

Cosmological baryon spread and impact on matter clustering in CAMELS

Matthew Gebhardt ,  1★ Daniel Anglés-Alcázar, ,  1,2 Josh Borrow ,  3,4 Shy Genel, ,  2,5 Francisco Villaescusa-Navarro, ,  2,6 Yueying Ni ,  7,8 Christopher C. Lovell ,  9,10 Daisuke Nagai ,  11 Romeel Davé ,  12,13 Federico Marinacci ,  14,15 Mark Vogelsberger ,  3,16 and Lars Hernquist 

¹Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA

²Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA

³Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁴Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

⁵Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA

⁶Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544 USA

⁷Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

⁸McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

⁹Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Road, Portsmouth PO1 3FX, UK

¹⁰Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK

¹¹Department of Physics, Yale University, New Haven, CT 06520, USA

¹²Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

¹³Department of Physics & Astronomy, University of the Western Cape, 7535 Cape Town, South Africa

¹⁴Dipartimento di Fisica e Astronomia 'Augusto Righi', Universit'a di Bologna, via Gobetti 93/2, I-40129 Bologna, Italy

¹⁵INAF, Astrophysics and Space Science Observatory Bologna, Via P. Gobetti 93/3, I-40129 Bologna, Italy

¹⁶The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Accepted 2024 March 14. Received 2024 March 8; in original form 2023 July 21

ABSTRACT

We quantify the cosmological spread of baryons relative to their initial neighbouring dark matter distribution using thousands of state-of-the-art simulations from the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project. We show that dark matter particles spread relative to their initial neighbouring distribution owing to chaotic gravitational dynamics on spatial scales comparable to their host dark matter halo. In contrast, gas in hydrodynamic simulations spreads much further from the initial neighbouring dark matter owing to feedback from supernovae (SNe) and active galactic nuclei (AGN). We show that large-scale baryon spread is very sensitive to model implementation details, with the fiducial SIMBA model spreading ~ 40 per cent of baryons > 1 Mpc away compared to ~ 10 per cent for the IllustrisTNG and ASTRID models. Increasing the efficiency of AGN-driven outflows greatly increases baryon spread while increasing the strength of SNe-driven winds can decrease spreading due to non-linear coupling of stellar and AGN feedback. We compare total matter power spectra between hydrodynamic and paired N -body simulations and demonstrate that the baryonic spread metric broadly captures the global impact of feedback on matter clustering over variations of cosmological and astrophysical parameters, initial conditions, and (to a lesser extent) galaxy formation models. Using symbolic regression, we find a function that reproduces the suppression of power by feedback as a function of wave number (k) and baryonic spread up to $k \sim 10 h \text{ Mpc}^{-1}$ in SIMBA while highlighting the challenge of developing models robust to variations in galaxy formation physics implementation.

Key words: galaxies: evolution – galaxies: formation – cosmology: large-scale structure of Universe.

1 INTRODUCTION

Investigating the distribution of matter in the Universe reveals many clues about its origin, content, and fate. Cosmological parameters such as the density of matter (Ω_m) and the present-day linear amplitude of matter fluctuations (σ_8) can be constrained by comparing theoretical predictions to observations from the cosmic microwave background (Planck Collaboration 2020), galaxy clustering (Cole

et al. 2005; Eisenstein et al. 2005), and weak lensing surveys (Hadzhiyska et al. 2021; Huang et al. 2021). In this new age of precision cosmology, simulations have become extremely valuable in the pursuit of tighter constraints on cosmological parameters by comparing their outputs to these surveys. As the next generation of surveys [e.g. CMB-S4¹ (Abazajian et al. 2016), DESI² (DESI

