT$_{\rm SN}$ and ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$.

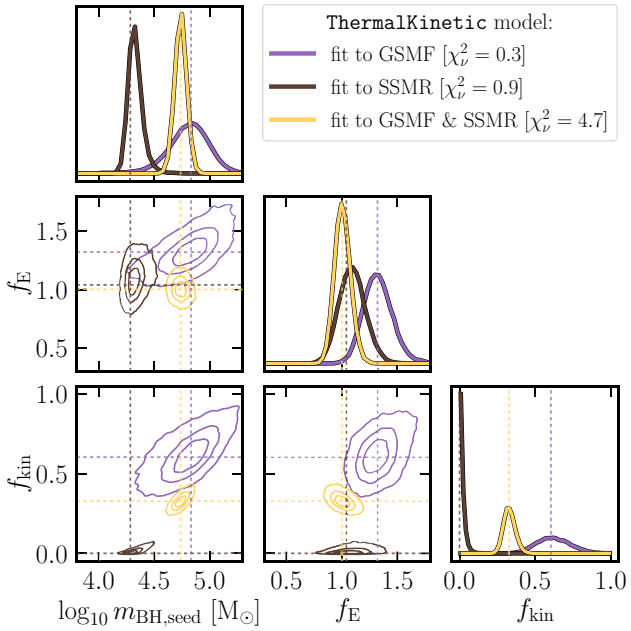

Figure 6. Posterior distributions of the parameters of the ThermalKinetic model resulting from fitting the emulator to the observed GSMF (purple), the observed SSMR (brown), or both the GSMF and SSMR (yellow), with the χ^2_ν value of each fit indicated in the legend. The three contours of each colour correspond to 34, 68, and 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 SSMR belong to very different regions of the parameter space. The model fit to both the GSMF and SSMR (yellow) is located in between the models fit to the GSMF and SSMR separately.

5.2.1 *Galaxy stellar mass function and galaxy sizes*

Fig. 7 shows the GSMF and SSMR at $z = 0$ for the best-fitting ThermalKinetic_varΔT$_{\rm SN}$ (navy-blue) and ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ (light-blue) models. The different symbols and line styles have the same meaning as in Fig. 2.

It is evident that both the ThermalKinetic_varΔT$_{\rm SN}$ and ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ models outperform the Basic and ThermalKinetic models, whose GSMF and SSMR were shown in Fig. 2. The GSMF in the ThermalKinetic_varΔT$_{\rm SN}$ model agrees with the observed GSMF for $M_* \lesssim 10^{9.5}\,{\rm M}_\odot$, overshoots it by up to ≈ 0.15 dex in the range $10^{9.5} \lesssim M_*/{\rm M}_\odot \lesssim 10^{10.5}$, and undershoots it by up to ≈ 0.25 dex at higher M_*. The best performance is achieved for the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model whose GSMF closely follows the observed GSMF across more than 4 dex in M_*, with deviations within the fitting range ($10^9 < M_*/{\rm M}_\odot < 10^{11.3}$) remaining within ≈ 0.1 dex.

The agreement between the SSMR in the ThermalKinetic_varΔT$_{\rm SN}$ model and the observed data is less satisfactory than for the GSMF. Similar to the ThermalKinetic model shown in Fig. 2, the ThermalKinetic_varΔT$_{\rm SN}$ model overshoots the observed sizes of galaxies with $M_* \lesssim 10^{9.5}\,{\rm M}_\odot$ and exhibits a dip in the SSMR at $M_* \approx 2 \times 10^{10}\,{\rm M}_\odot$ – albeit less pronounced than ThermalKinetic – which is not present in the observed relation. In contrast, this dip is absent in the SSMR of the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model, which closely

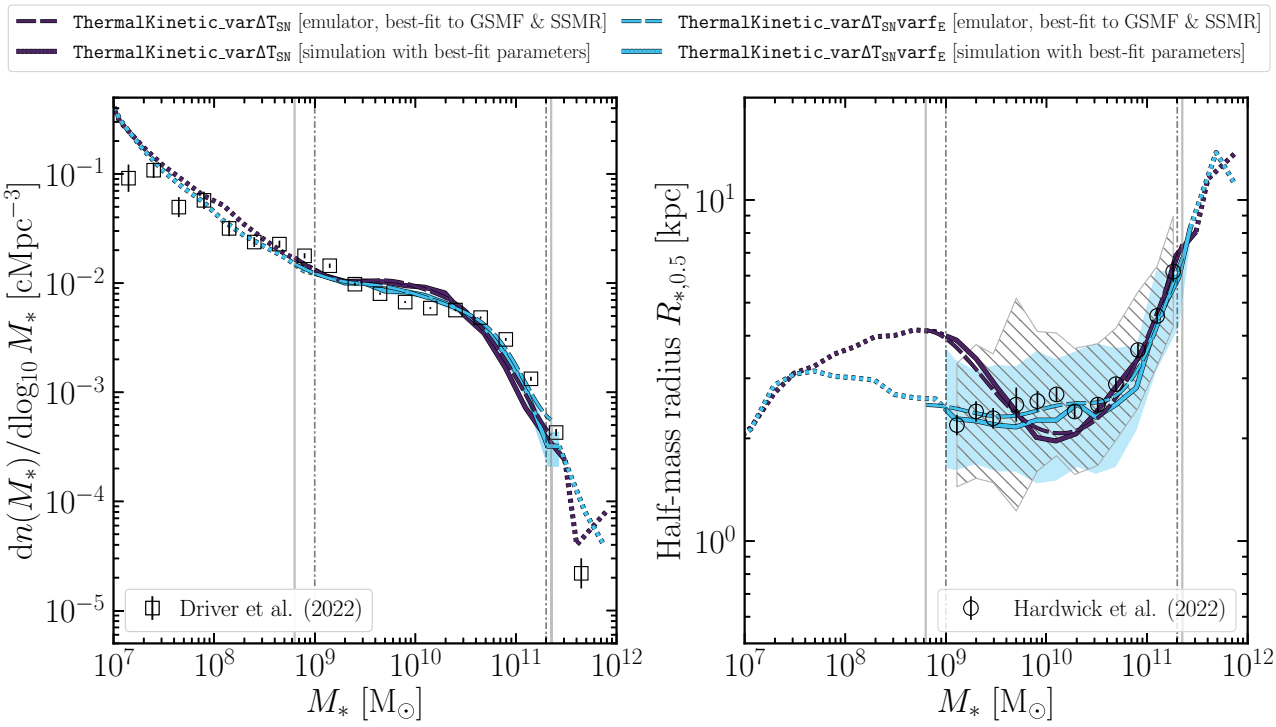

Figure 7. As Fig. 2 but showing the $z = 0$ GSMF (*left*) and median SSMR (*right*) for the ThermalKinetic_varΔT$_{SN}$ (navy-blue) and ThermalKinetic_varΔT$_{SN}$varf$_{E}$ (light-blue) models. While the ThermalKinetic_varΔT$_{SN}$ model shows only reasonably good agreement with the observed GSMF and disagrees in shape with the observed SSMR at $M_* \lesssim 10^{10.5}\,M_\odot$, the ThermalKinetic_varΔT$_{SN}$varf$_{E}$ model successfully reproduces both observational constraints across the entire fitting range ($10^9 < M_*/M_\odot < 10^{11.3}$).

follows the observed SSMR for all stellar masses in the fitting range, $10^9 < M_*/M_\odot < 10^{11.3}$, reproducing both the median galaxy half-mass size and its scatter at fixed M_*. For stellar masses $M_* < 10^9\,M_\odot$, where the models are not fit to observational data, the median half-mass radius does not drop below 2 kpc in either model. This 'floor' likely arises due to the relatively low numerical resolution of the simulations: at m7 resolution, a galaxy with a stellar mass $10^9\,M_\odot$ is sampled with only ~ 100 stellar particles.

To sum up, among the four considered models – Basic, ThermalKinetic, ThermalKinetic_varΔT$_{SN}$, and ThermalKinetic_varΔT$_{SN}$varf$_{E}$ – ThermalKinetic_varΔT$_{SN}$varf$_{E}$ is the only model that simultaneously reproduces both the observed GSMF and SSMR. Furthermore, as in previous figures that showed results from emulators and simulations for the same model parameters, Fig. 7 confirms that the emulator errors are negligibly small.

5.2.2 *Posterior distributions of the model parameters*

Fig. 8 displays the posterior distribution of the parameters in ThermalKinetic_varΔT$_{SN}$ (navy-blue) and ThermalKinetic_varΔT$_{SN}$varf$_{E}$ (light-blue) after each model was fit to both the observed GSMF and SSMR. Because one of the four parameters from ThermalKinetic_varΔT$_{SN}$ does not exist in ThermalKinetic_varΔT$_{SN}$varf$_{E}$, and vice versa, we plot the unique parameters of the two models in the bottom row of the figure at the same time: f_E for ThermalKinetic_varΔT$_{SN}$ and $P_{E,pivot}$ (in log) for ThermalKinetic_varΔT$_{SN}$varf$_{E}$. We show the same range of values for both parameters but attach two different labels to the panel axes, which for clarity are shown in the colours of the corresponding models (navy-blue and light-blue).

The ThermalKinetic_varΔT$_{SN}$ model prefers a BH seed mass of $\approx 10^{4.6}\,M_\odot$, while the ThermalKinetic_varΔT$_{SN}$varf$_{E}$ model favours a slightly higher value of $m_{BH,seed} \approx 10^{4.8}\,M_\odot$. The fraction of SN energy injected in kinetic form is close to 10 per cent in both models, which is lower than the $f_{kin} \approx 0.3$ in the ThermalKinetic model calibrated to the same observational data. This is likely because, unlike in ThermalKinetic, the heating temperature ΔT_{SN} in ThermalKinetic_varΔT$_{SN}$ and ThermalKinetic_varΔT$_{SN}$varf$_{E}$ can vary between $10^{6.5}$ and $10^{7.5}$ K. Thermal SN feedback with low ΔT_{SN} can reproduce some of the effects of kinetic feedback with low Δv_{kick}, reducing the need for a high f_{kin}. 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.

In both models, the best-fitting value of the pivot density in the SN thermal feedback with a variable heating temperature is $n_{SN,pivot} \approx 0.5\ cm^{-3}$, with ThermalKinetic_varΔT$_{SN}$varf$_{E}$ (ThermalKinetic_varΔT$_{SN}$) favouring slightly higher (lower) values. At $\Delta T_{SN} = \Delta T_{SN,pivot} = 10^{6.5}$ K, $n_{SN,pivot} = 0.5\,cm^{-3}$ corresponds to the C. Dalla Vecchia & J. Schaye (2012) critical density for $f_t \approx 2.5$ (see equation 6). The best-fitting value of f_E in the ThermalKinetic_varΔT$_{SN}$ model is ≈ 1.3, which is slightly higher than in the ThermalKinetic model ($f_E \approx 1.0$). The increase in f_E likely originates from the fact that the average ΔT_{SN} in ThermalKinetic_varΔT$_{SN}$ is lower than $10^{7.5}$ K, which results in weaker SN thermal feedback compared to ThermalKinetic, as more energy is radiated

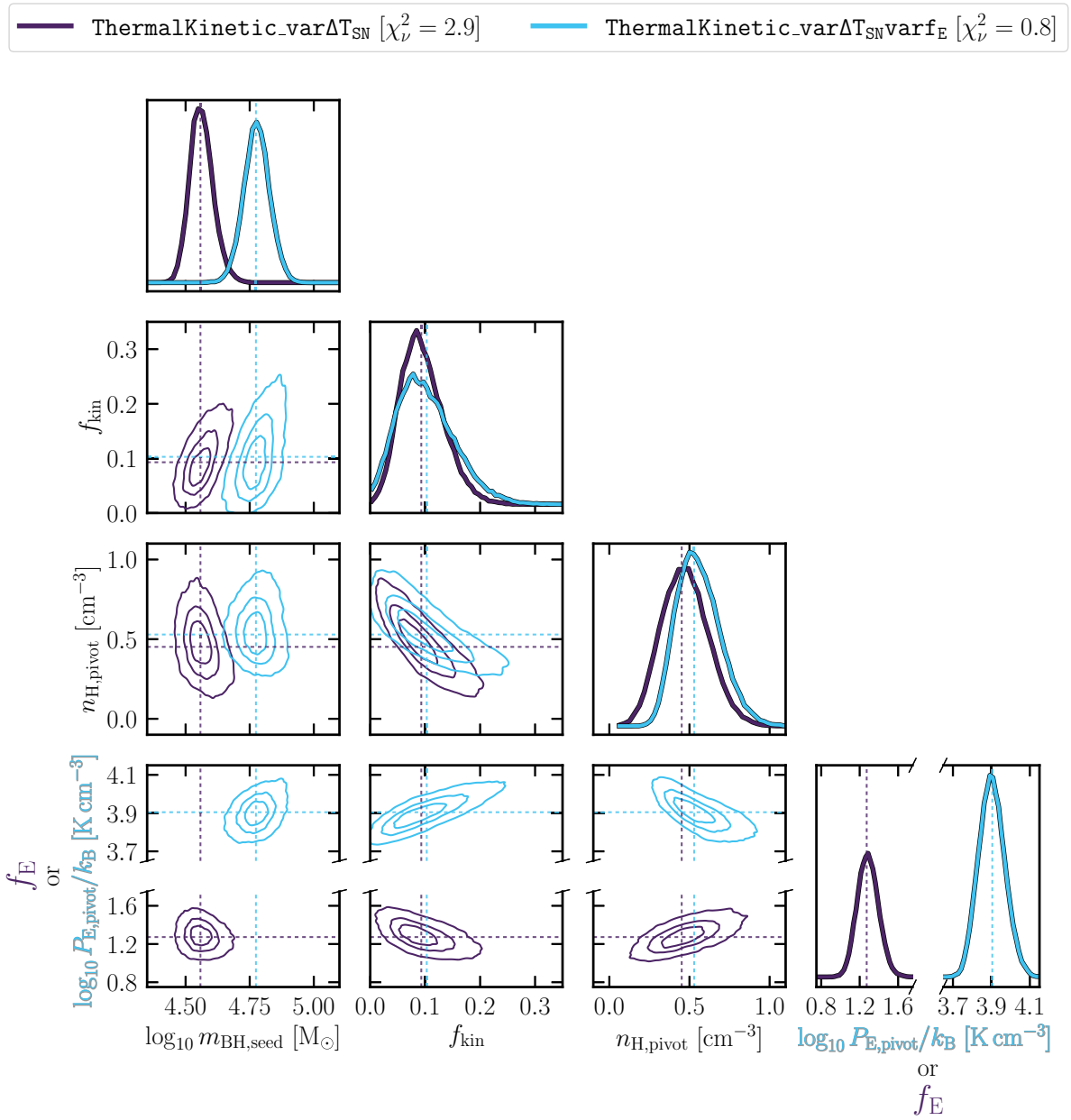

Figure 8. Posterior distributions of the parameters of the ThermalKinetic_varΔT$_{\rm SN}$ model (navy-blue) and the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model (light-blue), obtained by fitting the emulator to the observed $z = 0$ GSMF and SSMR. The contours of the same colour indicate the 34, 68, and 95 per cent credibility regions of the posterior distributions. Vertical and horizontal dotted lines mark the best-fitting parameter values for each model, corresponding to the maxima of their respective posterior distributions. While the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ (ThermalKinetic_varΔT$_{\rm SN}$) model favours a BH seed mass of $\approx 10^{4.8}$ M$_\odot$ ($10^{4.6}$ M$_\odot$), the fraction of SN energy injected in kinetic form, $f_{\rm kin}$, and the pivot density in SN thermal feedback, $n_{\rm SN,pivot}$, are similar in both models: $f_{\rm kin} \approx 0.1$ and $n_{\rm H,pivot} \approx 0.5\,{\rm cm}^{-3}$. In the bottom row, we show the parameters that are unique to each model: $f_{\rm E}$ for ThermalKinetic_varΔT$_{\rm SN}$, and $P_{\rm E,pivot}/k_{\rm B}$ for ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$; their best-fitting values are approximately 1.3 and $10^{3.9}\,{\rm K\,cm}^{-3}$, respectively.

away due to the enhanced radiative cooling rates of gas heated to lower $\Delta T_{\rm SN}$. To compensate for the weaker SN feedback, the energy per SN in units of 10^{51} erg is increased from ≈ 1.0 to 1.3.

The best-fitting value of the pivot birth pressure in the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model is $P_{\rm E,pivot}/k_{\rm B} \approx 10^{3.9}\,{\rm K\,cm}^{-3}$, which is close to the median stellar birth pressure in the simulation with this best-fitting model: $\approx 10^{3.75}\,{\rm K\,cm}^{-3}$. Substituting this median value into equation (2), along with the other model parameters that were not considered in the calibration ($f_{\rm E,min} = 0.1$, $f_{\rm E,max} = 4$, and $\sigma_{\rm P} = 0.3$), we obtain an SN energy at the median stellar birth pressure of $f_{\rm E}(P_{\rm birth}/k_{\rm B} = 10^{3.75}\,{\rm K\,cm}^{-3}) \approx 1.6$. The value of $f_{\rm E}$ averaged over all stellar particles formed in the simulation is slightly higher[17]:

[17]We note that, as described in Section 2.2.3, COLIBRE computes the energy in CC SN feedback per stellar particle by integrating the G. Chabrier (2003) IMF from $m_{\rm min,CCSN} = 8\,{\rm M}_\odot$ to $m_{\rm max,CCSN} = 100\,{\rm M}_\odot$, whereas EAGLE used a lower integration limit of $m_{\rm min,CCSN} = 6\,{\rm M}_\odot$. As a result, $f_{\rm E} = 1.8$ in COLIBRE corresponds to $f_{\rm E} \approx 1.2$ in EAGLE.

