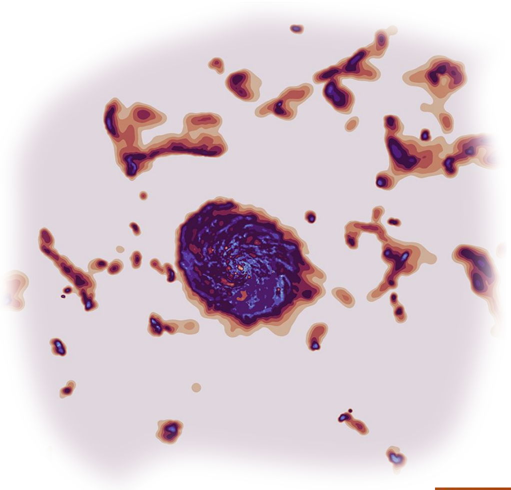
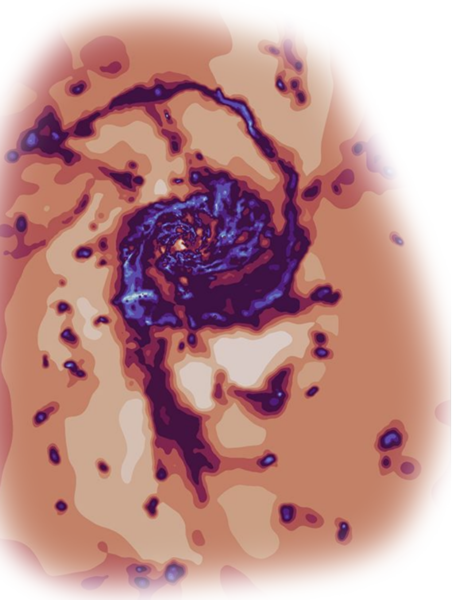
Calibrating stellar and AGN feedback in cosmological simulations that include a cold interstellar gas phase





Evgenii Chaikin

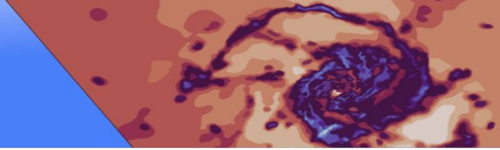
Leiden Observatory

University of Southampton

BiD4BESt, 6-8 February, 2024

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860744

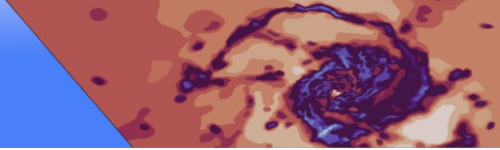
Overview



Calibrating subgrid feedback in galaxy simulations: the search for model parameter values for which the simulations best reproduce the target observational data

- A preliminary version of the COLIBRE galaxy formation model
- An emulator-based approach for tuning subgrid parameters of supernova (SN) and AGN feedback
- Bayesian inference to fit the simulations to observational data

The COLIBRE model



Code: swift (Schaller et al. 2023)

Hydro: SPH scheme Sphenix (Borrow et al. 2022)

Cooling: Non-equilibrium chemistry solver CHIMES (Richings et al. 2014) for H and He; metal cooling by a modified version of the Ploeckinger & Schaye (2020) cooling tables

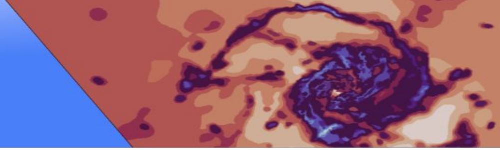
Dust: A live dust model coupled to the chemistry (Trayford et al., in prep.)

Star formation: Gravitational stability criterion (Nobels et al. 2023) with the Schmidt star-formation law

Feedback:

- Isotropic stochastic thermal-kinetic SN feedback (Chaikin et al. 2022, 2023)
- HII regions, stellar winds, and radiation pressure (Ploeckinger et al., in prep.)
- Thermal AGN feedback with a Bondi-Hoyle accretion model (Booth & Schaye 2009, Bahé et al. 2022)

The setup for calibration



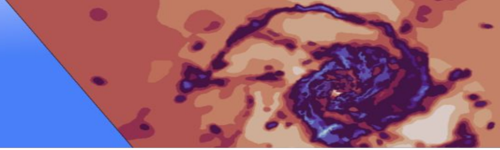
1. All simulations in a $(50\ \text{cMpc})^3$ **volume**

2. (COLIBRE low) **resolution**
   - Baryons: $1.5 \times 10^7\ M_\odot$
   - DM: $1.9 \times 10^7\ M_\odot$

3. Gaussian-process **emulators** (Kugel & Borrow, 2022) for calibrating
   - galaxy stellar mass function (GSMF) at z=0
   - size-stellar mass relation (SSM) at z=0

4. Use the emulators to **optimize**
   - *one* AGN feedback parameter (BH seed mass)
   - *one to three* SN feedback parameters (depending on the SN feedback prescription)

Four prescriptions for SN feedback



We calibrate four models with different prescriptions for SN feedback:

Complexity of SN feedback

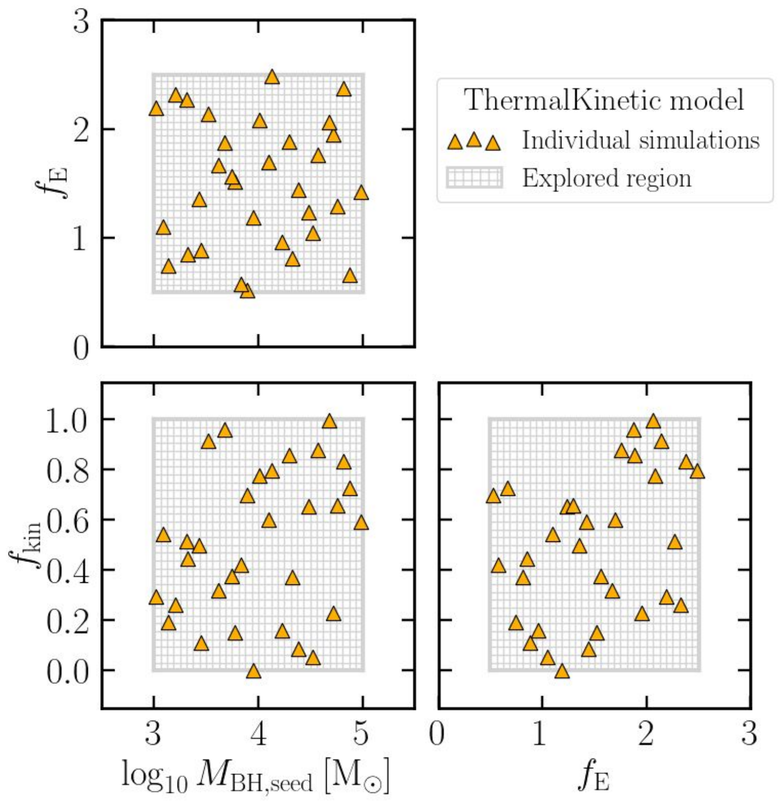
- **Basic**
  - Stochastic thermal SN feedback with a constant SN energy, f_E, and constant heating temperature of $\Delta T = 10^{7.5}$ K
- **ThermalKinetic**
  - As **Basic** but a fraction of the SN energy f_E, f_{kin}, is injected in kinetic form with 50 km/s velocity kicks
- **ThermalKineticVariableΔT**
  - As **ThermalKinetic** but ΔT is an increasing function of gas density
- **Reference (COLIBRE)**
  - As **ThermalKineticVariableΔT** but f_E is an increasing function of the gas pressure at stellar birth

Emulators' training data (ThermalKinetic model)

- Parameters of the **ThermalKinetic** model*:*
  1. BH seed mass, $M_{\mathrm{BH,seed}}$
  2. Energy in SN feedback, f_{E} (in units of 10^{51} erg)
  3. Fraction of SN energy injected in kinetic form, f_{kin}