* E-mail: matthew.gebhardt@uconn.edu

¹<https://cmb-s4.org>

²<https://www.desi.lbl.gov>

Collaboration 2016), eROSITA³ (Merloni et al. 2012), Euclid⁴ (Euclid Collaboration 2022), and Rubin Observatory⁵ (Ivezić et al. 2019)] provide greater statistical power via larger volumes and greater sensitivity, cosmological simulations must follow suit. As their resolution increases, however, simulations must model smaller scales at which matter clustering can no longer be explained purely by gravitational dynamics. At such scales, processes such as radiative cooling, galactic winds driven by supernovae (SNe), and active galactic nuclei (AGN) feedback play an important role in the evolution of galaxies and directly redistribute baryonic matter over a range of scales (Anglés-Alcázar et al. 2017b; Borrow, Anglés-Alcázar & Davé 2020), which can provide an important source of contamination when extracting information from cosmological surveys (van Daalen et al. 2011; Chisari et al. 2019; Schaye et al. 2023). Unfortunately, many key physical mechanisms in galaxy formation are still not well understood, and so it is a challenge to decouple astrophysical processes from the intrinsic effects of fundamental cosmological parameters on the matter distribution. The uncertainties and computational costs of these baryonic processes relegate their implementation in large-volume hydrodynamic simulations to extensively tuned free parameters in subgrid models (Somerville & Davé 2015). To extract the maximum amount of cosmological information from future surveys, the effects and uncertainties of these processes must be well accounted for.

Dark matter only ('*N*-body') simulations have seen great successes in reproducing the over-arching large scale structure of the Universe and achieving the large volumes (at sufficient resolution) required for comparisons to cosmological surveys (Springel et al. 2005; Klypin, Trujillo-Gomez & Primack 2011; Angulo et al. 2012). However, while the dark matter component is responsible for the majority of the gravitational potential to form large structures, baryonic matter is subject to various astrophysical processes and, as a result, does not simply follow the dark matter (Naab & Ostriker 2017; Vogelsberger et al. 2020). There have been a wide range of efforts to create models that approximate the effects of baryons in such simulations. Empirical models (e.g. Berlind & Weinberg 2002; van den Bosch et al. 2007; Behroozi, Conroy & Wechsler 2010) are computationally efficient and map observable properties of baryons to dark matter haloes without any explicit modeling of baryonic processes. Semi-analytical models (SAMs) are a more physically motivated approximation method (e.g. Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Croton et al. 2006; Guo et al. 2011) that predicts galaxy properties given simulated dark matter halo merger trees by solving bulk equations to track quantities such as gas accretion onto haloes, star formation rates, or gas ejected from galaxies (Baugh 2006; Somerville & Davé 2015), but still do not predict the total matter distribution in and around galaxies. Cosmological hydrodynamic simulations (e.g. Hirschmann et al. 2014; Schaye et al. 2015; Davé, Thompson & Hopkins 2016; Weinberger et al. 2017; Davé et al. 2019) are the most direct way of modeling the impact of baryonic physics on galaxy evolution and the total matter distribution, but suffer from uncertainties in baryonic physics models.

The predicted abundance, clustering, and concentration of dark matter haloes differs between *N*-body and hydrodynamic simulations (Cui, Borgani & Murante 2014; Cui et al. 2016; Beltz-Mohrmann & Berlind 2021; Lu & Haiman 2021; Sorini et al. 2022), and connecting

these can approximate the predictive power of hydrodynamic simulations. Two methods for such an approximation are 'halo models', which alter the radial density profiles of haloes in *N*-body simulations when calculating the total matter power spectrum to match that of hydrodynamic simulations (e.g. Seljak 2000; Semboloni, Hoekstra & Schaye 2013; Mead et al. 2015) and 'baryonification' methods, which go a step further to actually alter the 3D distribution particles to match the halo density profiles found in hydrodynamic simulations (e.g. Schneider & Teyssier 2015; Schneider et al. 2019; Weiss et al. 2019). Halo models are also used in modeling (e.g. Shaw et al. 2010; Osato & Nagai 2023) and interpreting Sunyaev-Zeldovich (SZ) surveys (e.g. Reichardt et al. 2012; Osato et al. 2018, 2020). However, additional cluster astrophysics, such as the feedback and baryonic effects, must be understood better to realize the statistical power of upcoming SZ surveys (Chisari et al. 2019).

In practice, to compare theory to observations, one typically computes a 'summary statistic', such as the matter power spectrum (e.g. Hildebrandt et al. 2017; Hikage et al. 2019), which describes how matter is clustering at different spatial scales. Relative to *N*-body simulations, matter power in hydrodynamic simulations is increased at smaller scales by radiative cooling and star formation, but is also broadly decreased by feedback processes inhibiting the clustering of matter (Chisari et al. 2018, 2019; van Daalen, McCarthy & Schaye 2020; Delgado et al. 2023). At larger scales in particular, feedback appears to play an important role in decreasing power, which has been supported by observations showing that stellar (Lynds & Sandage 1963; Madau et al. 1996; Martin 1998; Pettini et al. 2001) and AGN (Feruglio et al. 2010; Sturm et al. 2011; Fabian 2012; Greene, Zakamska & Smith 2012; Cicone et al. 2014) feedback-driven outflows are capable of ejecting gas significant distances away from dark matter haloes. Though not an exhaustive list, the above effects alone can significantly alter the distribution of matter as compared to an *N*-body simulation, which further complicates efforts to account for the effects of baryons in such simulations.