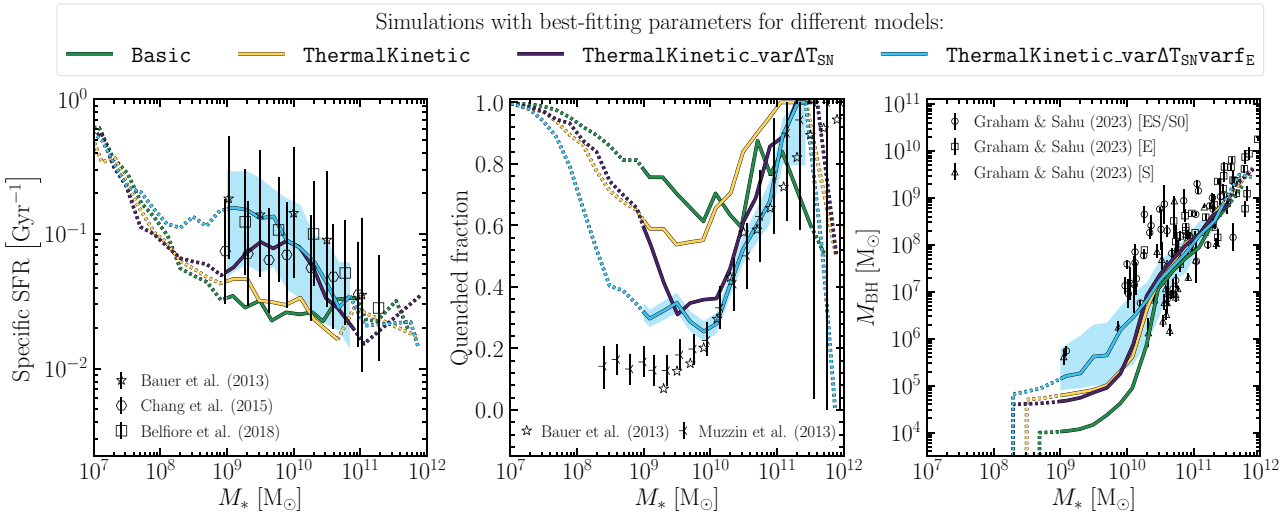

Figure 9. The median sSFR of active galaxies (sSFR $> 10^{-2}$ Gyr^{-1}) versus stellar mass (*left*), the fraction of quenched galaxies versus stellar mass (*middle*), and the median mass of SMBHs versus stellar mass (*right*), all shown at $z = 0$. Differently coloured solid curves show the results from the simulations with the best-fitting parameter values for the `Basic` (green), `ThermalKinetic` (yellow), `ThermalKinetic_varΔT_SN` (navy-blue), and `ThermalKinetic_varΔT_SNvarf_E` (light-blue) models. All models were fit to the observed $z = 0$ GSMF and SSMR. The solid curves turn into dotted curves at stellar mass below 10^9 M$_\odot$ indicating that those galaxies are poorly resolved, and when the number of objects per bin is less than 5 indicating the limit due to the simulated volume. The shaded light-blue region shows the 1σ scatter for the sSFR – M_* relation and the BSMR, and the 1σ confidence interval for the quenched fraction – M_* relation, all for the `ThermalKinetic_varΔT_SNvarf_E` model. The median BH mass drops to zero at $M_* \lesssim 10^{8.5}$ M$_\odot$ in all models because the corresponding haloes are not massive enough to be seeded with a BH particle. A compilation of observational data is shown as black symbols. Both the `ThermalKinetic_varΔT_SN` and `ThermalKinetic_varΔT_SNvarf_E` models show a good agreement with the comparison data for all three relations, with the `ThermalKinetic_varΔT_SNvarf_E` model exhibiting marginally better sSFR and quenched fractions at $M_* \lesssim 10^{10}$ M$_\odot$.

$\langle f_E \rangle \approx 1.8$. Both of these values of f_E are comparable to the best-fitting (constant) value in `ThermalKinetic_varΔT_SN`, $f_E = 1.3$, indicating that both models favour an (average) SN feedback energy slightly higher than the theoretical expectation, $f_E = 1$, corresponding to 10^{51} erg. In Appendix B, we provide further details on f_E, including the redshift evolution of the median values of f_E and $P_{\rm birth}$ as measured in the simulation using the best-fitting `ThermalKinetic_varΔT_SNvarf_E` model.

The legend of Fig. 8 lists the χ^2_ν values for the fits to the observational data: 2.9 for the `ThermalKinetic_varΔT_SN` model and 0.8 for `ThermalKinetic_varΔT_SNvarf_E`. This confirms that both models outperform the `Basic` and `ThermalKinetic` models, with `ThermalKinetic_varΔT_SNvarf_E` providing the best match to the GSMF and SSMR. We will use these results in Section 5.4 to define the fiducial COLIBRE model at m7 and higher resolutions.

Lastly, we note that based on isolated galaxy simulations at much higher resolution ($m_{\rm gas} = 10^5$ M$_\odot$), E. 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 H I velocity dispersion and the galaxy SFR surface density, as well as the observed spatially-resolved KS star-formation law (C. J. Kennicutt Robert C. et al. 2007). The latter was confirmed by F. S. J. Nobels et al. (2024), 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 the `ThermalKinetic_varΔT_SNvarf_E` best-fitting model prefers $f_{\rm kin} \approx 0.1$.

5.3 Predictions for galaxy properties that were not included in emulator-based calibration

In this section, we explore simulation predictions for galaxy properties that have not been previously discussed and therefore were not included in the emulator-based calibration of the model parameters. The following figures present the results for the four best-fitting models, `Basic`, `ThermalKinetic`, `ThermalKinetic_varΔT_SN`, and `ThermalKinetic_varΔT_SNvarf_E`, which have all been fit to both the observed GSMF and SSMR. We show only the results from the simulations, as the emulators were constructed exclusively for the $z = 0$ GSMF, SSMR, and SHMR.

5.3.1 *Star formation rates and quenched fraction*

Fig. 9 displays the specific star formation rates (sSFR) of active galaxies, the fraction of quenched galaxies, and the BH masses. All relations are shown at $z = 0$ and plotted versus galaxy stellar mass. All quantities are measured within 3D spherical apertures of radius 50 pkpc, and galaxy SFRs are computed using the instantaneous SFRs of gas particles. In a given stellar mass bin, we show the median sSFR and the median (subgrid) mass of the BHs. If a subhalo contains multiple BH particles, then the mass of the most massive BH is used to compute the median. We define a galaxy as 'active' if its instantaneous sSFR exceeds 10^{-2} Gyr^{-1} (e.g. A. R. Wetzel, J. L. Tinker & C. Conroy 2012); otherwise, it is considered quenched. The solid curves give the results from the

simulations with the four best-fitting models: `Basic` (green), `ThermalKinetic` (yellow), `ThermalKinetic_var`$\Delta T_{\rm SN}$ (navy-blue) and `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$f_{\rm E}$ (light-blue). The shaded light-blue region represents the uncertainty in the `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$f_{\rm E}$ model, computed as the 16$^{\rm th}$ to 84$^{\rm th}$ percentiles for the relations between sSFR and M_* and between BH mass and M_*, and using the Clopper-Pearson interval at the 68 per cent confidence level for the quenched fraction – M_* relation.

For comparison, we use the sSFR–stellar mass relations for star-forming galaxies from A. E. Bauer et al. (2013), Y.-Y. Chang et al. (2015), and F. Belfiore et al. (2018). The relation from A. E. Bauer et al. (2013) is based on $\sim 10^5$ galaxies from GAMA DR1 (S. P. Driver et al. 2011) within the redshift range $0.05 < z < 0.32$, where galaxies were classified as star-forming based on the flux and equivalent width of the H α line. Y.-Y. Chang et al. (2015) used $\sim 10^6$ SDSS galaxies with $z < 0.2$, combined with four-band WISE photometry (E. L. Wright et al. 2010), and identified star-forming galaxies using colour–colour diagram cuts. Finally, the relation from F. Belfiore et al. (2018) is based on $\sim 10^4$ galaxies at $0.01 < z < 0.15$ from the MaNGA survey (K. Bundy et al. 2014), where star-forming galaxies were classified using the Baldwin–Phillips–Terlevich diagram (J. A. Baldwin, M. M. Phillips & R. Terlevich 1981). For all three datasets, the y values represent the median sSFR, and the error bars indicate the 1σ scatter. For quenched fractions, we use $z \approx 0.1$ data from GAMA DR1 presented by A. E. Bauer et al. (2013), which were re-calculated by P. Behroozi et al. (2019) using our threshold for quiescent galaxies of $\mathrm{sSFR} < 10^{-2}\,\mathrm{Gyr}^{-1}$, as well as data from A. Muzzin et al. (2013) at $0.2 < z < 0.5$, based on the COSMOS/UltraVISTA galaxy catalogue, where quiescent galaxies were defined using UVJ colour selection. For the BSMR, we adopt the measurements[18] from A. W. Graham & N. Sahu (2023), whose sample is subdivided by morphological type: E, ES/S0, and S (shown with different black symbols as indicated in the legend). Where necessary, we correct for differences in the assumed stellar IMF by converting all data to the G. Chabrier (2003) IMF.

Overall, the agreement between the simulations and comparison data improves with increasing model complexity. First, the `Basic` and `ThermalKinetic` models predict a $z = 0$ sSFR–stellar mass relation with an unrealistically flat shape, offset by more than 0.4 dex toward lower values at low and intermediate stellar masses ($M_* \lesssim 10^{10.5}\,\mathrm{M}_\odot$) compared to observed trends. The fraction of quenched galaxies at similar stellar masses is significantly overestimated. This discrepancy in both models arises from the use of a constant heating temperature for SN thermal feedback, $\Delta T_{\rm SN} = 10^{7.5}$ K, which results in excessively large energy injections from (clustered) SNe. By $z = 0$, this overly powerful SN feedback in low- and intermediate-mass galaxies has likely disrupted, heated, and/or ejected most of the cold gas that would otherwise contribute to star formation, leading to lower sSFRs and higher quenched fractions at fixed M_*.

Switching to a density-dependent heating temperature in the `ThermalKinetic_var`$\Delta T_{\rm SN}$ and `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$f_{\rm E}$ models significantly improves agreement with the data. Both models reproduce the observed quenched fraction for galaxies with stellar masses $M_* \gtrsim 10^{9.5}\,\mathrm{M}_\odot$ and the sSFRs of active galaxies for $M_* \gtrsim 10^9\,\mathrm{M}_\odot$, with `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$f_{\rm E}$ exhibiting slightly better agreement for $M_* \sim 10^9\,\mathrm{M}_\odot$. At lower stellar masses ($M_* \lesssim 10^9\,\mathrm{M}_\odot$, corresponding to $\lesssim 100$ stellar particles), all four models show an increasing quenched fraction with decreasing stellar mass, which is likely driven by resolution effects. At m7 resolution, the ISM of such low-mass galaxies is sampled by only a few star-forming gas particles, if any, which makes them appear more quenched on average (e.g. J. Schaye et al. 2015), resulting in lower median sSFRs and higher quenched fraction at $M_* \lesssim 10^9\,\mathrm{M}_\odot$. In contrast, in the same mass range, the observed quenched fraction may be biased low due to selection effects, which are known to become progressively more important at lower M_* (e.g. S. Kaviraj et al. 2025).

Focusing on the right panel, we find that in all four models, galaxies with stellar masses $M_* \gtrsim 10^{10.5}\,\mathrm{M}_\odot$ host SMBHs that have grown to masses $\gtrsim 10^7\,\mathrm{M}_\odot$. The masses of these BHs follow a tight relation with the host galaxy's stellar mass, with a slope and normalization that closely match the observed scaling from A. W. Graham & N. Sahu (2023). The agreement with the observational data is expected, as the AGN feedback coupling efficiency, $\varepsilon_{\rm f} = 0.1$ (see equation 12), which determines the normalization of the BSMR, was chosen to produce realistic $z = 0$ SMBH masses in the high-mass galaxies for which BH masses can be measured observationally. The value of $\varepsilon_{\rm f}$ was set *independently* of the emulator-based calibration of SN and AGN feedback to the observed GSMF and SSMR. Due to the self-regulating nature of SMBHs, variations in $\varepsilon_{\rm f}$ primarily influence SMBH masses, while having little to no effect on other galaxy properties such as the GSMF and SSMR (e.g. C. M. Booth & J. Schaye 2009).

We do not observe any large differences between the models, which is expected since all models use the same numerical prescription for BH growth and AGN feedback. The only AGN-related parameter that differs between the models is the best-fitting value of the BH seed mass, $m_{\rm BH,seed}$. Among the four models, the `Basic` model uses the lowest seed mass, while the `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$f_{\rm E}$ model adopts the highest. A higher (lower) $m_{\rm BH,seed}$ results in somewhat faster (slower) BH growth in the latter (former) model, lasting until the BH enters the self-regulating regime where AGN feedback overtakes stellar feedback. As a consequence, at $M_* \lesssim 10^{10.5}\,\mathrm{M}_\odot$, the median BH mass in the `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$f_{\rm E}$ model is higher than in the `Basic` model, whereas at $M_* \gtrsim 10^{10.5}\,\mathrm{M}_\odot$ the two relations converge. The BH masses in the other two models, `ThermalKinetic` and `ThermalKinetic_var`$\Delta T_{\rm SN}$, fall between those in `Basic` and `ThermalKinetic_var`$\Delta T_{\rm SN}$`var`$f_{\rm E}$.

5.3.2 *Cosmic star formation rate density*

Fig. 10 displays the redshift evolution of the cosmic star formation rate density (SFRD), using the same four models as in Fig. 9. The SFRD is computed from the instantaneous SFRs of all star-forming gas particles in the simulations. For comparison, we include the $z \sim 0$ FUV-based SFRD estimate from S. P. Driver et al. (2012), derived from the GAMA DR1, as well as radio-based SFRD estimates from the LOFAR Deep Fields at $0 < z < 4$ (R. K. Cochrane et al. 2023), and rest-frame FIR observations collected with ALMA at $z \approx 4.5 - 5.5$ (Y. Khusanova et al. 2021). We also show results from the ALPINE multi-wavelength

[18] Following the erratum A. W. Graham & N. Sahu (2024), we applied a -0.15 dex correction to the stellar masses from A. W. Graham & N. Sahu (2023).

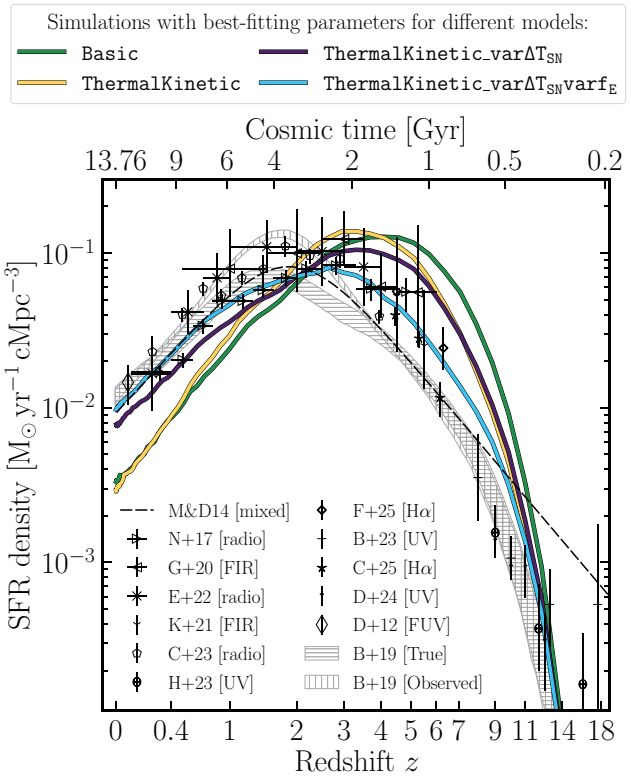

Figure 10. Cosmic SFRD versus redshift from the simulations with the best-fitting models to the $z = 0$ GSMF and SSMR: `Basic`, `ThermalKinetic`, `ThermalKinetic_varΔT_SN`, and `ThermalKinetic_varΔT_SNvarf_E` (differently coloured solid curves). For comparison, the black points show a compilation of observational data from S. P. Driver et al. (2012), M. Novak et al. (2017), C. Gruppioni et al. (2020), Y. Khusanova et al. (2021), A. Enia et al. (2022), R. K. Cochrane et al. (2023), R. Bouwens et al. (2023), Y. Harikane et al. (2023), C. T. Donnan et al. (2024), A. Covelo-Paz et al. (2025), S. Fu et al. (2025), the grey hatched regions indicate the intrinsic (labelled as 'true') and observed SRFD from the UniverseMachine (P. Behroozi et al. 2019), and the black dashed curve shows the best-fitting observed SFRD from P. Madau & M. Dickinson (2014). The use of a variable heating temperature in the SN feedback of the `ThermalKinetic_varΔT_SN` and `ThermalKinetic_varΔT_SNvarf_E` models greatly improves the agreement with the comparison data at $z < 1$. The inclusion of a stellar birth pressure-dependent SN energy in the `ThermalKinetic_varΔT_SNvarf_E` model results in a lower SFRD at $z > 2$, thereby further improving the agreement with the comparison data.

survey, based on 56 sub-mm continuum detections by ALMA in the ECDFS and COSMOS fields at $0.5 < z < 6$ (C. Gruppioni et al. 2020), deep VLA COSMOS radio data at $0.3 < z < 5$ (M. Novak et al. 2017), and VLA radio measurements from a subsample of the GOODS-N survey at $0.1 < z < 3$ (A. Enia et al. 2022). In addition, we include the best-fitting analytic fit for the SFRD evolution from P. Madau & M. Dickinson (2014, black dashed curve), derived from a compilation of IR and UV data across $0 < z < 8$, as well as predictions from the semi-empirical model UniverseMachine (P. Behroozi et al. 2019), showing both the 'true' and 'observed' SFRDs (grey hatched regions), where the latter accounts for systematic uncertainties in observationally inferred SFRs and the former is based on intrinsic SFR values predicted by UniverseMachine. Finally, we show a compilation of recent *JWST* measurements at high redshifts ($4 < z < 18$) from R. Bouwens et al. (2023), Y. Harikane et al. (2023), C. T. Donnan et al. (2024), A. Covelo-Paz et al. (2025), and S. Fu et al. (2025), which are based on UV or H α observations and whose SFRD values are calculated down to an absolute UV magnitude limit of ≈ -17 mag.[19]

The `Basic` and `ThermalKinetic` models predict significantly lower SFRD normalization than observed for $z < 1$, underpredicting the observed SFRD by ≈ 0.5 dex, which is consistent with both models having too low sSFR at $z = 0$ in Fig. 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 in SN thermal feedback: $\Delta T_{\rm SN} = 10^{7.5}$ K. Conversely, the SFRD in the other two models, `ThermalKinetic_varΔT_SN` and `ThermalKinetic_varΔT_SNvarf_E`, which incorporate the variable heating temperature, matches the $z \lesssim 1$ observed SFRD much better, with `ThermalKinetic_varΔT_SNvarf_E` predicting a $z = 0$ SFRD of $\approx 10^{-2}$ M$_\odot$ yr^{-1} cMpc^{-3} and showing the best agreement with the data.