- The training data consists of 32 simulations

- The Latin hypercube sampling technique



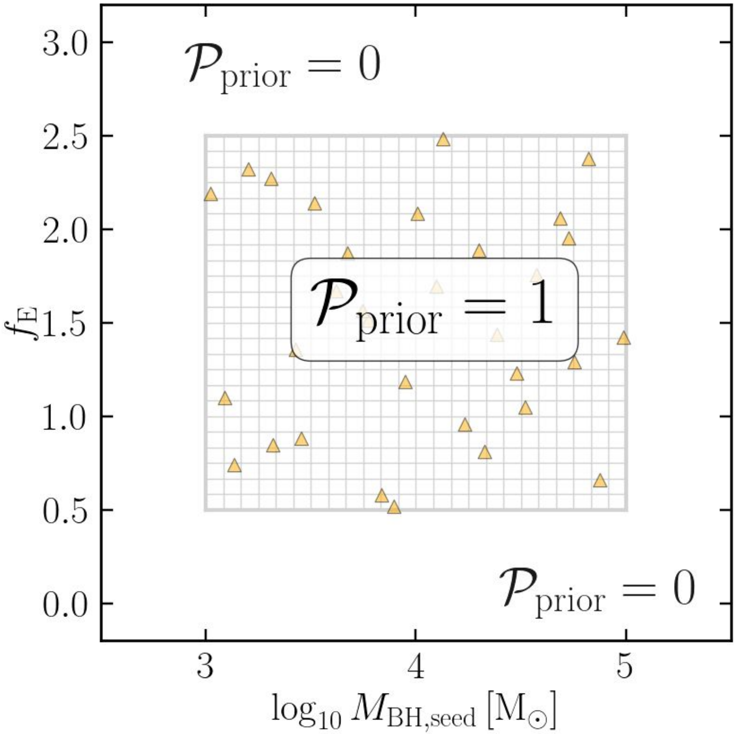
Prior and likelihood

- A **uniform prior** within the studied region of the parameter space

- The **likelihood function** with equal contributions from the size-stellar mass relation (SSM) and galaxy stellar mass function (GSMF)

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \ln \mathcal{L}_{\mathrm{GSMF}}(\boldsymbol{\theta}) + \ln \mathcal{L}_{\mathrm{SSM}}(\boldsymbol{\theta})$$

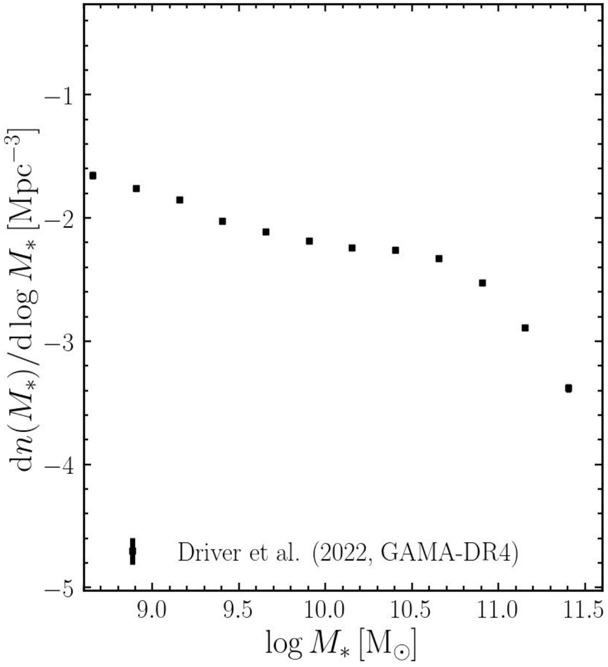
- **Errors** between observational data and emulator predictions are **Gaussian distributed**

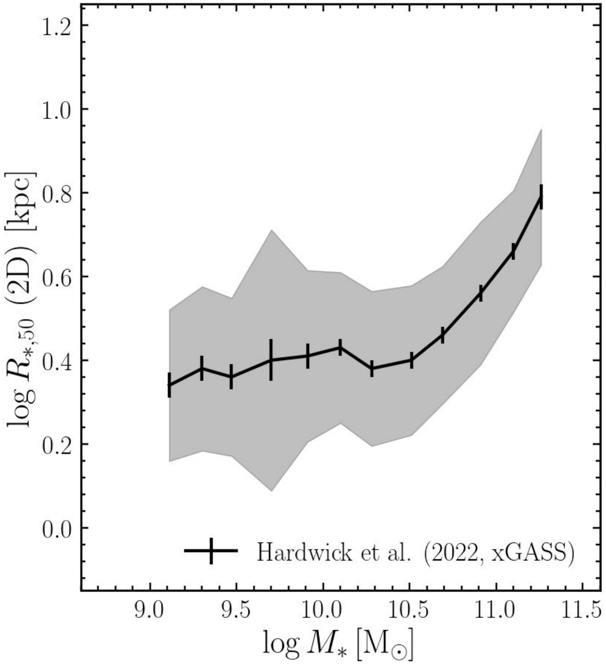


Target observational data

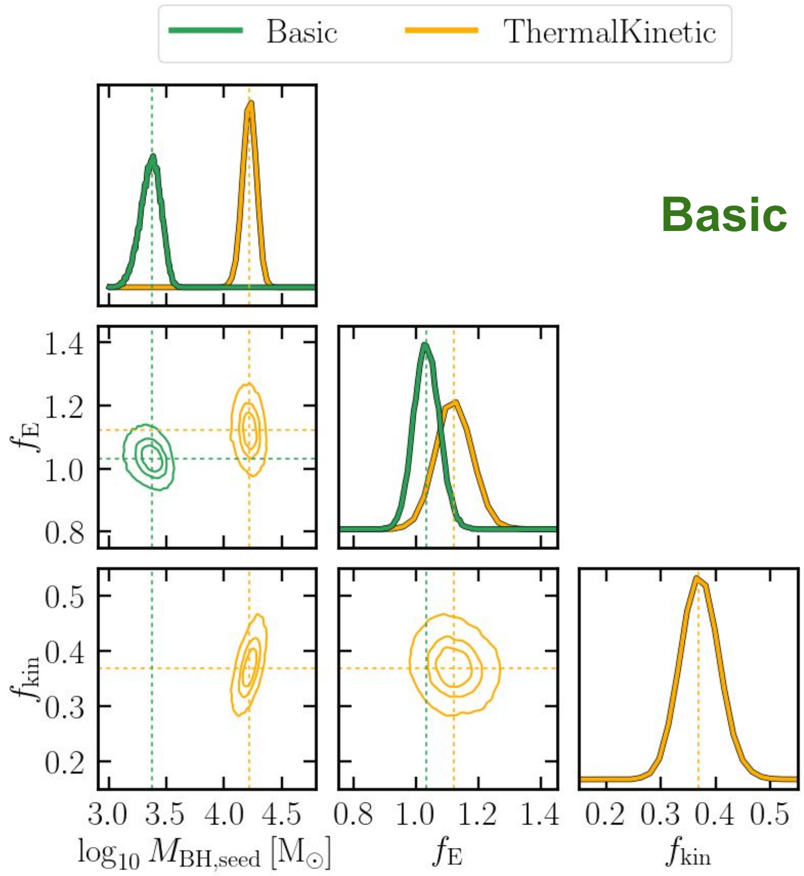
Two z~0 observational datasets:

- Galaxy stellar mass function (GSMF) from Driver et al. (2022)
- Galaxy size-stellar mass relation (SSM) from Hardwick et al. (2022)