One strategy to illuminate these complex feedback processes and to perhaps bypass the need to tightly constrain them is being carried out by the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project⁶ (Villaescusa-Navarro et al. 2021c). CAMELS contains thousands of hydrodynamic and *N*-body simulations with wide ranging variations of cosmological and subgrid feedback parameters. Using the large library of simulations, CAMELS data have been used to account for these uncertain feedback processes in a variety of ways. Promising results have arisen from attempts to predict cosmological parameters while marginalizing over astrophysical effects (Villaescusa-Navarro et al. 2020, 2021a, b; Perez et al. 2022; Shao et al. 2022; Villanueva-Domingo & Villaescusa-Navarro 2022; de Santi et al. 2023), estimate the mass of dark matter haloes from baryonic properties (Villanueva-Domingo et al. 2021d, 2022b) constrain subgrid feedback parameters, (Moser et al. 2022; Thiele et al. 2022; Pandey et al. 2023; Tillman et al. 2023), search for other summary statistics that may contain valuable cosmological information (Nicola et al. 2022; Villaescusa-Navarro et al. 2022a), reduce the scatter of scaling relations (Wadekar et al. 2023a, b), and more.

The hydrodynamic simulations in CAMELS include three different galaxy formation models: IllustrisTNG (Weinberger et al. 2017; Pillepich et al. 2018), SIMBA (Davé et al. 2019), and ASTRID (Bird et al. 2022; Ni et al. 2022). In CAMELS, the strengths of SNe and AGN feedback parameters as prescribed by the respective models

³<https://www.mpe.mpg.de/eROSITA>

⁴<https://www.euclid-ec.org>

⁵<https://www.lsst.org>

⁶<https://www.camel-simulations.org/>

are varied, which allows for systematic analysis of the effects that these processes have on, for example, the cosmic star formation rate history, the galaxy stellar mass function, and the large-scale distribution of matter as a whole.

In this work, we take advantage of the systematic model variations in CAMELS to quantify how far both dark and baryonic matter spread apart as a function of cosmological and feedback parameters by means of the Lagrangian matter *spread metric*. The spread metric was introduced in Borrow et al. (2020) and was used to quantify the redistribution of matter in the SIMBA cosmological simulation. It was shown that 40 per cent of the baryonic content of the simulated volume can spread more than $1 \text{ Mpc } h^{-1}$ away from the initial neighbouring dark matter distribution owing to the impact of large-scale AGN jets in SIMBA (see also Davé et al. 2019; Christiansen et al. 2020). Here, we present a detailed analysis of cosmological baryon spread including thousands of galaxy formation model variations in CAMELS, including the SIMBA, IllustrisTNG, and ASTRID implementations. Additionally, because feedback suppresses the matter power spectrum on large scales, we extend this analysis to investigate how the spreading of baryons correlates with the impact of feedback on the power spectrum.

This paper is organized as follows: In Section 2, we describe the CAMELS project, the data sets used, the spread metric, and other analysis techniques. In Section 3, we describe the results of analyzing the spread of dark matter and baryons in the SIMBA suite, as well as the correlation between cosmological baryonic spread and the total matter power spectrum. We also extend this analysis to simulations from the IllustrisTNG and ASTRID suites. In Section 4, we discuss the significance of these results in the context of the current state of the field. Finally, in Section 5, we summarize the conclusions of this work.

2 METHODOLOGY

For this work, we focus first on presenting a detailed study of cosmological matter spread using the SIMBA simulation suite in CAMELS, which we will then compare to the IllustrisTNG and ASTRID simulation suites. Our simulations and relevant analysis techniques are described below.