At high redshifts ($z \gtrsim 1$), the SFRD in the `Basic`, `ThermalKinetic`, and `ThermalKinetic_varΔT_SN` models rises steeply with increasing redshift, peaking at $3 < z < 5$. The fact that these models predict a peak SFRD at higher redshifts than observed ($z \approx 2$) suggests that high$-z$ star formation in these models may be overly efficient. Since all four models were calibrated to match the observed GSMF (and SSMR) at $z = 0$, an excess in the early SFRD necessitates a suppressed SFRD at later times to ensure that the correct total stellar mass is formed by $z = 0$.

The agreement with the data is noticeably improved in the `ThermalKinetic_varΔT_SNvarf_E` model, where the amount of stellar mass formed before $z = 3$ is significantly reduced compared to the other models. The SFRD in the `ThermalKinetic_varΔT_SNvarf_E` model exhibits a broad peak between $z = 4$ and 1 and starts steeply declining with cosmic time only thereafter. This improvement is caused by the dependence of the SN energy on the stellar birth pressure (see equation 2), which is incorporated only into the `ThermalKinetic_varΔT_SNvarf_E` model. Specifically, the SN feedback in the `ThermalKinetic_varΔT_SNvarf_E` is more energetic at higher redshifts, as the star formation at high z proceeds on average in higher gas-pressure environments (see Appendix B for further details). 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 $z = 0$ SSMR at $M_* \approx 2 \times 10^{10}$ M$_\odot$, which is absent only in the `ThermalKinetic_varΔT_SNvarf_E` model (see Figs 2 and 7).

5.3.3 *Galaxy stellar mass function and star formation rates at high redshift*

Fig. 11 shows the GSMF and sSFR versus stellar mass at $z = 2$ for the same four simulations as were shown in Figs 9 and 10. At a given stellar mass, we show the median sSFR considering both

[19] Unlike the other four studies, R. Bouwens et al. (2023) used a limit of -19 mag. We re-normalize their results to -17 mag by shifting the SFRD values reported in R. Bouwens et al. (2023) upward by 0.5 dex.

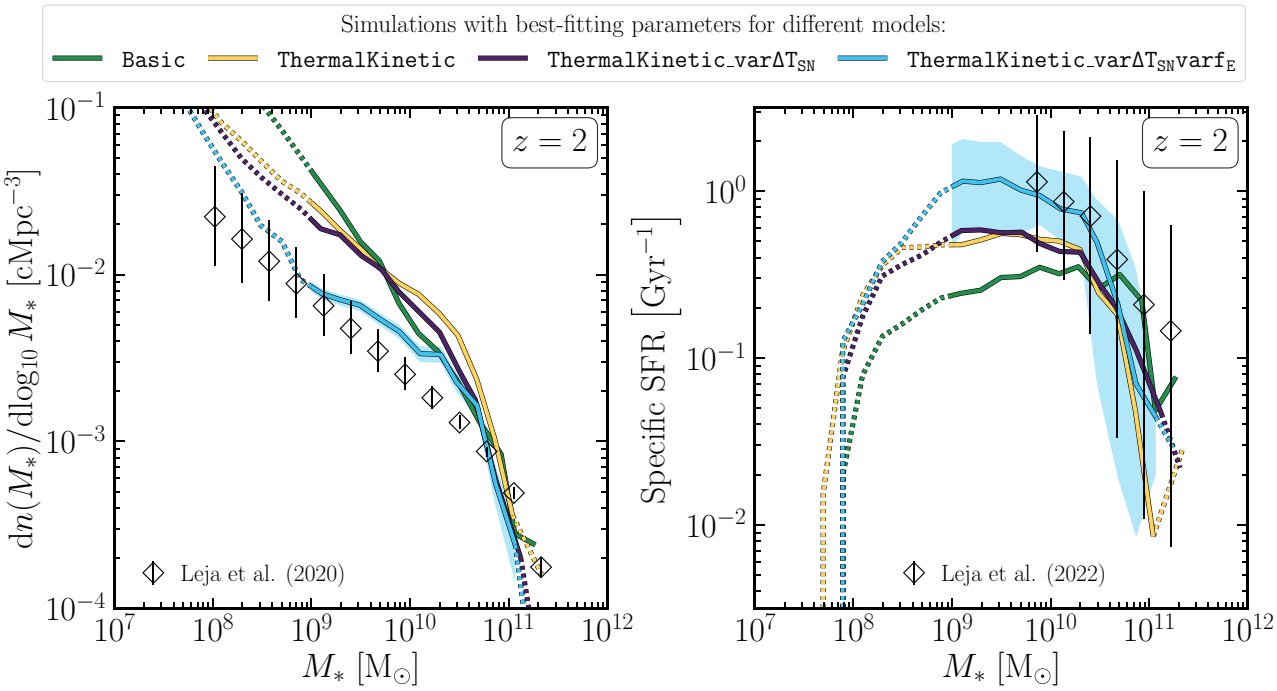

Figure 11. The $z = 2$ GSMF (*left*) and the $z = 2$ median sSFR of all galaxies (i.e. both star-forming and quenched; *right*) as a function of stellar mass. The solid curves show results from simulations using the best-fitting parameters for the `Basic`, `ThermalKinetic`, `ThermalKinetic_varΔT`$_{SN}$, and `ThermalKinetic_varΔT`$_{SN}$`varf`$_E$ models. For comparison, we include the observed $z = 2$ GSMF from J. Leja et al. (2020) and the observed $z = 2$ sSFR – M_* relation from J. Leja et al. (2022). The shaded light-blue region corresponds to the Poisson uncertainty in the left panel and the 16$^{\rm th}$-84$^{\rm th}$ percentile range in the right panel, both for the `ThermalKinetic_varΔT`$_{SN}$`varf`$_E$ model. Among the four models, `ThermalKinetic_varΔT`$_{SN}$`varf`$_E$ (in light-blue) provides the best match to the observed $z = 2$ GSMF and sSFR – M_* relation, although the simulated galaxies appear slightly overmassive – by about 0.2 dex – relative to J. Leja et al. (2020).

passive and active galaxies. For comparison, we display the $z = 2$ GSMF and median sSFR derived by J. Leja et al. (2020), J. Leja et al. (2022) who applied the spectral energy distribution (SED) fitting code Prospector to measure SFRs and stellar masses of $\sim 10^5$ galaxies at $0.2 < z < 3$ from COSMOS2015 and 3D-*HST* galaxy catalogues. The error bars in J. Leja et al. (2020, 2022) correspond to the 16$^{\rm th}$ and 84$^{\rm th}$ percentiles estimated by their best-fitting model.

At $z = 2$, the `ThermalKinetic_varΔT`$_{SN}$`varf`$_E$ model shows significantly better agreement with the data than its three simpler counterparts. Despite being calibrated only to $z = 0$ data, the model broadly reproduces the observed median sSFR, with discrepancies remaining ≈ 0.1 dex at $M_* < 10^{10.5}$ M$_\odot$. Additionally, it is systematically offset by just ≈ 0.2 dex from the observed GSMF, which is comparable to the systematic uncertainty in inferring M_* using SED-fitting codes at these redshifts (e.g. A. Katsianis et al. 2020). In contrast, the other three models perform less well: all systematically underpredict the observed sSFR by more than ≈ 0.3 dex at $M_* < 10^{10.5}$ M$_\odot$, and their GSMF values exceed the observed data by at least ≈ 0.5 dex at $M_* < 10^{10}$ M$_\odot$.

5.3.4 *Cold gas properties*

We next investigate the $z = 0$ properties of cold gas predicted by the simulations. The left and right panels of Fig. 12 show, respectively, the $z = 0$ ratios of galaxy H I 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, and the light-blue shaded regions show the 16$^{\rm th}$ to 84$^{\rm th}$ percentile range for the `ThermalKinetic_varΔT`$_{SN}$`varf`$_E$ model.

For reference, we show the $z \approx 0$ observed H I-to-stellar mass fractions from B. Catinella et al. (2018, xGASS survey) and D. Scholte et al. (2024, ALFALFA survey), both derived from 21 cm line observations. For H_2, we show mass fractions from xCOLD GASS (A. Saintonge et al. 2017), where H_2 masses were derived from CO(1-0) luminosity using the multivariate $\alpha_{\rm CO}$ conversion factor from G. Accurso et al. (2017). To ensure fair comparison with simulations, we divided H_2 masses from A. Saintonge et al. (2017) by 1.36 to remove a contribution from helium, which was included in the conversion factor $\alpha_{\rm CO}$ used by the authors. In both panels, black symbols indicate median observed gas fractions. Error bars for B. Catinella et al. (2018) and A. Saintonge et al. (2017) represent errors on the median, which we estimated via bootstrapping. Grey hatched regions show the 16$^{\rm th}$-84$^{\rm th}$ percentile scatter in the observations: in the left panel for B. Catinella et al. (2018), and in the right panel for A. Saintonge et al. (2017). In each panel, the upper hatched region (slanted top-right) shows the scatter when non-detections are treated as upper limits. The lower hatched region (slanted bottom-left) assumes non-detections are zeroes, resulting in a 16$^{\rm th}$ percentile extending to zero. Together, the lower percentiles of these two regions bracket the range for the true lower percentile.

We stress that the COLIBRE model does not impose an effective pressure and/or temperature floor and uses the non-equilibrium thermochemistry solver CHIMES, coupled to the COLIBRE dust grain model, to compute the abundances of primordial species.

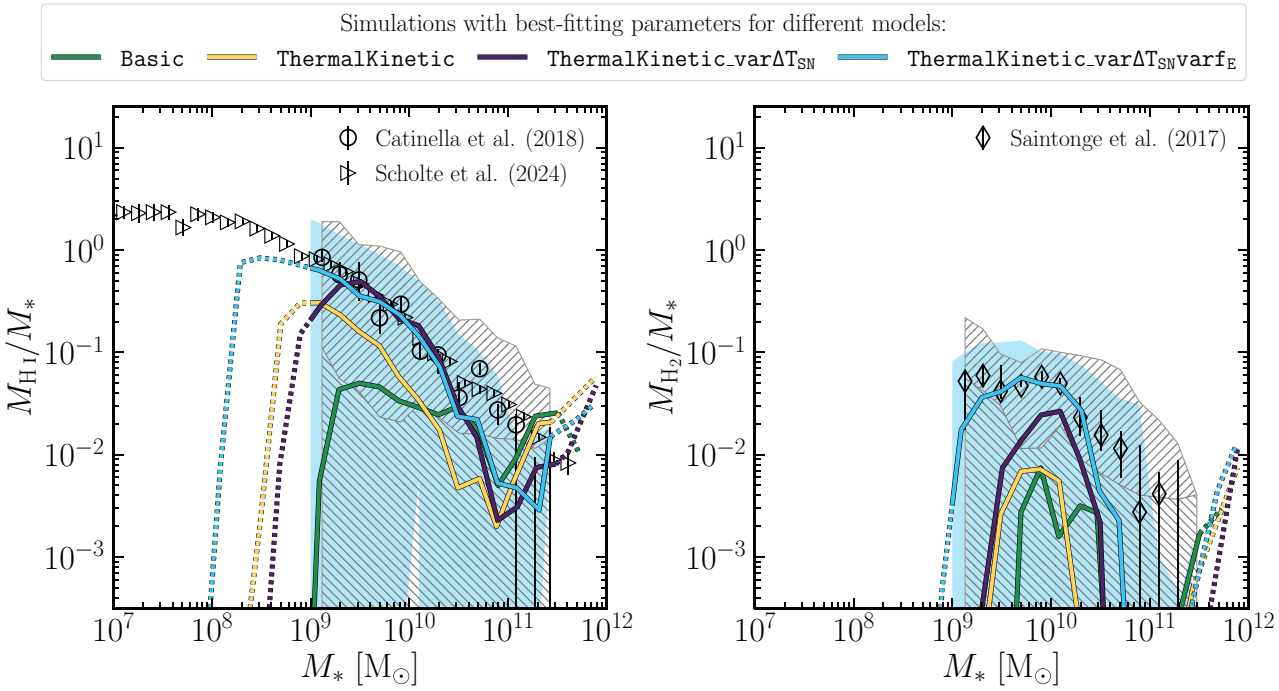

Figure 12. As in Fig. 11, but showing the median $z = 0$ H I-to-stellar mass ratio (*left*) and H_2-to-stellar mass ratio (*right*) as functions of galaxy stellar mass. In both panels, the light-blue shaded regions indicate the 16th to 84th percentile range for the ThermalKinetic_varΔT$_{SN}$varf$_E$ model. The shading extends to very low y values because the lower percentile is influenced by objects with negligible cold gas masses. For comparison, we show the observed H I mass fractions from B. Catinella et al. (2018) and D. Scholte et al. (2024), as well as H_2 measurements from A. Saintonge et al. (2017) (black symbols). We also include the scatter for the measurements from B. Catinella et al. (2018) and A. Saintonge et al. (2017), shown as grey hatched regions: the upper hatched region, slanted toward the top right, represents the 16th-84th percentile scatter when non-detections are treated as upper limits; the lower hatched region, slanted toward the bottom left, extends the 16th percentile of the upper region to zero, corresponding to the case where non-detections are treated as zeroes. In both simulations and observations, molecular gas masses do not include a contribution from helium. Only the ThermalKinetic_varΔT$_{SN}$varf$_E$ model reproduces both the H I and H_2 observed mass fractions for $M_* < 10^{10.5}$ M$_\odot$ (including the observed scatter), while at $M_* > 10^{10.5}$ M$_\odot$ all models undershoot the data.

This allows us to take the (non-equilibrium) H_2 and H I abundances directly from the simulations, which is in contrast to the previous generation of galaxy simulations of representative volumes, such as EAGLE and ILLUSTRISTNG, where the atomic and molecular gas fractions would need to be estimated in post-processing (see e.g. C. d. P. Lagos et al. 2015; B. Diemer et al. 2018; A. Manuwal & A. R. H. Stevens 2023). Such post-processing typically relies on (semi-)analytic models or fitting formulas calibrated using high-resolution radiative transfer simulations of smaller cosmological volumes (e.g. N. Y. Gnedin & A. V. Kravtsov 2011; A. Rahmati et al. 2013; M. R. Krumholz 2013), and may include an intermediate step to obtain the neutral (H_2 + H I) hydrogen fraction before the H_2 and H I fractions can be estimated (e.g. R. A. Crain et al. 2017).

Fig. 12 shows that while both the ThermalKinetic_varΔT$_{SN}$ and ThermalKinetic_varΔT$_{SN}$varf$_E$ models are consistent with the observational data for H I in the stellar mass range $10^9 < M_*/M_\odot < 10^{10.5}$, only ThermalKinetic_varΔT$_{SN}$varf$_E$ also reproduces the observed H_2 mass fractions within this range, whereas ThermalKinetic_varΔT$_{SN}$ systematically undershoots the H_2 data at all M_*. Moreover, comparing the light-blue shaded regions with the grey hatched regions shows that ThermalKinetic_varΔT$_{SN}$varf$_E$ also reproduces the observed scatter for both H I and H_2. By contrast, for the Basic and ThermalKinetic models, both the molecular and atomic gas fractions are, on average, too low compared to the data, consistent with these models underpredicting the observed sSFR at $z = 0$ (Fig. 9). The sharp downturn in the simulated molecular and atomic gas fractions in all four models below the stellar mass $M_* \sim 10^9$ M$_\odot$ is driven by limited numerical resolution, with H_2 mass fractions starting to drop at slightly higher M_* than H I.[20]

At the high-mass end ($M_* > 10^{10.5}$ M$_\odot$), all four models underpredict the observed gas fractions. In fact, there appears to be a tension in that the ThermalKinetic_varΔT$_{SN}$varf$_E$ model reproduces the quenched fraction and sSFRs in this mass range while undershooting the gas fractions. This discrepancy may highlight limitations in our relatively simple treatment of gas accretion onto SMBHs and/or AGN feedback, exacerbated by the relatively low resolution of the simulations ($m_{gas} \sim 10^7$ M$_\odot$), leading to overly efficient depletion of cold gas in massive galaxies. Indeed, J. Schaye et al. (2025) demonstrate that agreement with the data at the high-mass end improves significantly when using the higher m6 resolution of COLIBRE (see their fig. 19). Finally, we note that the dip in the H I mass fraction at $M_* \sim 5 \times 10^{10}$ M$_\odot$ may also be mitigated by mock observing the simulated galaxies, rather than using intrinsic values predicted by the simulations. This would account for observational effects such as H I emission blending, which can artificially boost the inferred

[20]Convergence tests show that higher resolution is required to robustly predict H_2 than to predict H I.