Posterior: Basic and ThermalKinetic



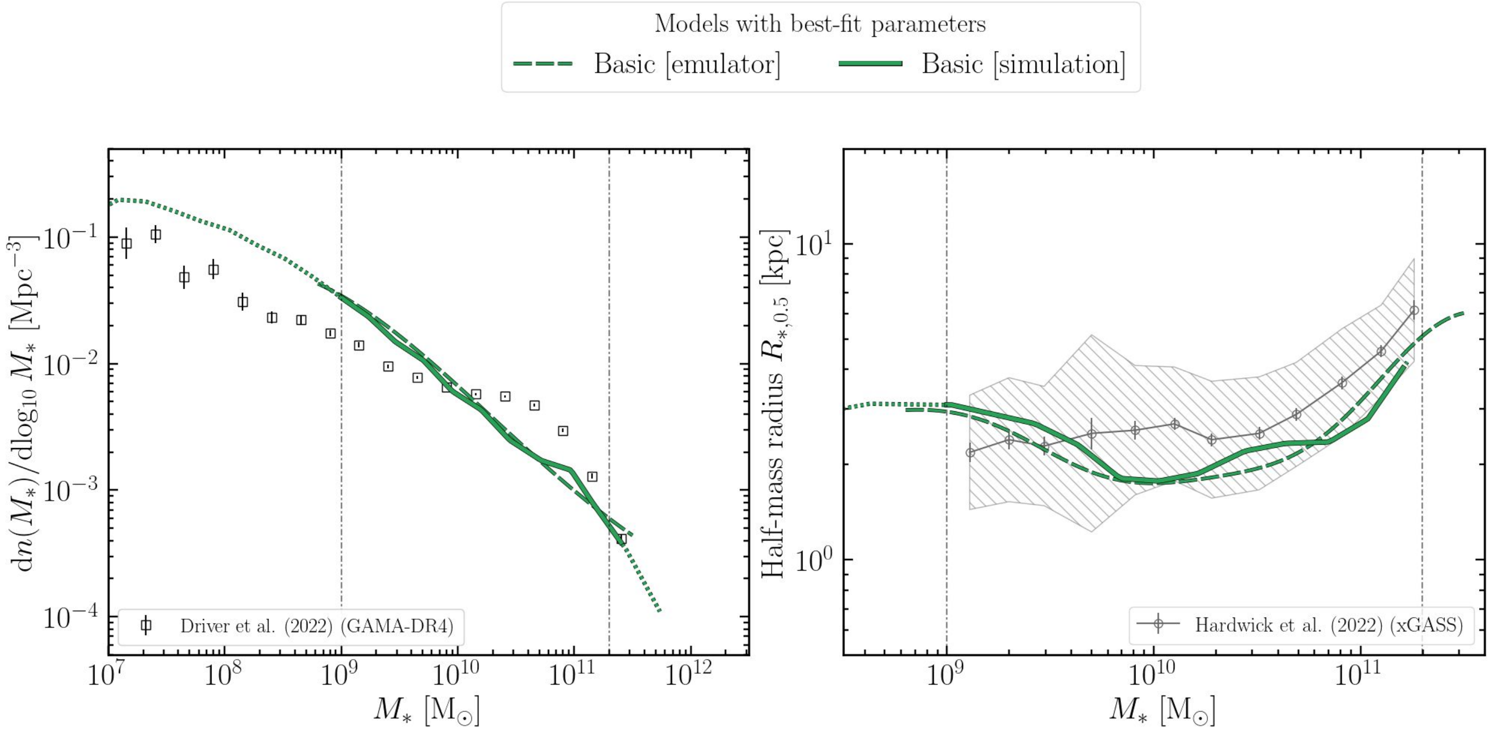
Basic

- $M_{BH,seed}$: the seed mass of BHs
- f_E: the energy in SN feedback (in units of 10^{51} erg)