2.1 CAMELS Simulations

CAMELS (Villaescusa-Navarro et al. 2021c) is a collection of 5516 (magneto)hydrodynamic cosmological simulations with sub-grid physics from the IllustrisTNG, SIMBA, and ASTRID galaxy formation models, and 5164 *N*-body simulations with matching initial conditions, each with a comoving volume of $(25 \text{ Mpc } h^{-1})^3$ containing 256^3 dark matter particles evolving from $z = 127$ to present day. Dark matter particles have a mass of $6.49 \times 10^7 (\Omega_m - \Omega_b)/0.251 h^{-1} \text{ M}_\odot$. Hydrodynamic simulations contain an additional 256^3 gas particles (which may form stars and black holes) with an initial mass of $1.27 \times 10^7 \Omega_b/0.049 h^{-1} \text{ M}_\odot$. The (sub)halo catalogs used in this work are generated with SUBFIND (Springel et al. 2001).

The SIMBA galaxy formation model builds on the MUFASA model (Davé et al. 2016), and uses the ‘Meshless Finite Mass’ mode of the *N*-body and hydrodynamics code GIZMO (Hopkins 2015). The gravitational dynamics is solved with a Tree-PM method adapted from the GADGET-III code (Springel 2005). Radiative cooling and photoionization are implemented using Grackle-3.1 (Smith et al. 2016). Stellar feedback is modeled similar to that of MUFASA: two-phase galactic winds with 30 per cent of wind particles being ejected with a temperature set by the difference in SNe energy and wind

kinetic energy. As an update from MUFASA, the mass loading factor and velocity of galactic winds scale with stellar mass following the relations found from FIRE zoom-in simulations (Muratov et al. 2015; Anglés-Alcázar et al. 2017b). Supermassive black hole (SMBH) growth is implemented in two phases, where the gravitational torque accretion model (Hopkins & Quataert 2011; Anglés-Alcázar, Özel & Davé 2013; Anglés-Alcázar et al. 2015, 2017a) is used for cold gas and the Bondi accretion model (Bondi 1952) is used for hot gas. Feedback from AGN is comprised of mechanical quasar-mode winds and high-speed collimated jets at fixed momentum flux following Anglés-Alcázar et al. (2017a) and X-ray feedback following Choi et al. (2012). For a more thorough description of SIMBA, see Davé et al. (2019).

The IllustrisTNG galaxy formation model builds on the Illustris model (Genel et al. 2014; Vogelsberger et al. 2014a), and uses the Arepo code (Springel 2010) with the Tree-PM method to solve the equations of gravity and a Voronoi moving-mesh method to solve for magnetohydrodynamics. Radiative cooling follows Katz, Weinberg & Hernquist (1996), Wiersma et al. (2009), and Rahmati et al. (2013). Stellar feedback galactic winds follow a kinetic scheme based on Springel & Hernquist (2003) in which particles are stochastically and isotropically ejected from star-forming gas. The SMBH model builds upon Springel (2005), Sijacki et al. (2007), and Vogelsberger et al. (2013), with SMBH mergers occurring when SMBH particles enter each other’s ‘feedback spheres’, and with gas accretion following the Eddington rate-limited Bondi parameterization (Bondi 1952). AGN feedback is implemented in three modes: thermal, kinetic, and radiative. The high accretion (thermal) mode injects thermal energy at a rate proportional to the mass accretion rate into a ‘feedback sphere’ around the SMBH, while the low accretion (kinetic) mode accumulates energy over time and injects kinetic energy in a random direction into the feedback sphere when a total amount of energy since last injection is accumulated. The radiative component is always active, and adds the SMBH’s radiation flux to the cosmic ionizing background. IllustrisTNG is fully described in Weinberger et al. (2017) and Pillepich et al. (2018).

The ASTRID galaxy formation model is implemented in the MP-Gadget simulation code (expanded from GADGET-III), using the Tree-PM method to solve for gravity and the smoothed particle hydrodynamics (SPH) method. Radiative cooling and heating are modeled from Katz et al. (1996), Vogelsberger et al. (2013), Faucher-Giguère (2020), and Rahmati et al. (2013). Stellar feedback is implemented kinetically, with wind particles sourced from newly formed star particles. The CAMELS version of ASTRID (Ni et al. 2023) uses the SMBH model adapted from the ASTRID production run Bird et al. (2022) and Ni et al. (2022), without explicit modelling of the BH dynamical friction, and is extended to include a two-mode SMBH feedback implementation using thermal or kinetic energy injection depending on the Eddington ratio of the SMBH accretion rate. The low accretion rate (kinetic) model follows Weinberger et al. (2017) but with different parameter values. The high accretion rate (thermal) model injects 5 per cent of the SMBH’s radiation energy into the surrounding gas within a spherical region with radius twice that of the SPH kernel. ASTRID is fully described in Bird et al. (2022) and Ni et al. (2022).