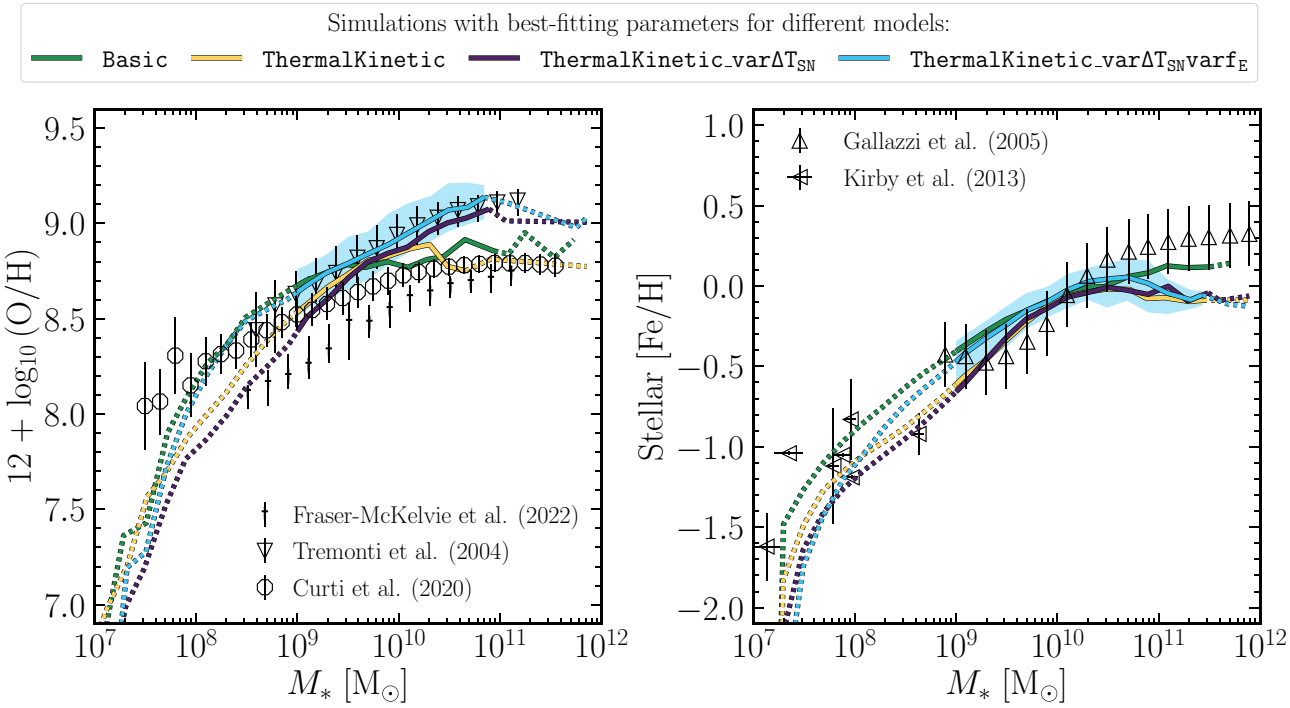

Figure 13. As Fig. 11, but showing the median gas-phase metallicity (*left*) and stellar metallicity (*right*) versus galaxy stellar mass at $z = 0$. The $z \approx 0$ comparison data are taken from A. Fraser-McKelvie et al. (2022), C. A. Tremonti et al. (2004), and M. Curti et al. (2020) for the gas-phase metallicity, and from A. Gallazzi et al. (2005) and E. N. Kirby et al. (2013) for the stellar metallicity. The gas-phase metallicity is computed in the gas that is sufficiently dense ($n_{\rm H} > 0.1\,{\rm cm}^{-3}$) and cool ($T < 10^{4.5}$ K) and only for star-forming galaxies (sSFR $> 10^{-2}\,{\rm Gyr}^{-1}$), excluding metals that are present in dust. All four models are consistent with the observations for both the gas-phase and stellar metallicities at $M_* \lesssim 10^{11}\,{\rm M}_\odot$.

neutral gas fractions at the high-mass end (e.g. A. R. H. Stevens et al. 2019).

5.3.5 *Metal content*

The left panel of Fig. 13 shows the relationship between the metallicity of the gas phase, in units of $12 + \log_{10}({\rm O/H})$, 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

$$\mathrm{O/H} = \frac{m_{\rm H}}{m_{\rm O}} \frac{\sum_i (X_{\rm O}/X_{\rm H})_i \, m_{{\rm gas},i}}{\sum_i m_{{\rm gas},i}}\,, \tag{19}$$

where $(X_{\rm O}/X_{\rm H})_i$ is the ratio of the oxygen and hydrogen mass fractions carried by gas particle i, $m_{{\rm gas},i}$ is the mass of gas particle i, and $m_{\rm H}/m_{\rm O}$ is the ratio of the masses of hydrogen and oxygen nuclei. To aid the comparison with observational data, in equation (19) we consider only those gas particles that are dense ($n_{\rm H} > 0.1\,{\rm cm}^{-3}$) and cool ($T < 10^{4.5}$ K). Furthermore, we apply a spatial mask by requiring the selected particles to be within 50 pkpc apertures centred on the galaxies. We do not include metals that are present in dust. In a given stellar mass bin, we show the median value of $12 + \log_{10}({\rm O/H})$ considering only star-forming galaxies (sSFR $> 10^{-2}\,{\rm Gyr}^{-1}$).

For comparison, we display the gas-phase metallicities of 472 star-forming galaxies at $0.04 < z < 0.128$ from the SAMI Galaxy Survey (A. Fraser-McKelvie et al. 2022), as well as metallicities for $\sim 10^5$ local star-forming SDSS galaxies from C. A. Tremonti et al. (2004) and M. Curti et al. (2020). The data from C. A. Tremonti et al. (2004) and A. Fraser-McKelvie et al. (2022) follow two distinct metallicity tracks, offset by ≈ 0.3 dex, while the measurements from M. Curti et al. (2020) gradually transition from the upper to the lower track with increasing stellar mass. The ≈ 0.3 dex systematic discrepancy between different observations arises from differences in the calibration of the methods used to infer gas-phase metallicities from galaxy spectra, as well as from differences in the methods themselves. In particular, the lower track typically corresponds to metallicities estimated using the so-called $T_{\rm e}$-method, while the upper track is based on calibrations using photoionization models (see e.g. Á. R. López-Sánchez et al. 2012; M. Curti et al. 2020).

We find that all four models are consistent with the observational data, including the normalization, slope, and scatter of the mass–metallicity relation. At stellar masses $M_* > 10^{9.5}\,{\rm M}_\odot$, `ThermalKinetic_varΔT`$_{SN}$ and `ThermalKinetic_varΔT`$_{SN}$`varf`$_{E}$ closely follow the upper track of the observations, whereas the metallicities in `Basic` and `ThermalKinetic` saturate at values corresponding to the lower track.

The right panel of Fig. 13 shows the relationship between galaxy stellar mass and 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 total numbers of iron and hydrogen nuclei, Fe/H, where we employ the same expression as for the gas-phase O/H (equation 19) but apply it to stellar particle-carried fields and replace oxygen with iron. Stellar particles that contribute to Fe/H are selected within 50 pkpc apertures. The resulting ratio is subsequently normalized by the solar value of Fe/H, assuming a

solar iron abundance of $12 + \log_{10}(\mathrm{Fe/H}) = 7.5$ (M. 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 A. Gallazzi et al. (2005) and dwarf irregular and spheroidal satellite galaxies of the Milky Way and M31 from E. N. Kirby et al. (2013). Where needed, the solar abundances used in the observations have been converted to the solar values reported by M. Asplund et al. (2009).

Overall, all four models are consistent with the observations within the stellar mass range of $\sim 10^7$ to $10^{11}\,\mathrm{M}_\odot$. At higher masses ($M_* \gtrsim 10^{11}\,\mathrm{M}_\odot$), the `ThermalKinetic`, `ThermalKinetic_varΔT_SN`, and `ThermalKinetic_varΔT_SNvarf_E` models undershoot the A. Gallazzi et al. (2005) data by up to ≈ 0.35 dex. This discrepancy is likely related to the fact that stellar metallicities in the simulations are derived from the Fe abundance, whereas the metallicities in A. Gallazzi et al. (2005) are estimated based on a combination of Mg and Fe absorption features in galaxy spectra, with the importance of Mg becoming higher in more massive galaxies due to $\alpha-$enhancement (e.g. M. C. Segers et al. 2016). J. Schaye et al. (2025) show that the agreement between COLIBRE and the A. Gallazzi et al. (2005) data at the high-mass end improves if stellar metallicities in the simulations are derived from Mg instead of Fe, although a small discrepancy remains.[21]

The stellar mass–metallicity relation in the `Basic` model saturates at a metallicity that is ≈ 0.25 dex higher than in the other three models. This results from the lower BH seed mass adopted in the `Basic` model, which weakens AGN feedback and consequently allows for more late-time star formation in massive galaxies, thereby increasing their present-day stellar metallicities.

5.4 The COLIBRE fiducial model at m7, m6, and m5 resolutions

In the previous sections, we showed that among the four models with the best-fitting parameter values identified by the emulators, `ThermalKinetic_varΔT_SNvarf_E` not only provides the closest match to the observed GSMF and SSMR, but is also in overall better agreement with the observational data that were not used in the calibration. If the COLIBRE model were designed solely for the resolution at which the emulators were employed (m7; $m_{\rm gas} \approx m_{\rm dm} \sim 10^7\,\mathrm{M}_\odot$), we could conclude our search for the best-fitting model here. However, since the COLIBRE suite also includes simulations at m6 ($m_{\rm gas} \approx m_{\rm dm} \sim 10^6\,\mathrm{M}_\odot$) and m5 ($m_{\rm gas} \approx m_{\rm dm} \sim 10^5\,\mathrm{M}_\odot$) resolutions, it is essential to verify that the `ThermalKinetic_varΔT_SNvarf_E` model performs well at these resolutions too before establishing it as the fiducial COLIBRE model.

[21] Another potential source of discrepancy with the A. Gallazzi et al. (2005) data concerns our fiducial choice of aperture size, within which all galaxy properties are calculated: 50 pkpc. In contrast, A. Gallazzi et al. (2005) derived stellar metallicities from SDSS spectra taken with 3 arcsec-diameter fibres, corresponding to a physical radius of ≈ 3 kpc at $z = 0.1$. We verified that reducing the aperture from 50 to 3 kpc increases the stellar metallicity at the high-mass end by no more than ≈ 0.1 dex, which is insufficient to account for the full offset with respect to the A. Gallazzi et al. (2005) measurements (≈ 0.35 dex).

5.4.1 Extension of the ThermalKinetic_varΔT$_{SN}$varf$_E$ model with best-fitting parameter values to higher resolutions

By applying the best-fitting `ThermalKinetic_varΔT_SNvarf_E` model calibrated at m7 resolution to m6 and m5 resolutions, without adjusting any parameter values except for gravitational softening (which scales with $m_{\rm gas}$ as $\varepsilon_{\rm soft} \propto m_{\rm gas}^{1/3}$), we found that the fit to the observed GSMF and SSMR at $z = 0$ becomes progressively worse at higher resolutions. The primary factor reducing the accuracy of the fit is the increasingly earlier onset of AGN feedback. This result could be anticipated, as the tail of the gas density distribution in higher-resolution simulations extends to higher values, which can significantly boost BH accretion rates (see equation 10). As a consequence, BH particles undergo faster growth and produce more aggressive AGN feedback, which ultimately results in galaxies with unrealistically low stellar masses and large sizes.

By manually adjusting the BH seed mass to suppress the overly efficient early BH growth and obtain a good fit to the observed GSMF and SSMR, we found that at m6 resolution the BH seed mass needs to be reduced by a factor of ~ 10 compared to its best-fitting value at m7 resolution, and by another factor of ~ 10 when increasing the resolution from m6 to m5, giving $m_{\rm BH,seed} \sim 10^3\,\mathrm{M}_\odot$ at m5 resolution. Because the BH accretion rate has a superlinear dependence on $m_{\rm BH}$ ($\dot{m}_{\rm BH} \propto m_{\rm BH}^2$), such a low value of $m_{\rm BH,seed}$ at m5 resolution yields two distinct populations of galaxies based on their stellar mass: (i) galaxies with $M_* \gtrsim 10^{10.5}\,\mathrm{M}_\odot$ that have efficient BH growth, and (ii) galaxies with $M_* \lesssim 10^{10.5}\,\mathrm{M}_\odot$ where BH accretion rates are very low. The transition from nearly no BH growth to efficient BH growth becomes a step-like function of M_*, which is undesirable. Fortunately, we found that a much smoother transition between the two regimes of BH growth, including an intermediate population of galaxies with moderately growing BHs, can be realised by adopting a much higher value of $m_{\rm BH,seed}$ alongside a *variable* AGN heating temperature, which depends linearly on the BH mass, instead of the constant value $\Delta T_{\rm AGN} = 10^9$ K.

To ensure that the COLIBRE model remains consistent across all resolutions, we decided to implement the change from $\Delta T_{\rm AGN} = 10^9$ K to a variable $\Delta T_{\rm AGN}$ not only at m5 resolution but also at m6 and m7. In the following text, we detail how the change in $\Delta T_{\rm AGN}$ is implemented and how it affects the calibration at m7 resolution described in the previous sections.

5.4.2 The COLIBRE model with a variable $\Delta T_{\rm AGN}$

We take the `ThermalKinetic_varΔT_SNvarf_E` model and modify its prescription for AGN feedback by replacing a fixed $\Delta T_{\rm AGN} = 10^9$ K with a function that depends on the BH (subgrid) mass,

$$\Delta T_{\rm AGN}(m_{\rm BH}) = \Delta T_{\rm AGN,pivot} \left(\frac{m_{\rm BH}}{m_{\rm BH,pivot}} \right), \tag{20}$$

where $\Delta T_{\rm AGN,pivot}$ and $m_{\rm BH,pivot}$ are (degenerate) parameters. The variable heating temperature is constrained within the bounds $\Delta T_{\rm AGN,min} < \Delta T_{\rm AGN}(m_{\rm BH}) < \Delta T_{\rm AGN,max}$, which at m7 resolution we set to $\Delta T_{\rm AGN,min} = 10^{6.5}$ K and $\Delta T_{\rm AGN,max} = 10^{9.5}$ K. These limits ensure that AGN feedback remains both efficient and well-sampled across a wide range of halo masses, following the same rationale used to set the variable heating temperature in SN feedback (see Section 2.2.5). The two remaining parameters, which are degenerate ($\Delta T_{\rm AGN,pivot} \propto m_{\rm BH,pivot}$), are set to

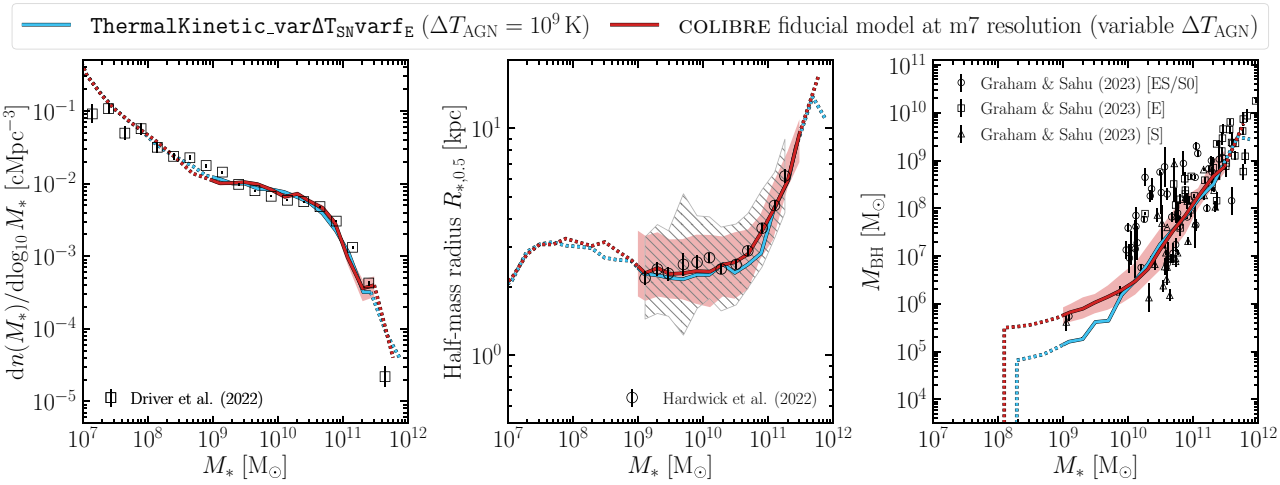

Figure 14. The GSMF (*left*), the median SSMR (*middle*), and the median BSMR (*right*), all shown at $z = 0$. The light-blue and dark-red solid curves are simulation predictions with the best-fitting ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model (using $\Delta T_{\rm AGN} = 10^9$ K) and its modified version with the variable $\Delta T_{\rm AGN}$ (i.e. the fiducial COLIBRE model), respectively. Both simulations were run in a (50 cMpc)^3 volume at m7 resolution. The black symbols and grey hatched region indicate observational data. For reference, the uncertainty in the predictions of the COLIBRE fiducial model is indicated by the dark-red shaded region, with boundaries corresponding to the Poisson uncertainty for the GSMF and the 16$^{\rm th}$ to 84$^{\rm th}$ percentile range for the SSMR and BSMR. While the GSMF and SSMR predicted by both models are nearly identical, each reproducing the observational data, the fiducial COLIBRE model predicts significantly higher BH masses in galaxies with $M_* \lesssim 10^{10}$ M$_\odot$.