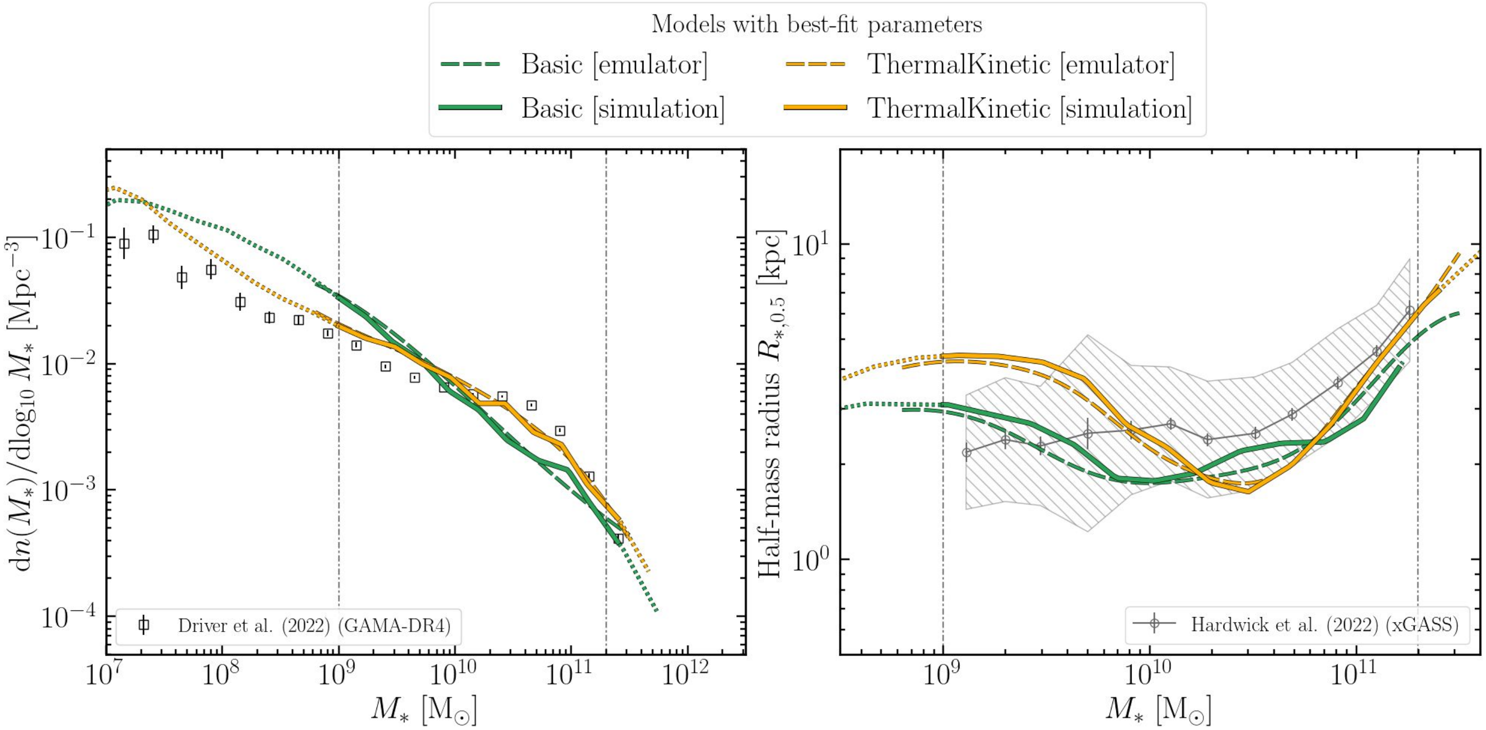
ThermalKinetic

- f_{kin}: the fraction of f_E that is injected in kinetic form

Performance: Basic vs. ThermalKinetic

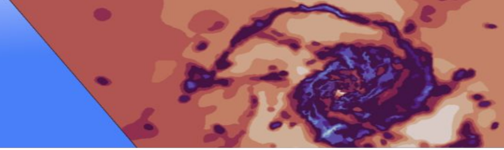


Performance: Basic vs. ThermalKinetic

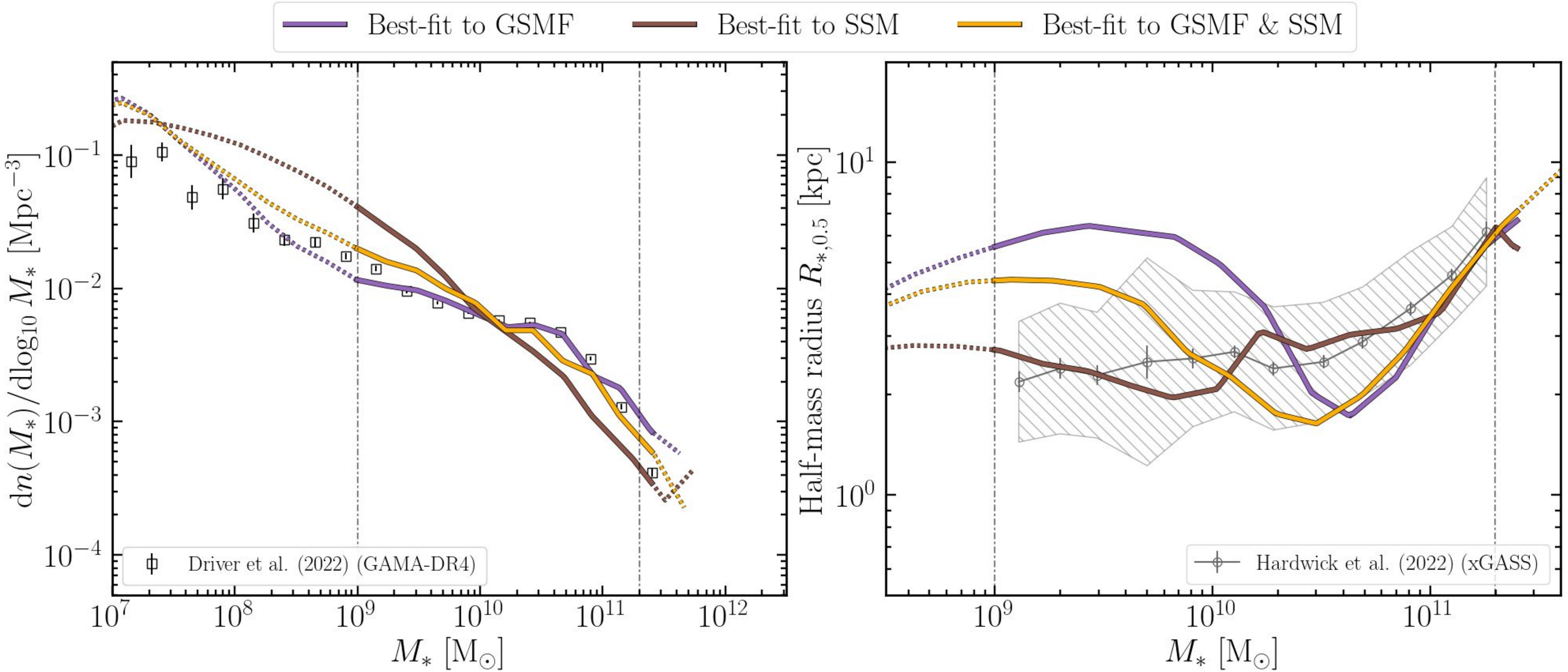


- Although **ThermalKinetic** produces a combined fit to the observed GSMF and SSM better than **Basic**, neither model is particularly good

ThermalKinetic model fit to GSMF and SSM *separately and together*



- Fitting only to the observed GSMF results in a good match to the GSMF but a poor match to the SSM
- Fitting to both observed relations at the same time produces only a reasonable match to both

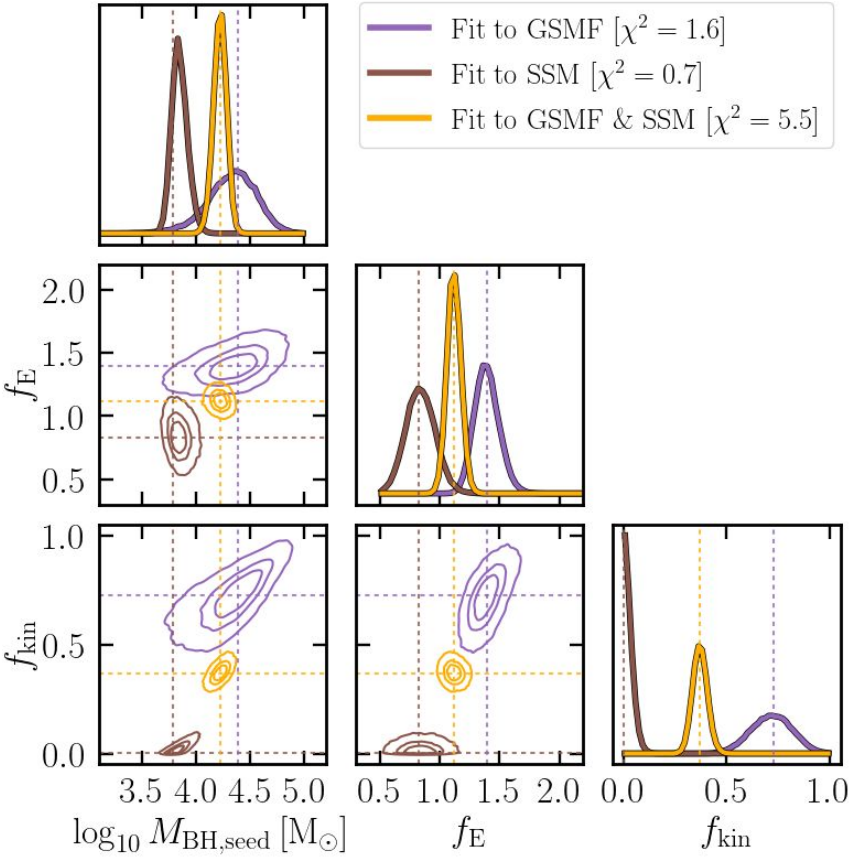


ThermalKinetic model fit to GSMF and SSM *separately and together*

- The best-fitting parameter values of **the model fit to the GSMF** and **the model fit to the SSM** belong to different regions of the parameter space

- **The model fit to both the GSMF and SSM** is located in between those fit to the GSMF and SSM separately

Implication: *more complex SN feedback may be needed*



SN feedback variation 3: **ThermalKineticVariableΔT**

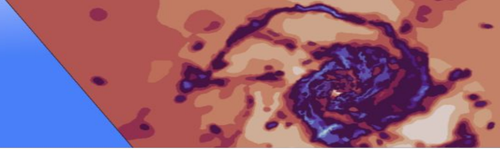
- **ThermalKineticVariableΔT**: as **ThermalKinetic** but the constant heating temperature, $\Delta T = 10^{7.5}$ K, is replaced with

$$\Delta T_{\rm SN}(n_{\rm H,SN}) = 10^{6.5}\,{\rm K}\left(\frac{n_{\rm H,SN}}{n_{\rm SN,0}}\right)^{2/3}$$

$n_{\rm H,SN}$ is the gas density, $10^{6.5} < \Delta T < 10^{7.5}$ K, and the slope of ⅔ is motivated by Dalla Vecchia & Schaye (2012)

- The parameter $n_{\rm SN,0}$ is optimized by the emulators
- In total, the model has 4 parameters that are optimized: $\boldsymbol{\theta} = (M_{\rm BH,seed}, f_{\rm E}, f_{\rm kin}, n_{\rm SN,0})$