Throughout the CAMELS suites, cosmological and feedback parameters are varied in different ways across four sets of simulations. We first focus on variations of two cosmological parameters (Ω_m and σ_8 with fixed $\Omega_b = 0.049$), two parameters governing SNe feedback (A_{SN1} and A_{SN2}), and two parameters governing AGN feedback (A_{AGN1} and A_{AGN2}). We explore the implications of varying these parameters in the SIMBA model and how they affect the spreading of

matter. In SIMBA, the feedback parameters represent the following quantities:

- (i) $A_{\text{SN}1}$ – Mass loading factor of galactic winds.
- (ii) $A_{\text{SN}2}$ – Speed of galactic winds.
- (iii) $A_{\text{AGN}1}$ – Momentum flux of quasar and jet-mode feedback.
- (iv) $A_{\text{AGN}2}$ – Speed of jet-mode feedback.

Feedback parameters are simply normalizations of feedback strength relative to the original feedback models (e.g. a SIMBA simulation with $A_{\text{SN}2} = 2$ will have twice the galactic wind speed as the original SIMBA subgrid model). Fiducial values of these six parameters are taken to be $\Omega_m = 0.3$, $\sigma_8 = 0.8$, and $A_{\text{SN}1} = A_{\text{SN}2} = A_{\text{AGN}1} = A_{\text{AGN}2} = 1$. Each parameter variation range is as follows: $0.1 \leq \Omega_m \leq 0.5$; $0.6 \leq \sigma_8 \leq 1.0$; $0.25 \leq (A_{\text{SN}1}, A_{\text{AGN}1}) \leq 4.00$; and $0.5 \leq (A_{\text{SN}2}, A_{\text{AGN}2}) \leq 2.0$. After a focused analysis of the effects variations on these six parameters have on the spreading of dark matter and gas in SIMBA, we will explore a broader range of up to 28 parameter variations (see Ni et al. 2023 for descriptions of all varied parameters) comparing the baryon spreading in SIMBA, IllustrisTNG, and ASTRID.

For each simulation, the total matter power spectrum $P(k)$ is calculated following Villaescusa-Navarro et al. (2021c). The masses for every particle type (dark matter, gas, stars, and black holes) are placed in a 512^3 voxel grid, which is Fourier transformed by averaging over k bins. The bin width is the fundamental frequency, $k_F = 2\pi/L$, where L is the length of the simulated box, $25 \text{ Mpc } h^{-1}$.

2.2 Data sets

This analysis uses several different data sets within CAMELS. Sections 3.1–3.3 all use the six parameter 1P set, CV set, and LH set from the SIMBA suite presented in Villaescusa-Navarro et al. (2021c), while Section 3.4 uses the full 28 parameter 1P set from the SIMBA suite, the 1P and LH sets from the ASTRID suite, and the 28 parameter 1P and SB28 sets from the IllustrisTNG suite presented in Ni et al. (2023). A short description for each simulation set is given below:

(i) The 1P ('one parameter') set contains simulations with the same initial conditions (random seed) but varying one parameter at a time. One simulation uses all fiducial parameter values while the remaining simulations correspond to variations of up to 28 parameters above and below their fiducial value while all others are held constant. The six parameters described above each have 10 variations (hereafter referred to as the six parameter 1P set), while the remaining 22 parameters each have 4 variations. The full list and descriptions of all 28 parameters can be found in Ni et al. (2023). The parameter variation spacing is linear for cosmological parameters and logarithmic for feedback parameters. The corresponding N -body 1P set contains 21 simulations varying only Ω_m and σ_8 . The 1P set is designed for determining the effects of specific parameters on various quantities such as the matter spread.

(ii) The CV ('cosmic variance') set contains 27 simulations with the same fiducial parameters, but different initial conditions. The CV set is used to quantify the effects of cosmic variance on observables, which is important given that the simulated volumes in CAMELS are small and not representative of the Universe as a whole.

(iii) The LH ('Latin Hypercube') set contains 1000 simulations each with different initial conditions, and with near-random parameters selected from a Latin hypercube. The main goal of the LH set is to train machine learning algorithms to make predictions given

cosmological and astrophysical inputs, account for cosmic variance, and marginalize over baryonic effects.