$\Delta T_{\rm AGN,pivot} = 10^9$ K and $m_{\rm BH,pivot} = 10^8$ M$_\odot$, such that a heating temperature of 10^9 K is reached for a BH with $m_{\rm BH} = 10^8$ M$_\odot$, the value that is typical for $M_* \sim 10^{11}$ M$_\odot$ galaxies (see Fig. 9). For BHs with masses below 10^8 M$_\odot$, the heating temperature is less than 10^9 K, leading to a better sampling of AGN feedback events compared to the case where $\Delta T_{\rm AGN}$ is fixed to 10^9 K. More generally, since the BH accretion rate scales as $\dot{m}_{\rm BH} \propto m_{\rm BH}^2$, while the accretion rate at the Eddington limit scales linearly with $m_{\rm BH}$, equation (20) ensures that the sampling of AGN feedback events is independent of the BH mass for BHs accreting at a fixed Eddington fraction.

Since the change from $\Delta T_{\rm AGN} = 10^9$ K to a variable $\Delta T_{\rm AGN}$ will impact the strength of AGN feedback, some of the best-fitting parameter values found for the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model with $\Delta T_{\rm AGN} = 10^9$ K may require adjustments to achieve similar goodness of fit to the observed GSMF and SSMR. As an initial test, we ran several simulations in a (50 cMpc)^3 volume using the variation of the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model with the variable $\Delta T_{\rm AGN}$ given by equation (20), keeping the three SN parameters at their best-fitting values from Table 2 found for $\Delta T_{\rm AGN} = 10^9$ K while exploring different values of $m_{\rm BH,seed}$. These tests showed that the SN parameters can indeed remain unchanged, while $m_{\rm BH,seed}$ needs to be increased. Specifically, we found that if $m_{\rm BH,seed}$ is increased from its best-fitting value for $\Delta T_{\rm AGN} = 10^9$ K of $10^{4.8}$ M$_\odot$ to $10^{5.5}$ M$_\odot$, the ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model with the variable $\Delta T_{\rm AGN}$ matches the observed GSMF and SSMR with a similar accuracy as the best-fitting ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model with $\Delta T_{\rm AGN} = 10^9$ K.

Fig. 14 shows the $z = 0$ GSMF (left panel), the $z = 0$ SSMR (middle panel) and the $z = 0$ BSMR (right panel) for the best-fitting ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model with $\Delta T_{\rm AGN} = 10^9$ K (light-blue) and its modified version with the variable $\Delta T_{\rm AGN}$ (dark-red). The only differences between the models are the treatment of $\Delta T_{\rm AGN}$ and the value of $m_{\rm BH,seed}$, which is set to $10^{4.8}$ M$_\odot$ for $\Delta T_{\rm AGN} = 10^9$ K and to $10^{5.5}$ M$_\odot$ for the variable $\Delta T_{\rm AGN}$.

The GSMF and SSMR predicted by both models are virtually indistinguishable, each providing an excellent match to the S. P. Driver et al. (2022) GSMF and the J. A. Hardwick et al. (2022) SSMR. However, the BSMRs show notable differences. While the predictions from both models converge for $M_* \gtrsim 10^{10}$ M$_\odot$, reproducing the measurements of A. W. Graham & N. Sahu (2023), the variable $\Delta T_{\rm AGN}$ model exhibits systematically more massive BHs at $M_* \lesssim 10^{10}$ M$_\odot$ – by up to ≈ 0.7 dex–due to its higher BH seed mass. As discussed in Section 5.4.1, the ability to increase $m_{\rm BH,seed}$ without compromising the fit to the observed GSMF and SSMR becomes crucial at m5 resolution, and is the main reason why the variable $\Delta T_{\rm AGN}$ model is preferred over its fixed $\Delta T_{\rm AGN}$ counterpart.

Fig. 15 presents a comparison between the predictions of the best-fitting ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$ model (light-blue) and its variation with the variable $\Delta T_{\rm AGN}$ (dark-red) for nine different relations that were not used to calibrate the models. These relations were previously shown in Section 5.3 in the context of comparing the four best-fitting models employing different SN feedback prescriptions. In Fig. 15, the panels, from left to right and top to bottom, show: the median sSFR versus stellar mass at $z = 0$ for active galaxies (sSFR $> 10^{-2}$ Gyr^{-1}), quenched fraction versus stellar mass at $z = 0$, gas metallicity versus stellar mass at $z = 0$ for star-forming galaxies, stellar metallicity versus stellar mass at $z = 0$, H I mass-to-stellar mass fraction versus stellar mass at $z = 0$, H$_2$ mass-to-stellar mass fraction versus stellar mass at $z = 0$, GSMF at $z = 2$, the median sSFR of all galaxies versus stellar mass at $z = 2$, and the cosmic SFRD versus redshift.

On average, the models with fixed and variable $\Delta T_{\rm AGN}$ perform similarly well in reproducing observational data at both $z = 0$ and $z = 2$. Small differences emerge in the $z = 0$ sSFR, H I fractions, H$_2$ fractions, and quenched fractions at $M_* \lesssim 10^{10.5}$ M$_\odot$, where the variable $\Delta T_{\rm AGN}$ model exhibits slightly higher gas fractions and SFRs and slightly lower quenched fractions,

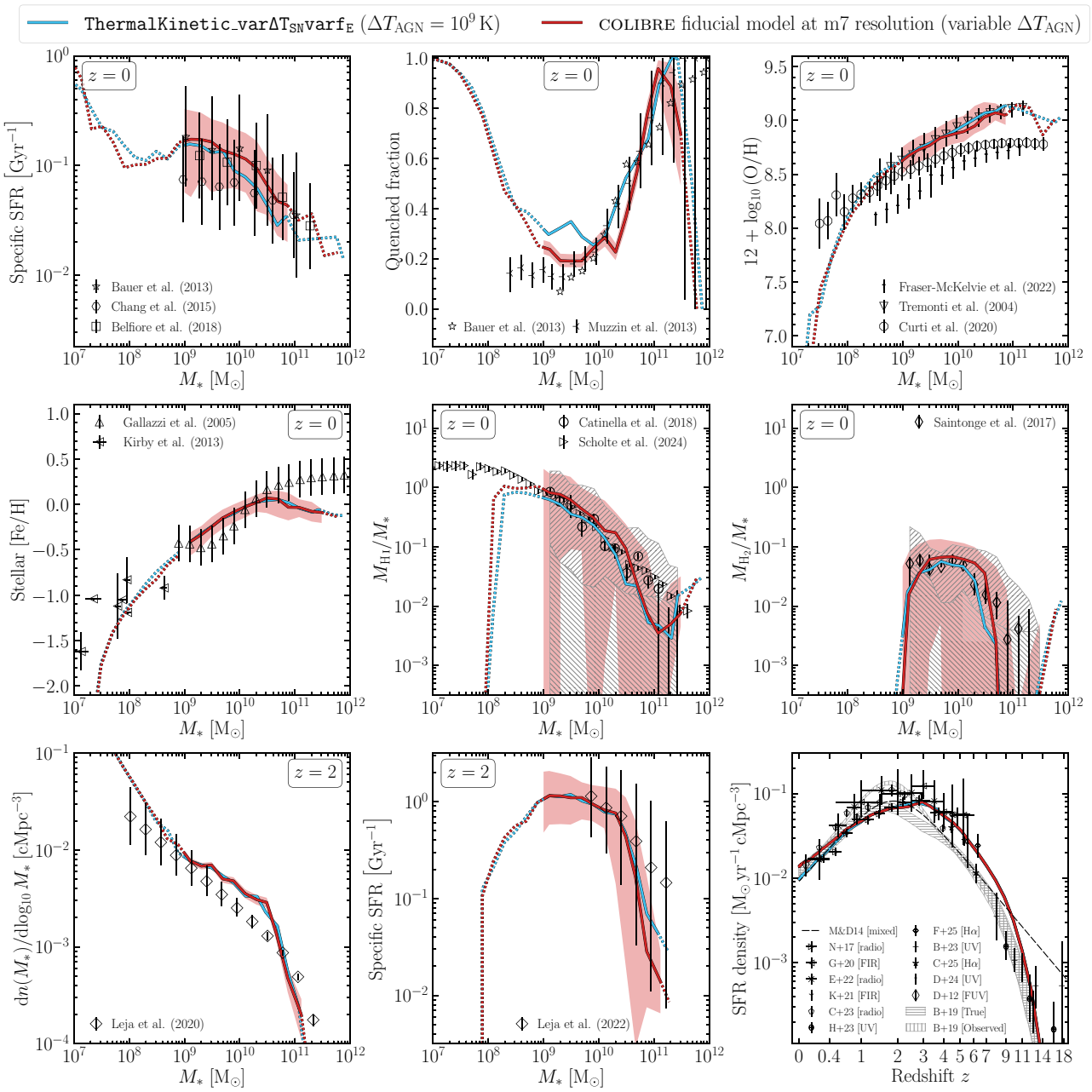

Figure 15. Comparison of the simulation predictions from the best-fitting `ThermalKinetic_varΔT_SNvarf_E` model (with $\Delta T_{\rm AGN} = 10^9$ K; light-blue) and its modified version with the variable $\Delta T_{\rm AGN}$ (i.e. the fiducial COLIBRE model; dark-red) for various relations that were not used to calibrate the models. Both simulations were run in a $(50\ {\rm cMpc})^3$ volume at m7 resolution. The panels, from left to right, top to bottom, display: the median sSFR versus stellar mass for active galaxies (sSFR $> 10^{-2}\,{\rm Gyr}^{-1}$) at $z = 0$, quenched fraction versus stellar mass at $z = 0$, the gas-phase metallicity versus stellar mass for star-forming galaxies at $z = 0$, the stellar metallicity versus stellar mass at $z = 0$, the median H I mass-to-stellar mass fraction versus stellar mass at $z = 0$, the median H_2 mass-to-stellar mass fraction versus stellar mass at $z = 0$, the GSMF at $z = 2$, the median sSFR versus stellar mass for all galaxies at $z = 2$, and the cosmic SFRD versus redshift. Where present, the dark-red shaded region indicates the 16$^{\rm th}$ to 84$^{\rm th}$ percentiles in the COLIBRE fiducial model, except for the quenched fraction, where it represents the 1σ confidence interval. The `ThermalKinetic_varΔT_SNvarf_E` and COLIBRE fiducial models perform similarly well in reproducing the observational data.

leading to marginal or no improvement in the agreement with the data. Overall, Fig. 15 confirms that switching from the best-fitting `ThermalKinetic_varΔT_SNvarf_E` model with fixed $\Delta T_{\rm AGN}$ to its variation with variable $\Delta T_{\rm AGN}$ does not degrade the agreement with observations in any of the explored relations.

Based on the above findings, we opted to establish the `ThermalKinetic_varΔT_SNvarf_E` model with the variable $\Delta T_{\rm AGN}$ as the *fiducial* COLIBRE model at all resolutions. At m7 resolution, the fiducial parameter values are therefore $m_{\rm BH,seed} = 10^{5.5}\,{\rm M}_\odot$, $f_{\rm kin} = 0.1$, $P_{\rm E,pivot}/k_{\rm B} = 8 \times 10^3\,{\rm K\,cm^{-3}}$, and $n_{\rm H,pivot} = 0.6\,{\rm cm^{-3}}$. For m6 and m5 resolutions, several subgrid parameter values are adjusted from their m7 values to improve the fit to the observed $z = 0$ GSMF and SSMR. Specifically, $m_{\rm BH,seed}$ is decreased from $m_{\rm BH,seed} = 10^{5.5}\,{\rm M}_\odot$ to $3 \times 10^4\,{\rm M}_\odot$, and $2 \times 10^4\,{\rm M}_\odot$,

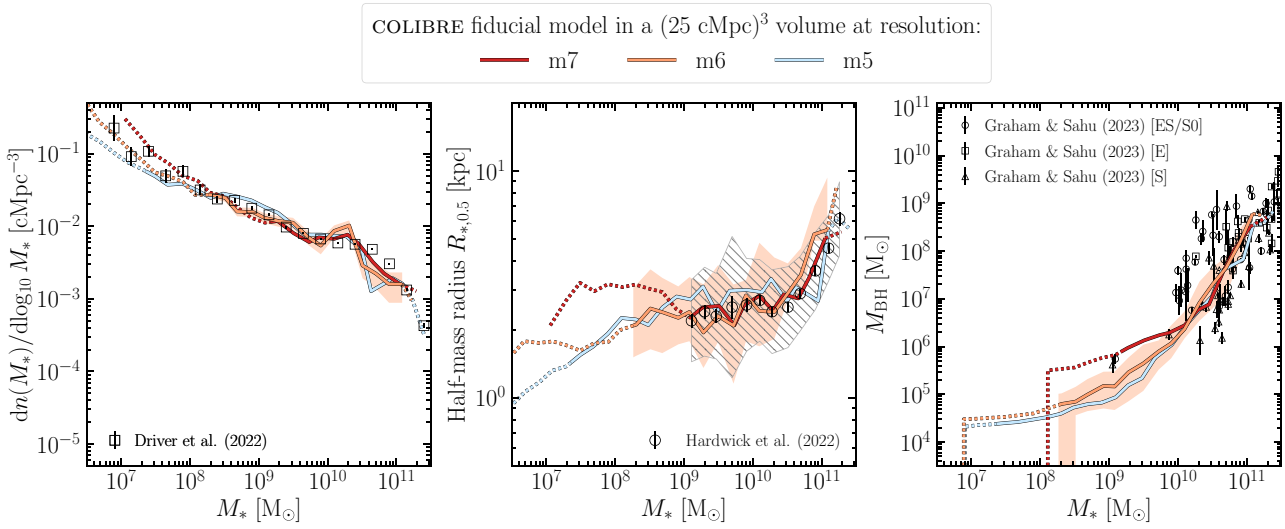

Figure 16. The $z = 0$ GSMF (*left*), the $z = 0$ median SSMR (*middle*), and the $z = 0$ median BSMR (*right*) in the simulations with the fiducial COLIBRE models (i.e. the best-fitting COLIBRE models with variable $\Delta T_{\rm AGN}$) at m7, m6, and m5 resolutions. All simulations were run in a (25 cMpc)^3 cosmological volume. Solid curves are converted to dotted curves below a stellar mass of 100 × the baryonic particle mass of the simulations, to indicate that the corresponding galaxies are poorly resolved, as well as at the high-mass end, where there are fewer than five galaxies per bin. The GSMF and SSMR predicted by COLIBRE exhibit very good convergence with resolution for $M_* \gtrsim 10^9$ M$_\odot$ and reproduce the observed GSMF from S. P. Driver et al. (2022) and SSMR from J. A. Hardwick et al. (2022). The COLIBRE BSMR agrees with the observational measurements from A. W. Graham & N. Sahu (2023) but converges only for $M_* \gtrsim 10^{10}$ M$_\odot$ due to the different BH seed mass values used at different resolutions.

respectively, while $P_{\rm E,pivot}/k_{\rm B}$ is increased from $P_{\rm E,pivot}/k_{\rm B} = 8 \times 10^3$ K cm^{-3} to 1×10^4 K cm^{-3} and 1.5×10^4 K cm^{-3}. The parameter $f_{\rm kin}$ remains fixed at 0.1 for all resolutions, while $n_{\rm H,pivot}$ is decreased from 0.6 cm^{-3} to 0.5 cm^{-3} at m6 resolution but increases to 1.0 cm^{-3} at m5 resolution. Among the SN and AGN parameters not used in the emulation, $\Delta T_{\rm SN,min}$, $\Delta T_{\rm SN,max}$, $\Delta T_{\rm AGN,max}$, and $f_{\rm E,min}$ are set to $10^{6.75}$ K, 10^8 K, 10^{10} K, and 0.3 at m6 resolution, and to 10^7 K, 10^8 K, 10^{10} K, and 0.8 at m5 resolution, respectively. Finally, the AGN feedback coupling efficiency, $\varepsilon_{\rm f}$, was reduced from 0.1 at m7 resolution to 0.05 at m6 and m5 resolutions to improve the agreement with the observed $z = 0$ BSMR. Without this adjustment, the normalization of the BSMR predicted by the m6 and m5 models would have been too high by a factor of ≈ 2 relative to the observations. A complete list of resolution-dependent subgrid parameter values is provided in table 1 of J. Schaye et al. (2025).

The adjustments to the SN and AGN parameter values at m6 and m5 resolutions were determined through an iterative trial-and-error process, involving ~ 100 simulations in 25^3 and 12.5^3 cMpc^3 volumes, respectively. Our initial guess assumed the same parameter values as in the fiducial m7 model, and we then refined them iteratively to arrive at the final values reported above. The resolution dependence of most parameters is physically intuitive. For example, at higher resolution, BHs start growing earlier due to the presence of denser gas, requiring a lower $m_{\rm BH,seed}$ to counterbalance the enhanced BH growth. Conversely, thermal SN feedback becomes less efficient due to stronger radiative cooling losses in the denser gas and because a fixed temperature increase $\Delta T_{\rm SN}$ of a single gas particle corresponds to a smaller increase in the particle's internal energy, necessitating an increase in $\Delta T_{\rm SN,min}$ and $\Delta T_{\rm SN,max}$ at higher resolutions to maintain feedback strength comparable to m7 resolution. Star formation at higher resolution occurs in gas environments with higher pressures, so $P_{\rm E,pivot}$, which is representative of the median stellar birth pressure in the simulation (see Appendix B), must increase. Finally, we note that the non-monotonic dependence of $n_{\rm H,pivot}$ on the resolution is due to its strong degeneracy with $P_{\rm E,pivot}$ and $\Delta T_{\rm SN,min}$, both of which change with resolution.

Fig. 16 shows the $z = 0$ GSMF (left panel), the $z = 0$ SSMR (middle panel) and the $z = 0$ BSMR (right panel) for the fiducial COLIBRE models at m7, m6 and m5 resolutions (differently coloured solid curves), all in a (25 cMpc)^3 cosmological volume. The observed GSMF from S. P. Driver et al. (2022), the SSMR from J. A. Hardwick et al. (2022), and the BSMR from A. W. Graham & N. Sahu (2023) are shown as black symbols.