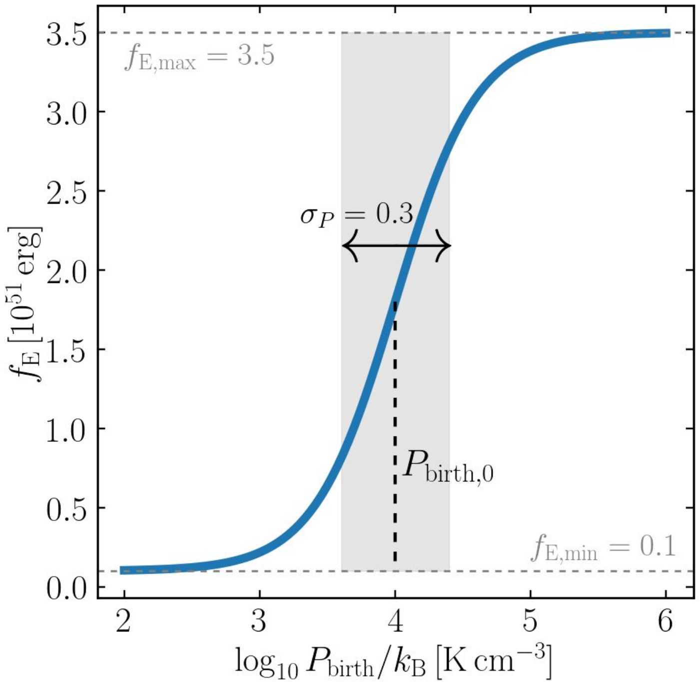
SN feedback variation 4: **Reference**



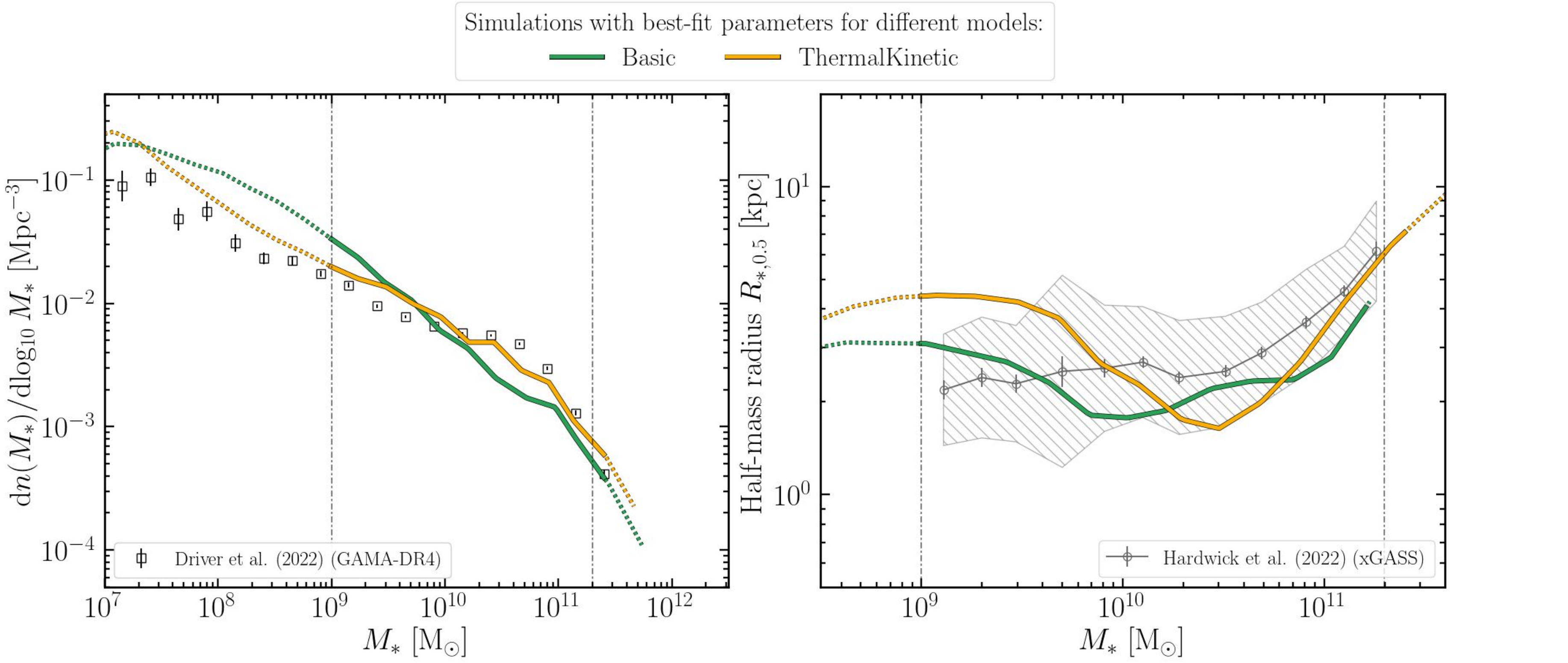
- **Reference**: as **ThermalKineticVariableΔT** but the constant SN energy, f_E , is replaced with a function $f_E(P_{birth})$ where P_{birth} is the pressure of the gas environment in which stellar particle are formed

The relation $f_E(P_{birth})$ depends on 4 parameters

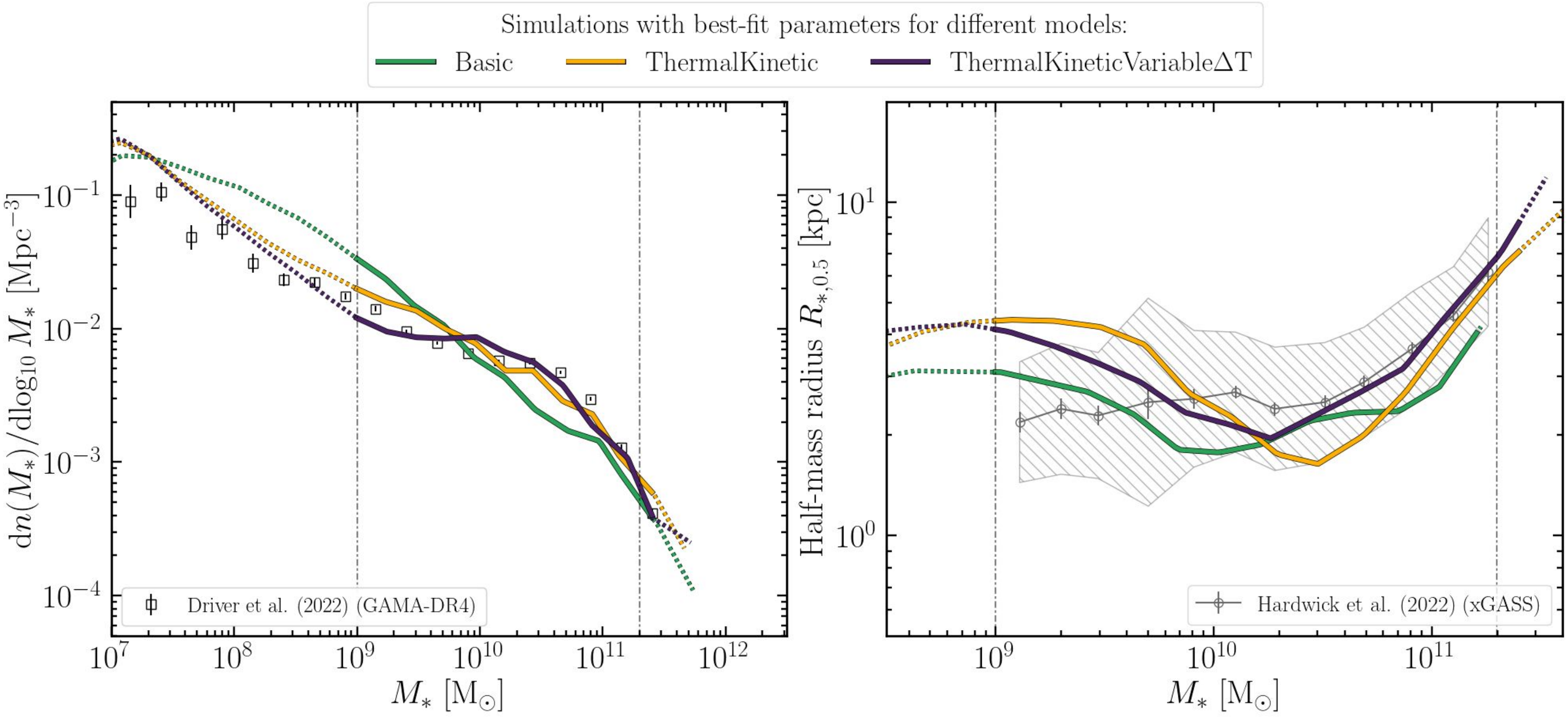
- We set σ_P = 0.3, $f_{E,min}$ = 0.1, and $f_{E,max}$ = 3.5
- The parameter $P_{birth,0}$ is optimized by the emulators