(iv) The SB28 ('Sobol Sequence') set is unique to IllustrisTNG and contains 1024 simulations with 28 parameters quasi-randomly selected from a Sobol sequence designed for machine learning applications in an expanded parameter space (Ni et al. 2023).

2.3 The spread metric

One way to quantify the decoupling of the baryonic and dark matter components is with the cosmological *spread metric* (Borrow et al. 2020). The spread metric for a particle (either gas or dark matter) is defined as the final distance at $z = 0$ between that particle and the dark matter particle it was nearest to in the initial conditions. For the spread of gas in IllustrisTNG, we use tracer particles that are designed to accurately follow the flow of gas (see Genel et al. 2013 for a full description) rather than tracking gas cells. For SIMBA and ASTRID, gas particles that turn into stars are not included in this analysis. The method for calculating the spread of a particle (which is the same for both gas and dark matter) is as follows:

- (i) In the initial conditions, find the particle's nearest dark matter neighbour by computing the distance to all dark matter particles.
- (ii) Store the IDs of these two neighbouring particles.
- (iii) By ID matching, find these particles in the $z = 0$ snapshot and compute the distance between them.

Only the initial conditions and final state of the simulation are required; here we focus on the spread of matter from $z = 127$ down to $z = 0$. The distributions of dark matter and gas particles are identical at $z = 127$ except for the systematic displacement of the staggered grids that define the initial conditions. The neighbouring dark matter particle is used as a reference point in this metric due to the ambiguity involved in measuring absolute distances in an evolving universe, where the gas and dark matter components in large structures can drift owing to peculiar motions. The spread metric is thus a measure of displacement relative to the initial surrounding dark matter distribution normalizing out peculiar motions. In this work, we calculate the 'spread' value for every gas and dark matter particle in all CAMELS simulations. The spread metric can also be computed for star particles, but here we focus on the spread of gas and refer to it as baryon spread interchangeably. With the spread output, particles may be selected and analyzed in various ways, such as investigating the spread of particles inside or outside of haloes, particles within a specified halo mass range, or particles within a specified spread percentile range. The median spread of particles for a given simulation can be computed to characterize with a single statistic the overall impact of baryonic effects on the distribution of matter.

2.4 Symbolic regression

Symbolic regression is a machine learning technique capable of finding an analytic mathematical formula to relate an input and an output. A distinct advantage of symbolic regression over a more standard 'least squares' regression is that the functional form does not need to be known ahead of time, and instead is constructed from a set of allowed operators. We use symbolic regression to find a function that can predict the impact of baryonic feedback on the matter power spectrum as a function of baryon spread and wave number k . We use the PySR package (Cranmer 2020) to train on SIMBA six parameter 1P set simulations that vary the four feedback parameters where the input is a 2D array with one axis consisting of median gas

spread values for each simulation and the other consisting of 40 k values between 0.5 and $10 h \text{ Mpc}^{-1}$. The output is the (negative) change in total matter power spectrum between the hydrodynamic and N -body simulations, $-\Delta P/P = -(P_{\text{Hydro}} - P_{\text{Nbody}})/P_{\text{Nbody}}$. The allowed operators are addition, multiplication, subtraction, division, exponential, inverse, absolute value, and square root. We use the loss function default to PySR, mean squared error:

$$f_{\text{loss}} = \frac{1}{N} \sum_i^N (\Delta P/P_i^{\text{true}} - \Delta P/P_i^{\text{predicted}})^2, \quad (1)$$

where N is the total number of data points. We use the measure of complexity default to PySR, which treats each operator with equal complexity. Finally, we use 100 iterations in each search.

3 RESULTS

In this section, we use the spread metric to analyze the redistribution of dark matter in N -body simulations and gas in SIMBA hydrodynamic simulations to evaluate how these change with varying cosmological parameters and (in the case of gas) feedback strength. We then investigate the correlation between the spread of gas and the impact of baryons on the total matter power spectrum. We conclude this section with comparisons between the baryon spread and impact on the matter power spectrum in SIMBA, IllustrisTNG, and ASTRID.

3.1 Dynamical spread in N -body simulations

All particles are subject to the force of gravity and may spread owing to chaotic gravitational dynamics. We begin by analyzing the spreading of dark matter in N -body simulations, which will serve as control for our subsequent analysis of baryonic spread to understand the relative roles of gravity and baryonic physics on the redistribution of matter.