The COLIBRE simulations reproduce the observed GSMF and SSMR not only at m7 resolution, but also at m6 and m5, despite the best-fitting parameter values for m6 and m5 having been determined manually rather than via emulation. In particular, the SSMR predicted by COLIBRE at m6 and m5 resolutions remains within ≈ 0.1 dex of the observed sizes across the full stellar mass range. In the range $10^9 \lesssim M_*/{\rm M}_\odot \lesssim 10^{10.5}$, the SSMR at m5 resolution appears slightly elevated compared to m6 and m7, but remains within ≈ 0.1 dex. At all three resolutions, the GSMF agrees with the observational data to within ≈ 0.1 dex at $M_* \lesssim 10^{10.5}$ M$_\odot$, while at $M_* \gtrsim 10^{10.5}$ M$_\odot$, the predictions are systematically offset from the observed GSMF, owing to the small number of massive galaxies in the (25 cMpc)^3 volume. This is, however, not a concern, as we have verified that increasing the simulated volume brings the GSMF predictions into much closer agreement with the data at the high-mass end (see Appendix A). The agreement between the COLIBRE predictions for the BSMR and the observational measurements of A. W. Graham & N. Sahu (2023) is excellent at all resolutions, matching both the normalization and the slope of the observed relation.

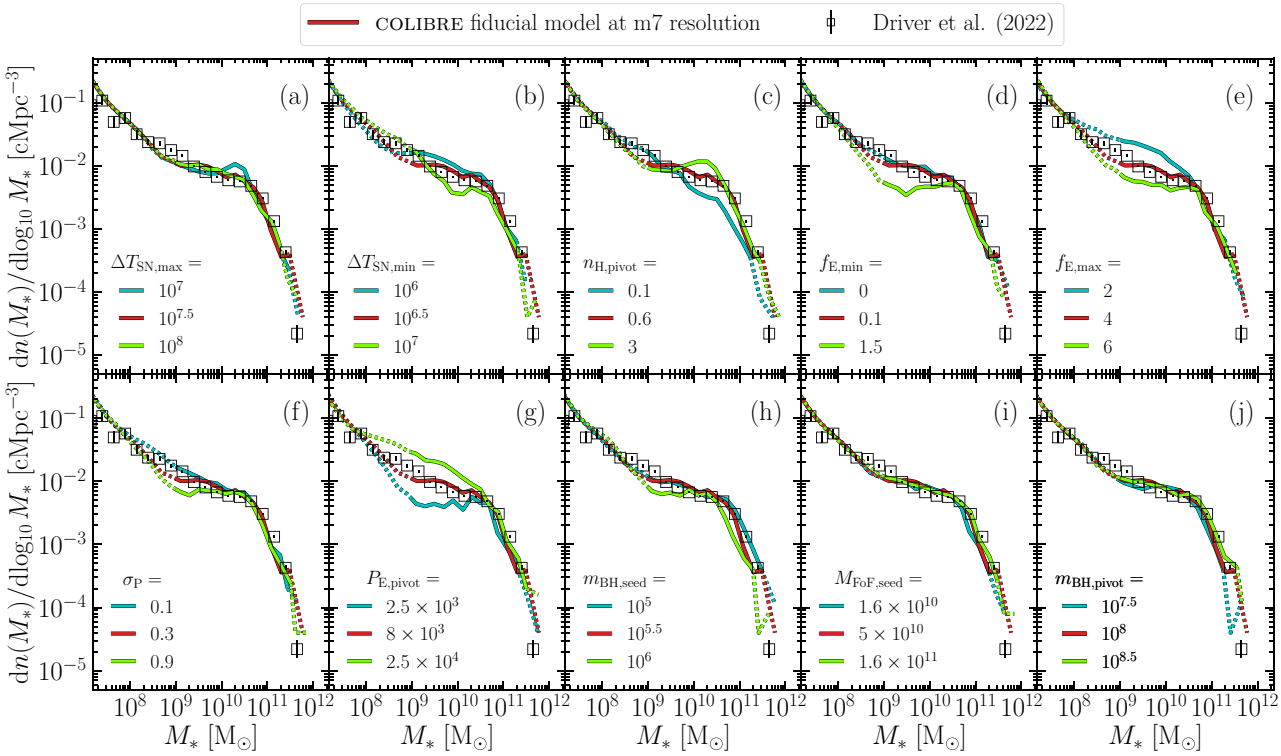

Figure 17. The GSMF at redshift $z = 0$. Different panels show the effect of varying different SN and AGN feedback-related subgrid parameters of the COLIBRE fiducial model at m7 resolution. For each variation, we run a separate $(50\,\mathrm{cMpc})^3$ volume simulation. Only one parameter is varied per panel, while the others are held fixed to their best-fitting values. Starting from the top-left panel and going left to right, top to bottom, the varied parameters are $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,min}$, $n_{\rm H,pivot}$, $f_{\rm E,min}$, $f_{\rm E,max}$, $\sigma_{\rm P}$, $P_{\rm E,pivot}$, $m_{\rm BH,seed}$, $M_{\rm FoF,seed}$, and $m_{\rm BH,pivot}$. The light-green (cyan) curve corresponds to the higher (lower) value of the parameter, relative to that in the fiducial m7 model, which is shown in dark-red. The comparison data from S. P. Driver et al. (2022) are displayed as black squares. Many parameters are degenerate with one another (i.e. they have a similar impact on the GSMF), while others have little effect on the GSMF.

The COLIBRE GSMF, SSMR, and BSMR exhibit very good convergence[22] with resolution for $M_* \gtrsim 10^8\,\mathrm{M}_\odot$, $M_* \gtrsim 10^9\,\mathrm{M}_\odot$, and $M_* \gtrsim 10^{10}\,\mathrm{M}_\odot$, respectively. The GSMF converges down to a lower M_* than the SSMR because the stellar mass is an integrated property, while the sizes depend on the spatial distribution of stellar particles within the galaxy and hence require more particles to be converged. At $M_* \lesssim 10^9\,\mathrm{M}_\odot$, the limit corresponding to $\lesssim 100$ stellar particles at m7 resolution, galaxy half-mass radii decrease with increasing resolution. Meanwhile, the BSMR converges only for $M_* \gtrsim 10^{10}\,\mathrm{M}_\odot$ due to differences in the BH seed mass used at different resolutions. Thanks to the variable $\Delta T_{\rm AGN}$, the lowest BH masses in the m5 model are $2 \times 10^4\,\mathrm{M}_\odot$. The cut-off in the BSMR at $M_* \sim 10^7 - 10^8\,\mathrm{M}_\odot$ is set by the minimum FoF mass, $M_{\rm FoF,seed}$, required for haloes to be seeded with a BH particle. $M_{\rm FoF,seed}$ is equal to $10^{10}\,\mathrm{M}_\odot$ at m5 and m6 resolutions but to $5 \times 10^{10}\,\mathrm{M}_\odot$ at m7 resolution,[23] shifting the cutoff of the BSMR to higher M_* in the m7 model compared to m6 and m5.

[22] As discussed earlier in Section 5.4.2, some of the parameter values in the fiducial COLIBRE model are adjusted between different resolutions. Therefore, the resolution convergence should be interpreted as 'weak convergence', in the language of J. Schaye et al. (2015).

[23] Setting $M_{\rm FoF,seed}$ to $10^{10}\,\mathrm{M}_\odot$ at m7 resolution, which is the value used at m6 and m5 resolutions, would result in massive BH particles having gravitational masses that exceed their subgrid masses, which is undesirable (see section 3.8.1 of J. Schaye et al. 2025 for a discussion).

5.5 Parameter variations in the fiducial m7 COLIBRE model

In the previous section, we showed that the fiducial COLIBRE models at m7, m6, and m5 resolutions reproduce the observational data to which they were calibrated: the $z = 0$ GSMF and SSMR using Gaussian process emulators and, independently, the $z = 0$ BSMR by manually adjusting the coupling efficiency $\varepsilon_{\rm f}$ in equation (12). Because the m7 model was calibrated to the GSMF and SSMR using emulators by optimizing the best-fitting values of a set of *four* subgrid parameters, $\boldsymbol{\theta} = (f_{\rm kin}, n_{\rm H,pivot}, P_{\rm E,pivot}, m_{\rm BH,seed})$, an important question arises: why were these four parameters optimized, while other subgrid parameters related to stellar and AGN feedback were held fixed during emulation? This question was discussed in Section 4.2, though without presenting results supporting the choice of subgrid parameters. In this section, we further address this question by showing that the feedback-related subgrid parameters excluded from emulator-based calibration either have a negligible impact on the calibrated galaxy properties or are degenerate with parameters already included in $\boldsymbol{\theta}$.

Figs 17 and 18 show how the $z = 0$ GSMF and SSMR in the COLIBRE fiducial m7 model respond to variations in $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,min}$, $n_{\rm H,pivot}$, $f_{\rm E,min}$, $f_{\rm E,max}$, $\sigma_{\rm P}$, $P_{\rm E,pivot}$, $m_{\rm BH,seed}$, $M_{\rm FoF,seed}$, and $m_{\rm BH,pivot}$. 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

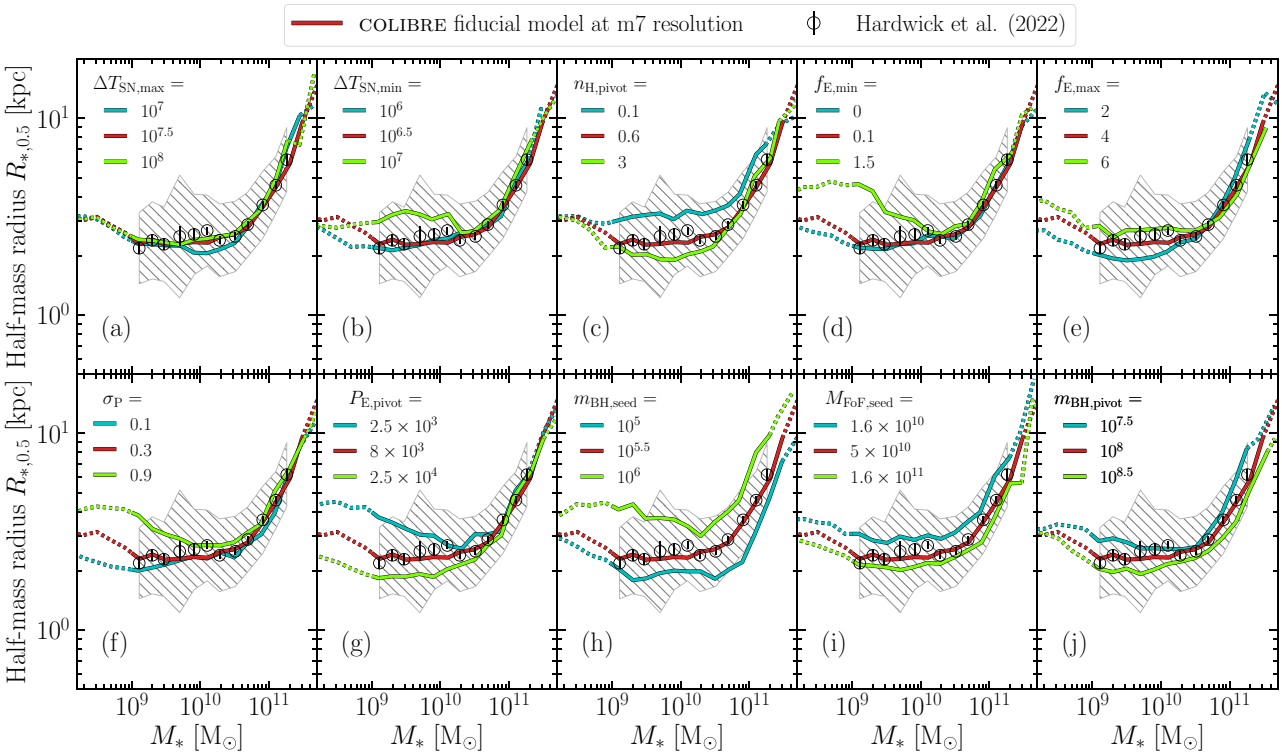

Figure 18. As Fig. 17, but showing the $z = 0$ median galaxy SSMR. The median observed relation from J. A. Hardwick et al. (2022) is displayed as black circles. The error bars show the 1σ error on the median observed sizes, while the grey hatched region indicates the galaxy population-wide scatter in J. A. Hardwick et al. (2022).

separate $(50\,\mathrm{cMpc})^3$ volume simulation (that is, what is shown in the figures are the results from simulations, not from emulators). We show the results for three different values of each parameter: the value from the fiducial model (dark-red), a lower value (cyan), and a higher value (light-green). The values of the varied parameters are indicated in each panel's legend.

First, we recall that the density-dependent heating temperature in the SN thermal feedback of the COLIBRE fiducial model depends on the subgrid parameters $\Delta T_{\rm SN,max}$, $\Delta T_{\rm SN,min}$, and $n_{\rm H,pivot}$ (see equation 8). The effects of varying these parameters are shown in panels (a), (b), and (c), respectively. We observe that increasing $\Delta T_{\rm SN,max}$ by 0.5 dex has no impact on either the GSMF or galaxy sizes. In contrast, decreasing $\Delta T_{\rm SN,max}$ dex by 0.5 dex introduces a bump in the GSMF and a dip in galaxy sizes around $M_* = 10^{10.5}\,\mathrm{M}_\odot$, indicating inefficient SN feedback at the mass scale just below where BHs take over. Similarly, changing $\Delta T_{\rm SN,min}$ by ±0.5 dex has a pronounced effect on both the GSMF and SSMR. As expected, a higher (lower) $\Delta T_{\rm SN,min}$ enhances (weakens) SN thermal feedback, 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 SSMR up (down). By comparing panels (a) and (b) with panel (c), where $n_{\rm H,pivot}$ is varied, we observe that most effects on the GSMF and SSMR caused by changes in $\Delta T_{\rm SN,min}$ and/or $\Delta T_{\rm SN,max}$ can be captured by adjusting $n_{\rm H,pivot}$ alone. Furthermore, we recall that $\Delta T_{\rm SN,min}$ cannot be much lower than $10^{6.5}$ K (otherwise the SN thermal feedback would 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 would become poor, see Section 2.2.5). 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 (at m7 resolution) 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 $P_{\rm birth}$ (equation 2). The parameters are: $f_{\rm E,min}$, $f_{\rm E,max}$, $\sigma_{\rm P}$ and $P_{\rm E,pivot}$. First, we observe that varying the parameters $f_{\rm E,min}$ and $f_{\rm E,max}$ – which specify the energy injected by SN feedback from stellar particles formed in low- and high-pressure gas environments, respectively – modulates the overall strength of SN feedback, predominantly affecting galaxies with $M_* \lesssim 10^{10.5}\,\mathrm{M}_\odot$. Decreasing (increasing) the SN energy results in more (less) stellar mass formed and more (less) compact galaxies. Next, a nearly order-of-magnitude variation in the parameter $\sigma_{\rm P}$, which controls the width of the transition from $f_{\rm E,min}$ at low $P_{\rm birth}$ to $f_{\rm E,max}$ at high $P_{\rm birth}$, primarily impacts the low-mass end of the GSMF and SSMR ($M_* \lesssim 10^{10}\,\mathrm{M}_\odot$) in a manner similar to $f_{\rm E,min}$. Finally, the combined effect on the GSMF and SSMR of varying $f_{\rm E,min}$, $f_{\rm E,max}$, and $\sigma_{\rm P}$ can be well captured by solely changing $P_{\rm E,pivot}$ (compare panels d, e, and f versus panel g), thereby justifying our choice to optimize $P_{\rm E,pivot}$ while keeping $f_{\rm E,min}$, $f_{\rm E,max}$, and $\sigma_{\rm P}$ fixed. Since lower values of $f_{\rm E,min}$ and $\sigma_{\rm P}$ yield a slightly better SSMR at low stellar masses (see panels d and f in Fig. 18), but $\sigma_{\rm P}$ must also not be too small to avoid an excessively sharp transition from $f_{\rm E,min}$ to $f_{\rm E,max}$, we set $f_{\rm E,min}$ to 0.1 and $\sigma_{\rm P}$ to 0.3. As for $f_{\rm E,max}$, we set its value to 4, which gives an average energy per SN within the range $(1.5 - 2) \times 10^{51}$ erg. Alternatively, we could choose a somewhat different value for $f_{\rm E,max}$ by adjusting $P_{\rm E,pivot}$ due to the degeneracy between these two parameters (compare panel e vs. panel g).

Lastly, the remaining three panels – (h), (i), and (j) – illustrate the effect of three BH-related parameters: $m_{\rm BH,seed}$, $M_{\rm FoF,seed}$, and $m_{\rm BH,pivot}$, respectively. Physically, increasing $m_{\rm BH,seed}$ promotes BH growth through gas accretion and mergers, leading to stronger AGN feedback, while decreasing $m_{\rm BH,seed}$ suppresses BH growth, resulting in weaker AGN feedback. Changing $M_{\rm FoF,seed}$ in the same direction as $m_{\rm BH,seed}$ has the opposite effect: larger values of $M_{\rm FoF,seed}$ delay BH seeding and reduce the overall number of seeded BHs, resulting in slower BH growth and weaker AGN feedback, while smaller values increase the rate of BH seeding, leading to faster BH growth and stronger AGN feedback. Finally, increasing (decreasing) the normalization of the relation between the BH mass and $\Delta T_{\rm AGN}$, $m_{\rm BH,pivot}$, leads to lower (higher) $\Delta T_{\rm AGN}$ at fixed BH mass, producing less (more) bursty – and less (more) efficient – AGN feedback.[24] Because our fiducial value for the BH seed mass is relatively large ($m_{\rm BH,seed} = 10^{5.5}\,{\rm M}_\odot$), varying these BH-related parameters affects all galaxies whose haloes are massive enough to have been endowed with a BH particle ($M_* \gtrsim 10^{8.5}\,{\rm M}_\odot$). Galaxies experiencing stronger (weaker) AGN feedback on average form fewer (more) stars, and their stellar half-mass radii are larger (smaller).