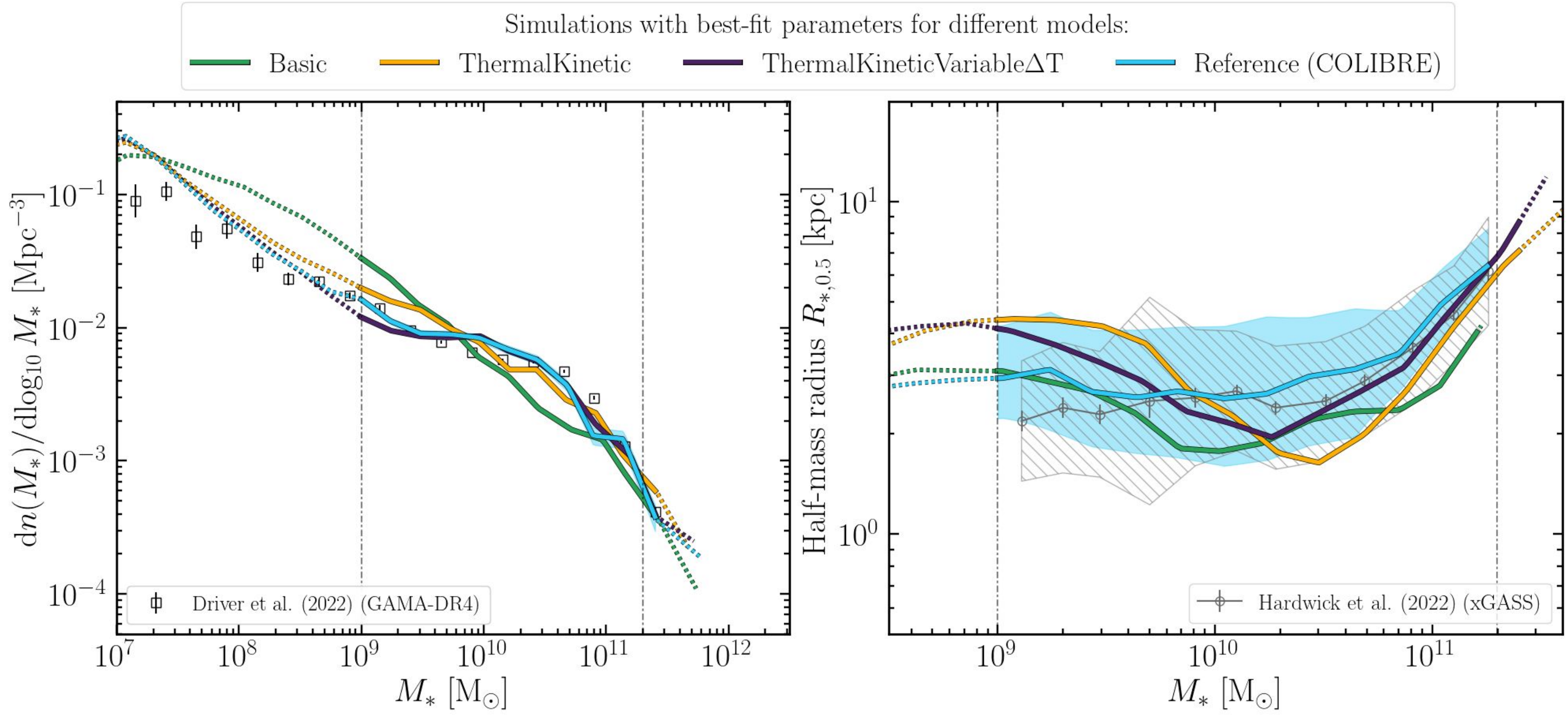
GSMF and SSM in the four models [split in three slides]



GSMF and SSM in the four models

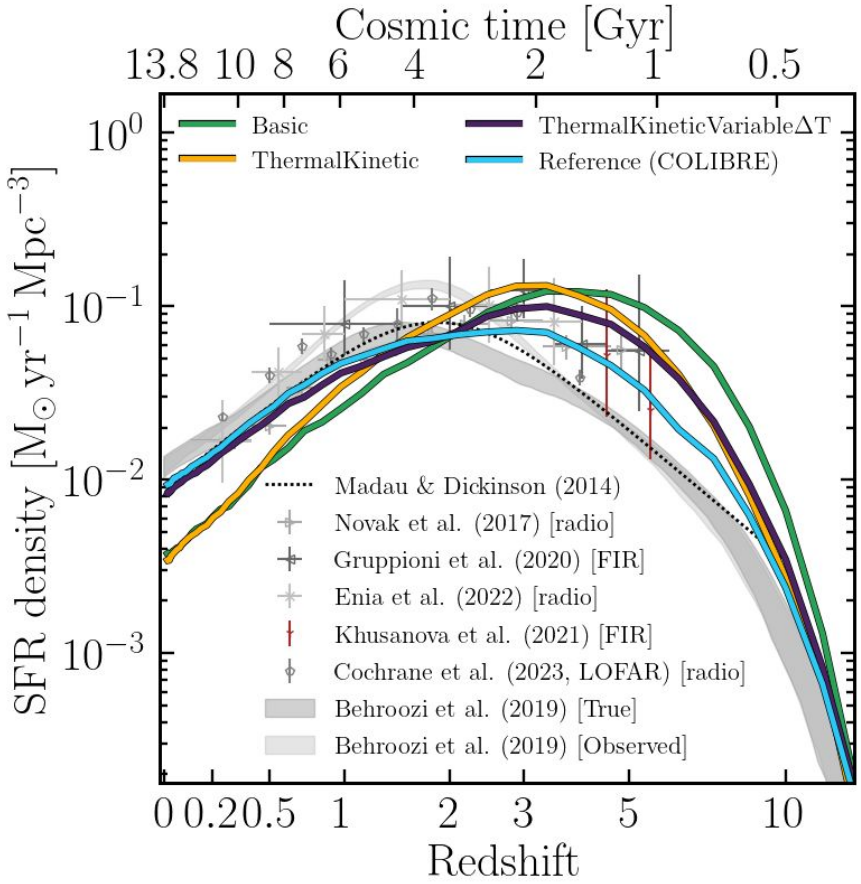


GSMF and SSM in the four models



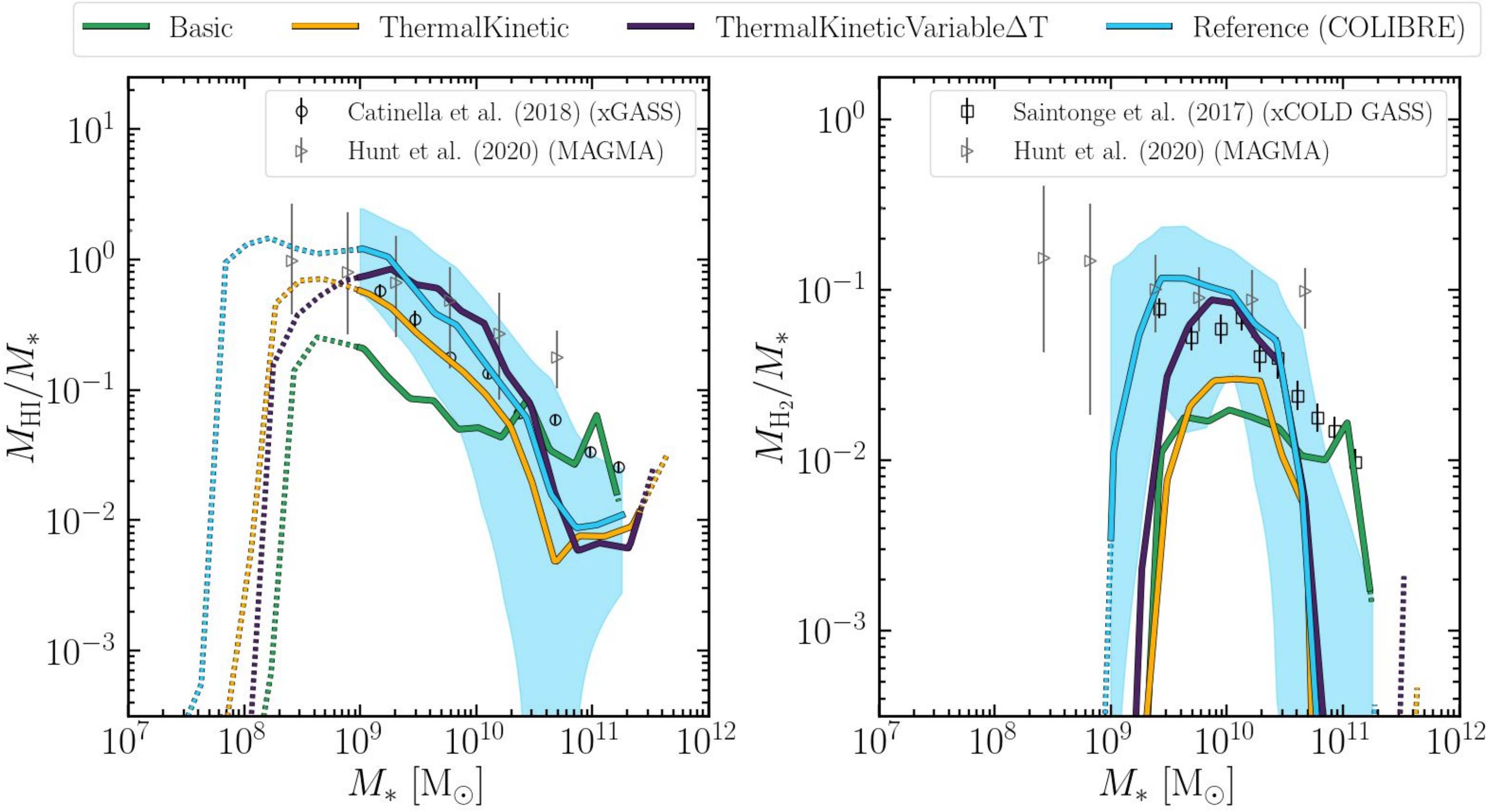
- **The reference model** is the only model that successfully reproduces the observed GSMF and SSM