Using the N -body 1P set, we can investigate how the spreading of dark matter depends on cosmological parameters. Fig. 1 shows the distribution of the spread of all dark matter particles in each simulation as the cosmological parameters Ω_m and σ_8 are varied. Each line corresponds to a different simulation and the colour denotes the parameter value. We see here that higher parameter values for Ω_m and σ_8 correspond to greater spread of dark matter (more particles are spreading farther from their initial neighbours). At the greatest values of Ω_m , some particles are spreading more than $7 \text{ Mpc } h^{-1}$, whereas the farthest spread particles in the lowest Ω_m run spread less than $3 \text{ Mpc } h^{-1}$. Variations in σ_8 show a slightly tighter distribution, with the farthest spread particles in the highest σ_8 run spreading just over $6 \text{ Mpc } h^{-1}$, and around $3 \text{ Mpc } h^{-1}$ in the lowest σ_8 run.

Dark matter halos represent gravitational potential wells where chaotic orbits can make the trajectories of dark matter particle neighbours diverge over time. The dark matter particles that spread the most are then likely to reside around the most massive structures in the simulation. It is thus informative to explore the relationship between spread, parameter value, and halo size. Fig. 2 shows the median spread of dark matter particles from the 10 most massive haloes (but excluding the most massive as it is often significantly larger than the others) as a function of their average virial radius for simulations varying Ω_m , with the shaded region representing the 25th–75th percentile of spread. This depicts a clear positive (sub-linear) correlation between Ω_m and the average virial radius, which increases from $150 \text{ kpc } h^{-1}$ to $395 \text{ kpc } h^{-1}$ with $\Omega_m = 0.1 \rightarrow 0.5$ and also correlates with an overall increase in the spread of particles. This suggests that increasing Ω_m and σ_8 yield wider spreading of

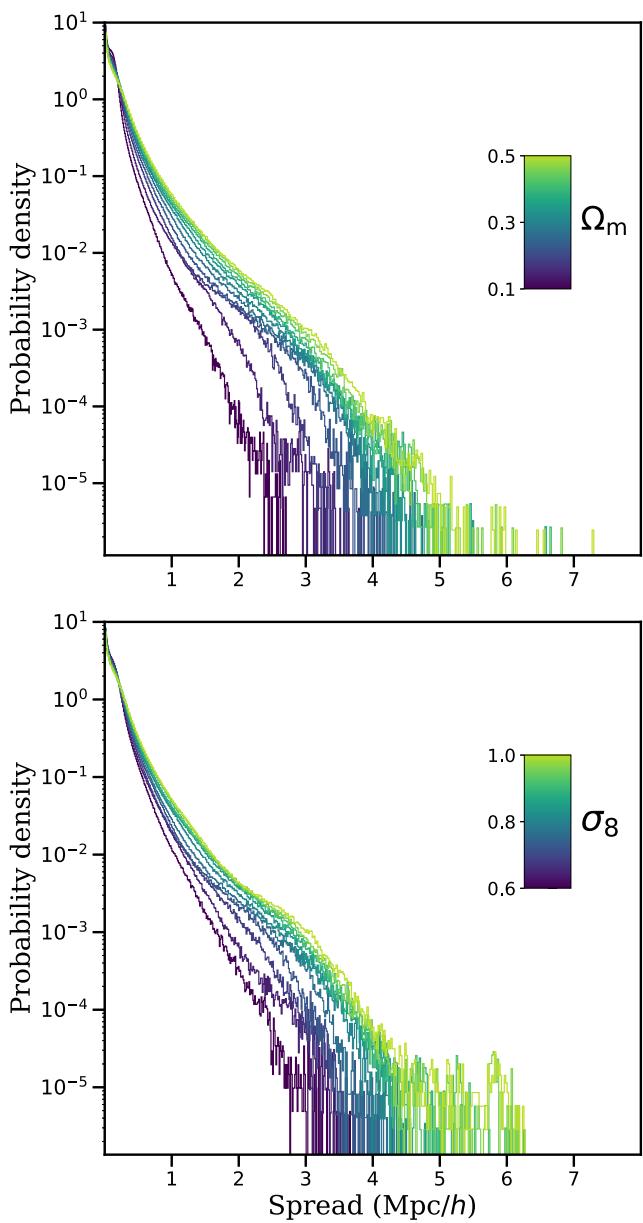


Figure 1. Distribution of spread distances for all dark matter particles in the 1P set N -body simulations at $z = 0$. The top panel shows the spread distribution in simulations varying Ω_m (with all other parameters held constant) and the bottom panel shows the impact of varying σ_8 alone. Colour bars indicate the parameter value for each simulation. Dark matter spreads further from the initial neighbouring distribution with increasing Ω_m and σ_8 .

matter by increasing the mass and abundance of massive halos (see Villaescusa-Navarro et al. 2021c).