Overall, comparing panel (h) to panels (i) and (j), we find that varying $m_{\rm BH,seed}$ has a qualitatively similar effect on the GSMF and SSMR as changing $M_{\rm FoF,seed}$ or $m_{\rm BH,pivot}$ in the opposite direction. This indicates that these three parameters are degenerate, meaning that optimizing one is sufficient. However, the GSMF and SSMR are quantitatively more sensitive to variations in $m_{\rm BH,seed}$ than in $M_{\rm FoF,seed}$ or $m_{\rm BH,pivot}$. We remind that we chose to optimize $m_{\rm BH,seed}$ while fixing $M_{\rm FoF,seed}$ at $5 \times 10^{10}\,{\rm M}_\odot$ and $m_{\rm BH,pivot}$ at $10^8\,{\rm M}_\odot$ in the COLIBRE fiducial m7 model; or equivalently, fixing $M_{\rm FoF,seed}$ at $5 \times 10^{10}\,{\rm M}_\odot$ and $\Delta T_{\rm AGN}$ at 10^9 K in the $\texttt{ThermalKinetic\_var}\Delta\texttt{T}_{\rm SN}\texttt{var}f_{\rm E}$ model.

Lastly, we note that $m_{\rm BH,pivot}$ cannot be significantly smaller than $10^8\,{\rm M}_\odot$, as this would correspond to $\Delta T_{\rm AGN} \ll 10^9$ K in massive objects according to equation (20), leading to implausibly high gas fractions therein, which are known to be sensitive to $\Delta T_{\rm AGN}$ (e.g. A. M. C. Le Brun et al. 2014; I. G. McCarthy et al. 2017; R. Kugel et al. 2023). Although we did not explicitly calibrate the models to cluster gas fractions, a few test simulations in a $(100\,{\rm cMpc})^3$ volume confirmed that $\Delta T_{\rm AGN} = 10^9$ K in the $\texttt{ThermalKinetic\_var}\Delta\texttt{T}_{\rm SN}\texttt{var}f_{\rm E}$ model and $m_{\rm BH,pivot} = 10^8\,{\rm M}_\odot$ in the COLIBRE fiducial model yield plausible gas fractions in galaxy clusters. The analysis of the properties of galaxy clusters will be presented in future work.

6 CONCLUSIONS

We have presented the calibration of the new COLIBRE subgrid model for cosmological hydrodynamical simulations of galaxy formation (J. Schaye et al. 2025). COLIBRE is available at three resolutions: $m_{\rm gas} \approx m_{\rm dm} \sim 10^7\,{\rm M}_\odot$ (m7), $10^6\,{\rm M}_\odot$ (m6), and $10^5\,{\rm M}_\odot$ (m5). It has evolved from the OWLS (J. Schaye et al. 2010) and EAGLE (J. Schaye et al. 2015) galaxy formation models 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 energy transfer from DM to baryons (by using four times more DM particles than baryonic particles); (iii) a model for the formation and evolution of dust grains 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 all subgrid physics processes, including the prescriptions for radiative cooling, star formation, stellar mass loss, BHs, and feedback from stars and SMBHs.

We used Gaussian process emulators to calibrate SN and AGN feedback in the COLIBRE model. The emulators were trained on ~ 200 simulations at m7 resolution in a $(50\,{\rm cMpc})^3$ volume. Each simulation was run with a unique combination of subgrid parameter values governing the strengths of SN and AGN feedback, enabling the emulators to learn how galaxy properties vary as functions of these parameters. These four 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, $n_{\rm H,pivot}$; (iii) the pivot stellar birth pressure in the relation between the SN energy and stellar birth pressure, $P_{\rm E,pivot}$; and (iv) the BH seed mass, $m_{\rm BH,seed}$. By fitting the trained emulators to the observed $z = 0$ GSMF and to the observed $z = 0$ galaxy SSMR in the stellar mass range $10^9 < M_*/{\rm M}_\odot < 10^{11.3}$, we found the values of the subgrid parameters that result in the best agreement with the target observational data.

The emulator-based calibration used a fixed $\Delta T_{\rm AGN} = 10^9$ K. After the emulation was completed, we updated the model to use an AGN heating temperature that increases linearly with BH mass (equation 20), as this proved important for higher resolution simulations. Paired with an increase in the BH seed mass, this change has a negligible effect on any of the calibration diagnostics. In the following, we first summarize our conclusions concerning the calibration with emulators at m7 resolution, using the COLIBRE model with $\Delta T_{\rm AGN} = 10^9$ K (referred to as $\texttt{ThermalKinetic\_var}\Delta\texttt{T}_{\rm SN}\texttt{var}f_{\rm E}$), and then discuss the transition to the fiducial COLIBRE model, which uses a variable $\Delta T_{\rm AGN}$.

In the prescription for SN feedback in the COLIBRE model, (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 (equation 8), and (iii) the energy per SN in units of 10^{51} erg, $f_{\rm E}$, is an increasing function of stellar birth gas pressure, $P_{\rm birth}$ (equation 2). In order to demonstrate that these model ingredients are all necessary to reproduce the observed GSMF and SSMR, 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 (i.e. $f_{\rm E}$ is independent of $P_{\rm birth}$) 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 a target 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, $\texttt{ThermalKinetic\_var}\Delta\texttt{T}_{\rm SN}$, the heating temperature $\Delta T_{\rm SN}$ increases with the density of the gas surrounding the SNe. Compared to the $\texttt{ThermalKinetic\_var}\Delta\texttt{T}_{\rm SN}\texttt{var}f_{\rm E}$ model, this model uses a constant energy per SN, as opposed to the SN

[24] We verified that varying $\Delta T_{\rm AGN}$ by 0.5 dex relative to 10^9 K in the $\texttt{ThermalKinetic\_var}\Delta\texttt{T}_{\rm SN}\texttt{var}f_{\rm E}$ model has a qualitatively similar effect on the GSMF and SSMR as varying $m_{\rm BH,pivot}$ in the fiducial COLIBRE model, which uses a variable $\Delta T_{\rm AGN}$.

energy increasing with the stellar birth pressure as adopted in the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model.

These three simplified models were fit to the observed $z = 0$ GSMF and SSMR using emulators in the same manner as the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model, and for each model, the best-fitting subgrid parameter values were found. For the `Basic` model, we optimized the parameters $m_{\rm BH,seed}$ and $f_{\rm E}$; for `ThermalKinetic`, the parameters $m_{\rm BH,seed}$, $f_{\rm E}$, and $f_{\rm kin}$; and for $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}$, $m_{\rm BH,seed}$, $f_{\rm E}$, $f_{\rm kin}$, and $n_{\rm H,pivot}$. In total, we ran ≈ 200 simulations for various combinations of the subgrid parameters and models. Our main results with regard to the calibration are as follows:

(i) The `Basic` model fails to produce a good fit to the observed $z = 0$ GSMF (Fig. 2). The GSMF exhibits a power-law shape as opposed to the observed P. 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. 4).

(ii) The `ThermalKinetic` model can successfully reproduce the observed $z = 0$ GSMF or the observed SSMR separately but cannot fit both relations simultaneously (Fig. 5). The fact that `ThermalKinetic` can provide a good match to the observed GSMF is a consequence of the ability to combine the large energy injections of the thermal channel of SN feedback with the low-energy kicks of the kinetic channel (Fig. 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 (SSMR) prefers $f_{\rm kin} \approx 0.6$ ($f_{\rm kin} \approx 0$), while fitting to both constraints gives an intermediate value of $f_{\rm kin} \approx 0.3$ (Fig. 6).

(iii) Adopting a density-dependent heating temperature $\Delta T_{\rm SN}$ (the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}$ model) improves the combined fit to the GSMF and SSMR, while additionally introducing a stellar birth pressure dependence of the SN energy (the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model) results in excellent agreement with the observed GSMF and SSMR (Fig. 7).

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

(i) The observed $z = 0$ sSFR and the galaxy quenched fractions are broadly matched by the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}$ and $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ models, but not by `Basic` and `ThermalKinetic` (Fig. 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.

(ii) The observed $z = 0$ cold gas fractions in the stellar mass range $10^9 < M_*/{\rm M}_\odot < 10^{10.5}$ are reproduced by both the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}$ and $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ models for H I, but only by $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ for H_2 (Fig. 12). In addition, $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ reproduces the observed scatter for both H I and H_2. At higher stellar masses, $M_* > 10^{10.5}\,{\rm M}_\odot$, all four models underestimate the gas fractions (Fig. 9), though, as shown by J. Schaye et al. (2025), the agreement with the data at the high-mass end can be improved by using higher resolution.

(iii) Owing to its stellar birth pressure dependence of the energy in SN feedback, the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model is the only model that provides a reasonably good match to the observed GSMF and sSFR at $z = 2$ (Fig. 11).

(iv) Similarly, due to the pressure dependence of its SN feedback, the cosmic SFRD in the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model is suppressed at high z relative to the other three models (Fig. 10). As a result, the SFRD in the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ 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 $z \approx 3$.

(v) The observed $z = 0$ relations between galaxy stellar mass and stellar and gas-phase metallicities are reproduced at $M_* \lesssim 10^{11}\,{\rm M}_\odot$ in all four models (Fig. 13).

(vi) 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. 9). This agreement is expected as the value of the AGN feedback coupling efficiency, $\varepsilon_{\rm f}$, was chosen to match the $z = 0$ observed BSMR at the high galaxy masses for which dynamical BH mass measurements are possible. Because $\varepsilon_{\rm f}$ primarily affects the normalization of the BSMR but has less impact on the GSMF and SSMR, its value was set independently of the other subgrid parameters and without using emulators.

Having demonstrated that the best-fitting $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model outperforms its three counterparts with simplified SN feedback, we applied it to simulations at two higher COLIBRE resolutions: m6 and m5. Using the best-fitting parameter values at m7 resolution as an initial guess, we manually adjusted the subgrid parameter values to achieve a fit to the $z = 0$ GSMF, SSMR, and BSMR at m6 and m5 resolutions that is similarly good as at m7. At m5 resolution, we found that $\Delta T_{\rm AGN} = 10^9$ K – the heating temperature used in the AGN feedback of the $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model–requires a low BH seed mass ($m_{\rm BH,seed} \sim 10^3\,{\rm M}_\odot$), yielding a probably unrealistically steep relation between galaxy stellar mass and BH mass; otherwise, AGN feedback is excessively strong. To allow for a higher $m_{\rm BH,seed}$ without compromising the fit to the observed GSMF and SSMR, we replaced the fixed $\Delta T_{\rm AGN} = 10^9$ K with a variable $\Delta T_{\rm AGN}$ (equation 20). This change was applied at all three COLIBRE resolutions to maintain consistency. The updated model (i.e. the modified version of the best-fitting $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$ model that uses the variable $\Delta T_{\rm AGN}$) was established as the fiducial COLIBRE model. At m7 resolution, the fiducial values of the four calibrated subgrid parameters of SN and AGN feedback are: $m_{\rm BH,seed} = 10^{5.5}\,{\rm M}_\odot$, $f_{\rm kin} = 0.1$, $P_{\rm E,pivot}/k_{\rm B} = 8 \times 10^3\,{\rm K\,cm^{-3}}$, and $n_{\rm H,pivot} = 0.6\,{\rm cm^{-3}}$. In relation to both the change in $\Delta T_{\rm AGN}$ and the extension of the calibrated m7 model to higher resolutions, we demonstrated that:

(i) The fit of the COLIBRE fiducial m7 model, which uses a variable $\Delta T_{\rm AGN}$, to the observed GSMF, SSMR, and BSMR remains as good as that of $\texttt{ThermalKinetic\_var}\Delta T_{\rm SN}\texttt{var}f_{\rm E}$, which employs $\Delta T_{\rm AGN} = 10^9$ K and was calibrated using emulators (Fig. 14). Additionally, the fiducial model demonstrates a similar level of agreement with observational data that were not used in the calibration (Fig. 15).

(ii) The fiducial m6 and m5 COLIBRE models, both of which also use a variable $\Delta T_{\rm AGN}$ but were calibrated manually, exhibit a similar level of agreement with the observed GSMF, SSMR, and BSMR as the fiducial m7 model (Fig. 16).

(iii) At m7 resolution, switching from a fixed to a variable $\Delta T_{\rm AGN}$ allows for an increase in the BH seed mass, leading to more massive BHs in low-mass galaxies by up to 0.7 dex (right-hand panel of Fig. 14). At m5 resolution, the BH mass never falls below 2×10^4 M$_\odot$, ensuring that the median BH mass grows smoothly with increasing galaxy stellar mass (right-hand panel of Fig. 16).

Having calibrated the COLIBRE fiducial model at all resolutions, we proceeded to investigate the effect of changing individual subgrid parameters in the fiducial m7 model, including the parameters that were not optimized by the emulators. We confirmed that the latter parameters are either degenerate with the parameters that were optimized and/or have little impact on the $z = 0$ GSMF and SSMR (Figs 17 and 18).

In closing, we stress that calibrating a galaxy formation model is a numerically demanding process with no guarantee of success. The fact that the COLIBRE fiducial m7 model fit to the $z = 0$ GSMF and SSMR reproduces many observed relations that were not considered during the calibration (Fig. 15) is an encouraging result. Part of this agreement can be attributed to the choices made during the development and testing of various physical prescriptions implemented in COLIBRE, including the model for radiative cooling (S. Ploeckinger et al. 2025), the model for dust grains (J. W. Trayford et al. 2026), the star formation prescription (F. S. J. Nobels et al. 2024), the model for metal enrichment and turbulent diffusion of element mass fractions (Correa et al. in preparation), the prescription for early stellar feedback (A. Benítez-Llambay et al. 2025), and the model for SN feedback (E. Chaikin et al. 2023). For a detailed discussion of the performance of the COLIBRE simulations at higher resolutions and in larger cosmological volumes, we refer the reader to J. Schaye et al. (2025), who present results from the fiducial COLIBRE models at m7, m6, and m5 resolutions in the largest available cosmological volumes at $z = 0$: 400^3 cMpc^3, 200^3 cMpc^3, and 25^3 cMpc^3, respectively.

ACKNOWLEDGEMENTS

The authors of this work acknowledge the pioneering impact that the late Richard Bower had on the use of emulation techniques in galaxy formation simulations, which led to the highly successful emulation campaign for FLAMINGO, and now for the COLIBRE model. His overwhelmingly positive demeanor and sharp physical insight are sorely missed by the COLIBRE team and across the astronomical community. We thank the referee for a constructive report. 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. This project has received funding from the Netherlands Organization for Scientific Research (NWO) through research programme Athena 184.034.002. ABL acknowledges support by the Italian Ministry for Universities and Research (MUR), program 'Dipartimenti di Eccellenza 2023-2027' within the Centro Bicocca di Cosmologia Quantitativa (BiCoQ), and support by UNIMIB's Fondo di Ateneo Quota Competitiva (project 2024-ATEQC-0050). CGL acknowledges support from STFC grants ST/T000244/1 and ST/X001075/1. CSF acknowledges support from European Research Council (ERC) Advanced Grant DMIDAS (GA 786910). EC was supported by the funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 860744 (BiD4BESt). JT acknowledges support of STFC Grant ST/X004651/1. RAC acknowledges support from STFC grants ST/Y002482/1 and ST/Y001907/1. SP acknowledges support from the Austrian Science Fund (FWF) through project V 982-N. VJFM acknowledges support by NWO through the Dark Universe Science Collaboration (OCENW.XL21.XL21.025). YMB acknowledges support from UK Research and Innovation through a Future Leaders Fellowship (grant agreementMR/X035166/1) and financial support from the Swiss National Science Foundation (SNSF) under project 200021_213076. The research in this paper made use of the SWIFT open-source simulation code (http://www.swiftsim.com, M. Schaller et al. 2018) version 1.0.0. The data analysis was carried out with the use of SWIFTSIMIO (J. Borrow & A. Borrisov 2020; J. Borrow & A. J. Kelly 2021), NUMPY (C. R. Harris et al. 2020), SCIPY (P. Virtanen et al. 2020), MATPLOTLIB (J. D. Hunter 2007), and SEABORN (M. L. Waskom 2021).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. The public version of the SWIFT simulation code can be found on www.swiftsim.com. The SWIFT modules related to the COLIBRE galaxy formation model will be integrated into the public version after the public release of COLIBRE. The CHIMES astrochemistry code is publicly available at https://richings.bitbucket.io/chimes/home.html.