Star formation rate density (SFRD)

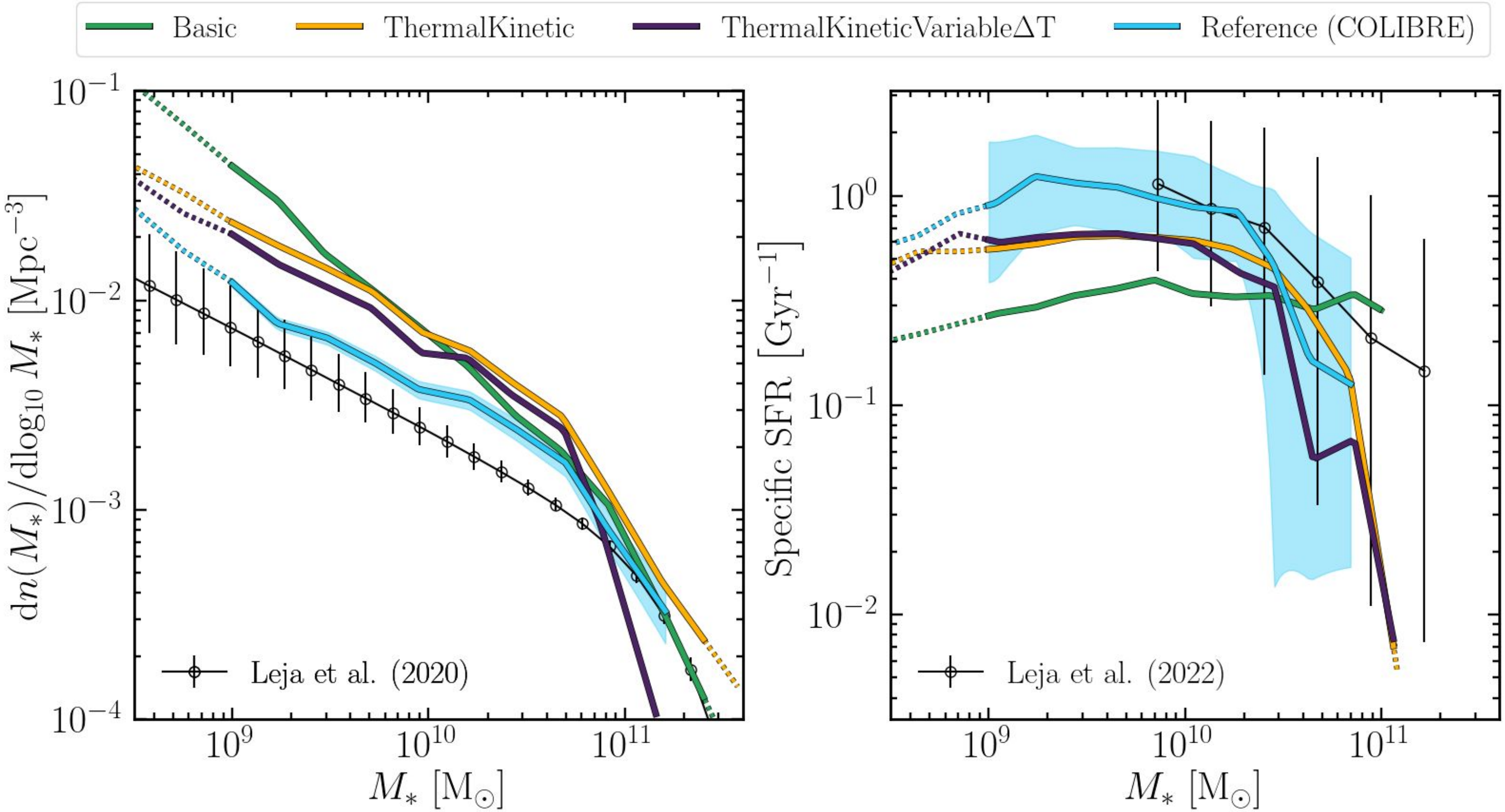


- The use of the variable heating temperature in **ThermalKineticVariableΔT** and **Reference** greatly improves the agreement with the data at low z
- The inclusion of a stellar birth pressure-dependent SN energy in **Reference** results in a lower SFRD at z>2 thereby further improving the agreement