REFERENCES

Abbott T. M. C. et al., 2022, Phys. Rev. D, 105, 023520
Accurso G. et al., 2017, MNRAS, 470, 4750
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Bahé Y. M. et al., 2016, MNRAS, 456, 1115
Bahé Y. M. et al., 2022, MNRAS, 516, 167
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Barber C., Crain R. A., Schaye J., 2018, MNRAS, 479, 5448
Barber C., Schaye J., Crain R. A., 2019, MNRAS, 483, 985
Bate M. R., Burkert A., 1997, MNRAS, 288, 1060
Bauer A. E. et al., 2013, MNRAS, 434, 209
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Belfiore F. et al., 2018, MNRAS, 477, 3014
Benítez-Llambay A. et al., 2025, preprint (arXiv:2509.25309)
Béthermin M. et al., 2023, A&A, 680, L8
Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, MNRAS, 512, 3703
Black J. H., 1987, in Hollenbach D. J., Thronson H. A., eds, Astrophysics and Space Science Library, Vol. 134, Interstellar Processes. Springer, Dordrecht, p. 731
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Booth C. M., Schaye J., 2010, MNRAS, 405, L1
Borrow J., Borrisov A., 2020, J. Open Source Softw., 5, 2430
Borrow J., Kelly A. J., 2021, preprint (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, 526, 2441
Bouwens R., Illingworth G., Oesch P., Stefanon M., Naidu R., van Leeuwen I., Magee D., 2023, MNRAS, 523, 1009
Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, MNRAS, 407, 2017
Bregman J. N., Harrington J. P., 1986, ApJ, 309, 833
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80

Bundy K. et al., 2014, ApJ, 798, 7
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
Cochrane R. K. et al., 2023, MNRAS, 523, 6082
Cook R. H. W., Cortese L., Catinella B., Robotham A., 2019, MNRAS, 490, 4060
Covelo-Paz A. et al., 2025, A&A, 694, A178
Crain R. A., van de Voort F., 2023, ARA&A, 61, 473
Crain R. A. et al., 2015, MNRAS, 450, 1937
Crain R. A. et al., 2017, MNRAS, 464, 4204
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., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
de Graaff A., Trayford J., Franx M., Schaller M., Schaye J., van der Wel A., 2022, MNRAS, 511, 2544
Dehnen W., Aly H., 2012, MNRAS, 425, 1068
Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, ApJ, 676, 33
Diemer B. et al., 2018, ApJS, 238, 33
Dolag K. et al., 2025, preprint (arXiv:2504.01061)
Donnan C. T. et al., 2024, MNRAS, 533, 3222
Driver S. P. et al., 2011, MNRAS, 413, 971
Driver S. P. et al., 2012, MNRAS, 427, 3244
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
Eddington A. S., 1913, MNRAS, 73, 359
Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263
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
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Forouhar Moreno V. J., Helly J., McGibbon R., Schaye J., Schaller M., Han J., Kugel R., 2025, MNRAS, 543, 1339
Fraser-McKelvie A. et al., 2022, MNRAS, 510, 320
Fu S. et al., 2025, ApJ, 987, 186
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Gardner J. et al., 2006, Space Sci. Rev., 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, Commun. Appl. Math. Comput. Sci., 5, 65
Graham A. W., Sahu N., 2023, MNRAS, 518, 2177
Graham A. W., Sahu N., 2024, MNRAS, 530, 3429
Greengard L., Rokhlin V., 1987, J. Comput. Phys., 73, 325
Gruppioni C. et al., 2020, A&A, 643, A8
Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, Astrophysics Source Code Library, record ascl:2008.024
Hahn O., Rampf C., Uhlemann C., 2021, MNRAS, 503, 426
Han J., Cole S., Frenk C. S., Benitez-Llambay A., Helly J., 2018, MNRAS, 474, 604
Hardwick J. A., Cortese L., Obreschkow D., Catinella B., Cook R. H. W., 2022, MNRAS, 509, 3751
Harikane Y. et al., 2023, ApJS, 265, 5
Harris C. R. et al., 2020, Nature, 585, 357
Henden N. A., Puchwein E., Shen S., Sijacki D., 2018, MNRAS, 479, 5385
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Huško F. et al., 2026, MNRAS, 547, stag324
Ikeda R. et al., 2022, ApJ, 933, 11
Katsianis A. et al., 2020, MNRAS, 492, 5592
Kaviraj S., Lazar I., Watkins A. E., Laigle C., Martin G., Jackson R. A., 2025, MNRAS, 538, 153
Kennicutt Robert C. J., 1998, ApJ, 498, 541
Kennicutt Robert C. J. et al., 2007, ApJ, 671, 333
Khusanova Y. et al., 2021, A&A, 649, A152
Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, ApJ, 779, 102
Krumholz M. R., 2013, MNRAS, 436, 2747
Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36
Krumholz M. R., McKee C. F., Klein R. I., 2006, ApJ, 638, 369
Kugel R., Borrow J., 2022, J. Open Source Softw., 7, 4240
Kugel R. et al., 2023, MNRAS, 526, 6103
Lagos C. d. P. et al., 2015, MNRAS, 452, 3815
Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, MNRAS, 441, 1270
Le Fèvre O. et al., 2020, A&A, 643, A1
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
Manuwal A., Stevens A. R. H., 2023, MNRAS, 523, 2738
Martín-Navarro I., La Barbera F., Vazdekis A., Falcón-Barroso J., Ferreras I., 2015, MNRAS, 447, 1033
Mathis J. S., Mezger P. G., Panagia N., 1983, A&A, 128, 212
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
McGibbon R., Helly J., Schaye J., Schaller M., Vandenbroucke B., 2025, The J. Open Source Softw., 10, 8252
McKay M. D., Beckman R. J., Conover W. J., 1979, Technometrics, 21, 239
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
Muzzin A. et al., 2013, ApJ, 777, 18
Nobels F. S. J., Schaye J., Schaller M., Ploeckinger S., Chaikin E., Richings A. J., 2024, MNRAS, 532, 3299
Novak M. et al., 2017, A&A, 602, A5
Ostriker E. C., 1999, ApJ, 513, 252
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., 2024, MNRAS, 528, 2930
Ploeckinger S., Richings A. J., Schaye J., Trayford J. W., Schaller M., Chaikin E., 2025, MNRAS, 543, 891
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. MIT Press, Cambridge, MA

Richings A. J., Schaye J., Oppenheimer B. D., 2014a, MNRAS, 440, 3349
Richings A. J., Schaye J., Oppenheimer B. D., 2014b, MNRAS, 442, 2780
Robertson B. E., Kravtsov A. V., 2008, ApJ, 680, 1083
Robotham A. S. G., Bellstedt S., Lagos C. d. P., Thorne J. E., Davies L. J., Driver S. P., Bravo M., 2020, MNRAS, 495, 905
Saintonge A. et al., 2017, ApJS, 233, 22
Schaller M., Dalla Vecchia C., Schaye J., Bower R. G., Theuns T., Crain R. A., Furlong M., McCarthy I. G., 2015, MNRAS, 454, 2277
Schaller M. et al., 2018, Astrophysics Source Code Library, record ascl:1805.020
Schaller M. et al., 2024, MNRAS, 530, 2378
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Schaye J. et al., 2010, MNRAS, 402, 1536
Schaye J. et al., 2015, MNRAS, 446, 521
Schaye J. et al., 2023, MNRAS, 526, 4978
Schaye J. et al., 2025, preprint (arXiv:2508.21126)
Schechter P., 1976, ApJ, 203, 297
Schmidt M., 1959, ApJ, 129, 243
Scholte D. et al., 2024, MNRAS, 535, 2341
Segers M. C., Schaye J., Bower R. G., Crain R. A., Schaller M., Theuns T., 2016, MNRAS, 461, L102
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Smith M. C. et al., 2024, MNRAS, 527, 1216
Springel V., Hernquist L., 2003, MNRAS, 339, 289
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
Stevens A. R. H. et al., 2019, MNRAS, 483, 5334
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, ARA&A, 58, 157
Teyssier R., 2002, A&A, 385, 337
Thomas J. et al., 2011, MNRAS, 415, 545
Trayford J. W. et al., 2026, MNRAS, 545, staf2040
Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, MNRAS, 470, 1121
Tremonti C. A. et al., 2004, ApJ, 613, 898
Truelove J. K., Klein R. I., McKee C. F., Holliman J. H. II, Howell L. H., Greenough J. A., 1997, ApJ, 489, L179
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
Waskom M. L., 2021, J. Open Source Softw., 6, 3021
Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, MNRAS, 399, 574
Wilkinson M. J., Ludlow A. D., Lagos C. d. P., Fall S. M., Schaye J., Obreschkow D., 2023, MNRAS, 519, 5942
Woosley S. E., Eastman R. G., Schmidt B. P., 1999, ApJ, 516, 788
Wootten A., Thompson A. R., 2009, Proc. IEEE, 97, 1463
Wright E. L. et al., 2010, AJ, 140, 1868
Zibetti S., Charlot S., Rix H.-W., 2009, MNRAS, 400, 1181

APPENDIX A: THE EFFECT OF THE SIMULATION BOX SIZE

Fig. A1 shows the $z = 0$ GSMF (left panel) and the $z = 0$ median SSMR (right panel) from simulations using the fiducial COLIBRE m7 model in different cosmological volumes: 25^3, 50^3, 100^3, and 200^3 cMpc^3 (shown in progressively darker shades of red). The dark-red shaded region indicates the Poisson uncertainty for the GSMF and the 16^{th}–84^{th} percentile range for the SSMR in the $(200\,\text{cMpc})^3$ simulation.

The fiducial m7 model in 100^3 and 200^3 cMpc^3 volumes exhibits excellent agreement with the observed $z = 0$ GSMF and SSMR, despite being calibrated to these data using emulators trained on $(50\,\text{cMpc})^3$ volume simulations. In contrast, the simulation in the smallest volume, $(25\,\text{cMpc})^3$, shows noticeable deviations from the observations, particularly at the high-mass end, due to the limited number of massive haloes. In other words, this comparison demonstrates that, although not ideal, a $(50\,\text{cMpc})^3$ volume is sufficient for performing calibration to the observed $z = 0$ GSMF and SSMR, while $(25\,\text{cMpc})^3$ would have been too small.

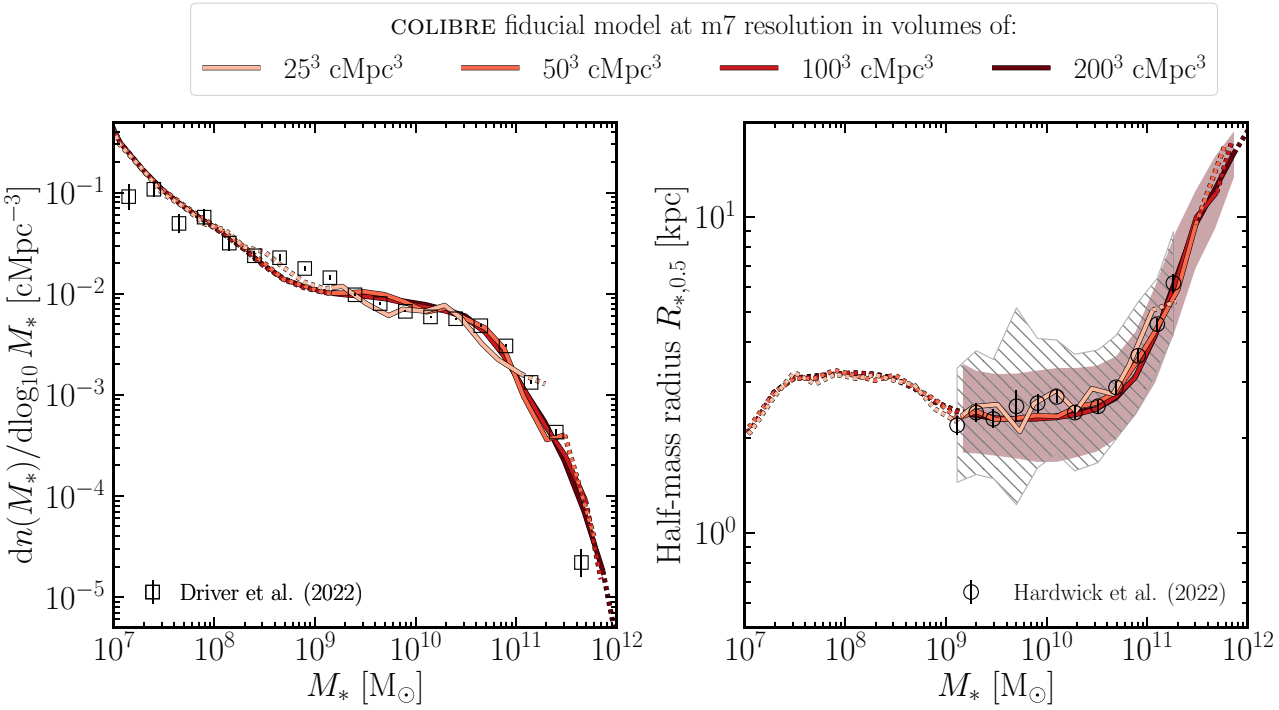

Figure A1. The $z = 0$ GSMF (*left*) and the $z = 0$ median SSMR (*right*) for the COLIBRE fiducial COLIBRE m7 model in different cosmological volumes: 25^3, 50^3, 100^3, and 200^3 cMpc^3 (shown in progressively darker shades of red). The dark-red shaded region represents Poisson uncertainty for the GSMF and the 16$^{\mathrm{th}}$ to 84$^{\mathrm{th}}$ percentile scatter for the SSMR in the largest simulation. Observational data are shown as black squares in the left panel (GSMF from S. P. Driver et al. 2022) and black circles in the right panel (SSMR from J. A. Hardwick et al. 2022), with the grey hatched region indicating the galaxy population-wide scatter from J. A. Hardwick et al. (2022). Despite being calibrated using emulators trained on 50^3 cMpc^3 volume simulations, the COLIBRE model shows excellent agreement with the observed $z = 0$ GSMF and SSMR also in 100^3 and 200^3 cMpc^3 cosmological volumes.

APPENDIX B: REDSHIFT EVOLUTION OF THE ENERGY IN SN FEEDBACK

In Section 5, we showed that among the four models with best-fitting parameter values identified by the emulators, `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` provides the closest match to the observational data. A major reason for its superior performance over the other three models is the stellar birth pressure-dependent energy in SN feedback, employed only in `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` (the other three models use a fixed energy per SN). In this section, we provide additional details on the pressure dependence in the `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` model and show its impact on the energy released by SNe at different redshifts.

Fig. B1 compares the best-fitting `ThermalKinetic_varΔT$_{\rm SN}$` (navy-blue) and `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` (light-blue) models at m7 resolution in a (50 cMpc)^3 volume. The top panel shows the cosmic SFRD, the middle panel shows the median energy per CC SN (in units of 10^{51} erg) as a function of redshift, and the bottom panel shows the median stellar birth pressure versus redshift. The light-blue shaded regions in the middle and bottom panels denote the 16$^{\mathrm{th}}$ to 84$^{\mathrm{th}}$ percentile scatter in the `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` model. In the middle panel, the thin horizontal light-blue dotted lines indicate the minimum and maximum allowed SN energies in the `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` model, while the horizontal long-dashed line marks the average SN energy across the entire simulation, all in units of 10^{51} erg. In the bottom panel, the light-blue dash–dotted line indicates the value of the subgrid parameter $P_{\rm E,pivot}$ used in the `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` model.

We find that in the `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` model, the median CC SN energy increases monotonically with redshift due to the corresponding rise in stellar birth pressure, as dictated by equation (2). At $z > 2$, the median SN energy in `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` significantly exceeds the constant CC SN energy in `ThermalKinetic_varΔT$_{\rm SN}$`, resulting in stronger SN feedback and causing `ThermalKinetic_varΔT$_{\rm SN}$varf$_{\rm E}$` to agree better with the observed SFRD at these redshifts.

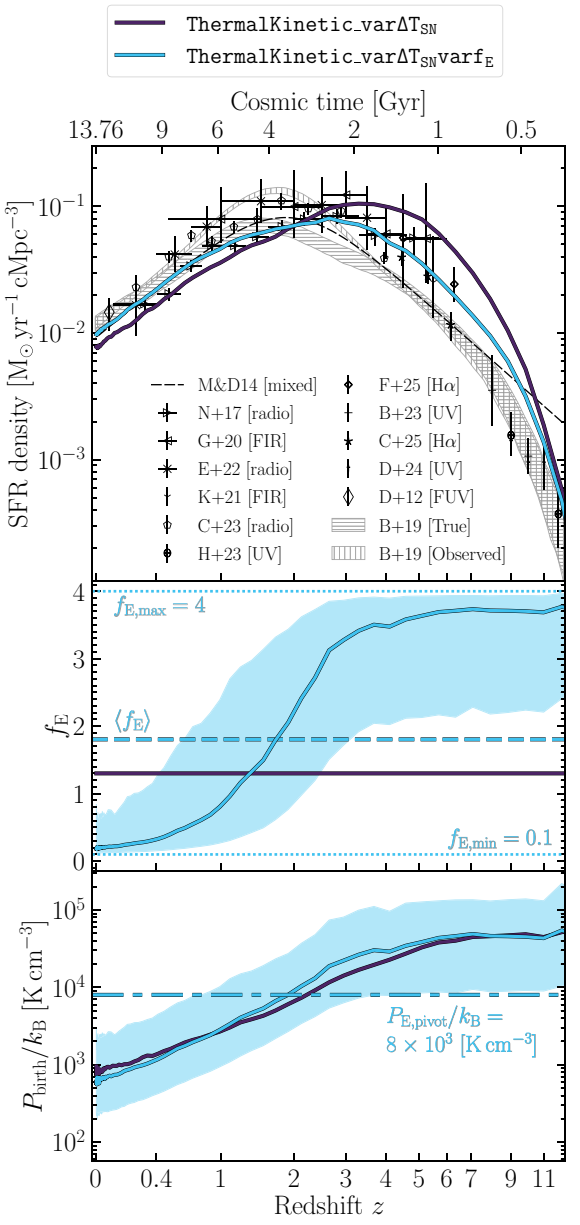

Figure B1. Comparison of the best-fitting `ThermalKinetic_varΔT_SN` (navy-blue) and `ThermalKinetic_varΔT_SNvarf_E` (light-blue) models at m7 resolution in a (50 cMpc)^3 volume. The top, middle, and bottom panels show, respectively, the cosmic SFRD, the median energy per CC SN in units of 10^{51} erg, and the median stellar birth pressure versus redshift. In the middle and bottom panels, the shaded regions represent the 16$^{\rm th}$ to 84$^{\rm th}$ percentile scatter in the `ThermalKinetic_varΔT_SNvarf_E` model. For reference, in the middle panel, the horizontal light-blue dotted lines mark the minimum and maximum allowed CC SN energy values in `ThermalKinetic_varΔT_SNvarf_E`, while the horizontal long-dashed line indicates the average SN energy over the entire simulation, all in units of 10^{51} erg. The dash–dotted line in the bottom panel indicates the value of the subgrid parameter $P_{\rm E,pivot}/k_{\rm B}$ used in `ThermalKinetic_varΔT_SNvarf_E`. The median SN energy in the `ThermalKinetic_varΔT_SNvarf_E` model increases monotonically with redshift, causing it to better match the observed SFRD at high z compared to `ThermalKinetic_varΔT_SN`, where the CC SN energy is constant.

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