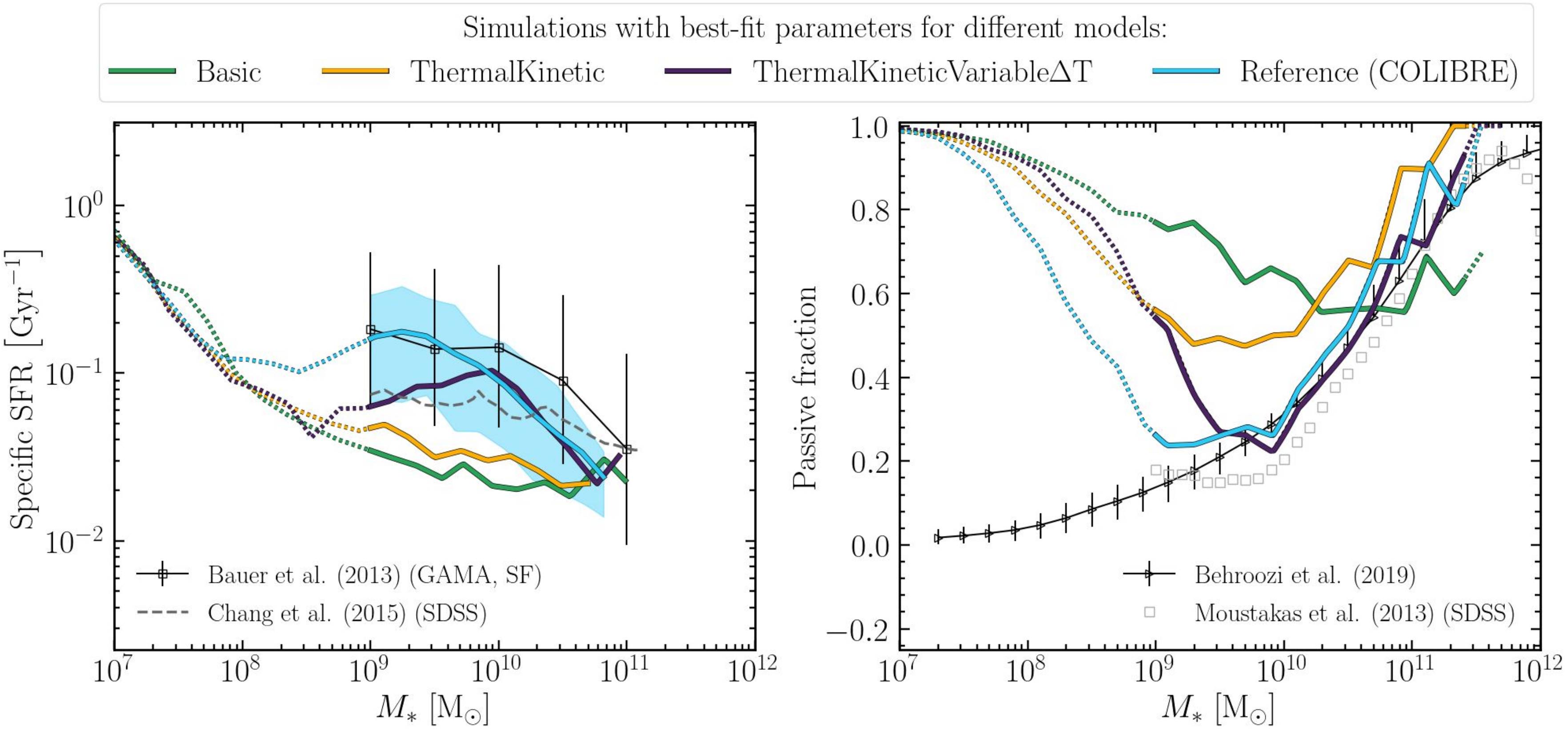
Atomic and molecular gas at z=0

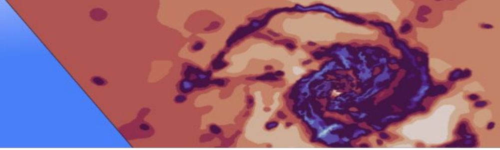


Comparison with data at z=2



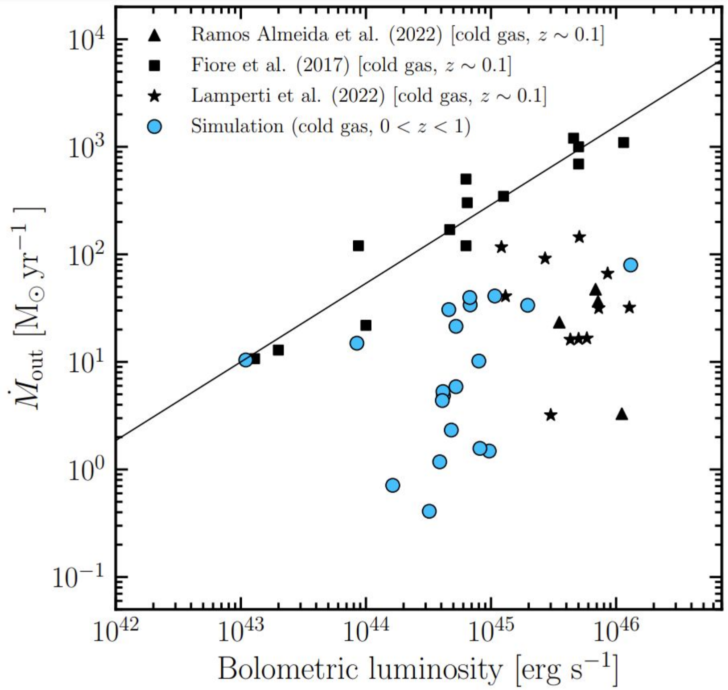
Specific star formation rates and passive fractions

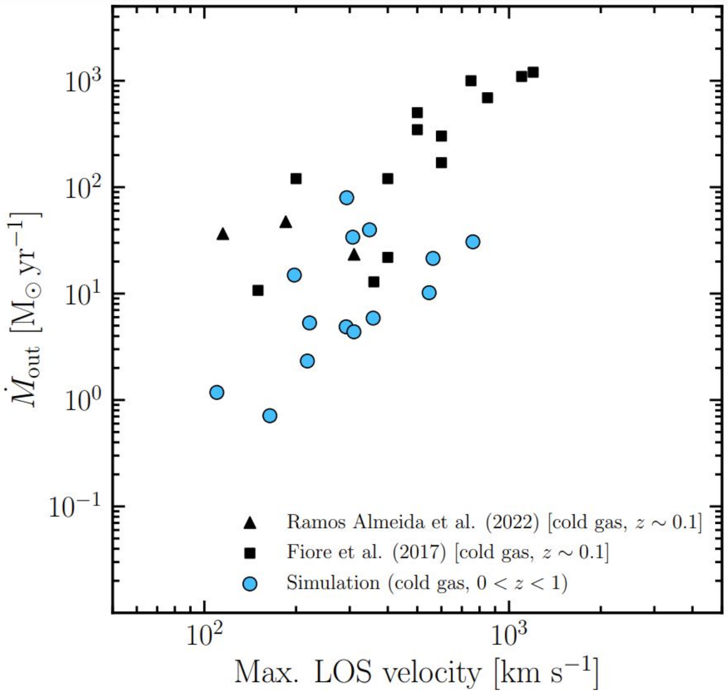




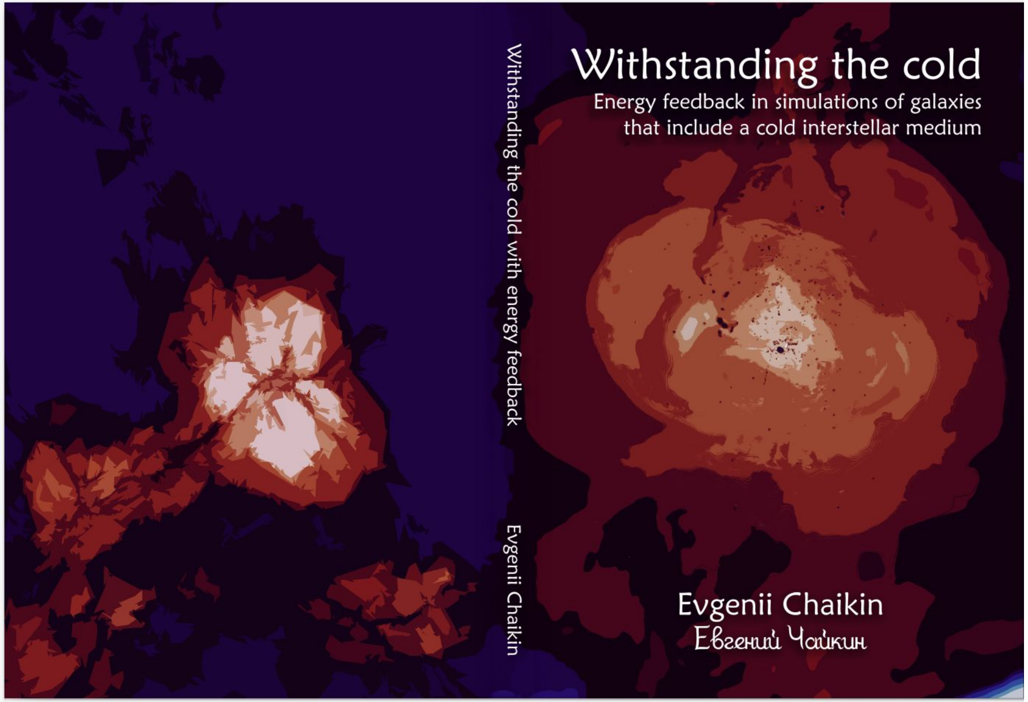
An example of applications for AGN studies

Comparing the properties of AGN-driven winds: simulations vs. observations (together with Blessing Musiimenta and Giovanna Speranza)





Thesis is finished



Summary

We used Gaussian-process emulators to calibrate the COLIBRE subgrid feedback based on the observed z=0 galaxy stellar mass function and size-stellar mass relation

- We considered four prescriptions for SN feedback with different complexities
- We used emulators to optimize one AGN feedback parameter and up to 3 SN feedback parameters
- We demonstrated how to rule out models lacking the complexity to fit the data
- We showed that *if and only if* a model successfully reproduces observed galaxy stellar masses and sizes, does it also reproduce galaxy properties to which it was not calibrated

Likelihood

We construct the log likelihood for the GSMF and SSM as

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

- $\{x_{\mathrm{R},n}, y_{\mathrm{R},n}, \sigma_{\mathrm{R},n}\}$ are the observational data points (x, y, and errors on y)
- $\hat{f}_{\mathrm{R}}(x, \boldsymbol{\theta})$ is the prediction of the trained emulator at x for the subgrid parameters $\boldsymbol{\theta}$
- $\varepsilon_{\mathrm{emu}}$ is the emulator error

The total log likelihood is

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \ln \mathcal{L}_{\mathrm{GSMF}}(\boldsymbol{\theta}) + \ln \mathcal{L}_{\mathrm{SSM}}(\boldsymbol{\theta})$$

Star formation rates, passive fractions, and BHs