To investigate further, Fig. 3 shows the median spread of particles residing within haloes of specified mass ranges as a function of parameter value. Regardless of cosmological parameter variation, particles in larger haloes spread farther. When looking at haloes of the same mass, the values of Ω_m and σ_8 have little impact on the median spread of dark matter. Dark matter spread does, however, slightly decrease at fixed halo mass with increasing Ω_m , which may be explained by an increase in halo concentration (see Section 4). The clear relationship between halo mass and dark matter spread provides support for the increased dark matter spread at higher Ω_m and σ_8 simply reflecting the formation of more massive haloes. This

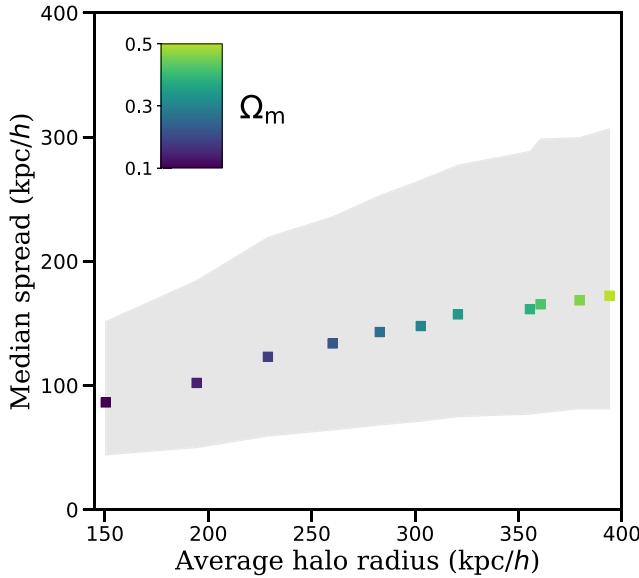


Figure 2. Median spread of all dark matter particles in the 10 most massive haloes (excluding the most massive halo) as a function of the average virial radius among these haloes at $z = 0$ in each of the N -body simulations varying Ω_m . The colour bar indicates the value of Ω_m for each simulation. As Ω_m increases, haloes get larger and spreading increases.

relationship is further supported by the full spread distribution of particles in these haloes, which shows a clear dependence on halo mass at fixed cosmology (see Appendix A).

3.2 Baryonic spread in cosmological hydrodynamic simulations

The spreading of gas becomes more complicated with the addition of radiative cooling, hydrodynamic forces, and galaxy formation feedback in cosmological hydrodynamic simulations. Fig. 4 shows the distribution of spread distances for all gas particles in each hydrodynamic simulation in the six parameter SIMBA 1P set, which now includes more simulations varying feedback parameters. As expected, the distance to which baryons spread relative to the initial neighbouring dark matter distribution increases with higher values of Ω_m and σ_8 , reflecting the formation of more massive halos and the spread of matter owing to chaotic gravitational dynamics as seen for N -body simulations (Fig. 1). Indeed, when investigating the median spread of gas particles in haloes as a function of Ω_m and halo radius, the trend is similar to that of dark matter, and this trend largely disappears when normalizing to the spread of each gas particle's initial dark matter neighbour (see Appendix B). There is a notable difference in the spread of baryons compared to dark matter, with gas particles spreading up to $\sim 11 \text{ Mpc } h^{-1}$ (compared to $\sim 4.5 \text{ Mpc } h^{-1}$ for dark matter) in the fiducial simulation.

Increasing the strength of AGN feedback systematically increases the spread of gas, with the maximum distance reached varying from $\sim 10\text{--}14 \text{ Mpc } h^{-1}$ when increasing either $A_{\text{AGN}1}$ or $A_{\text{AGN}2}$ from their minimum to maximum parameter values explored here. In contrast, increasing the strength of SNe feedback shows complex results. Greater mass loading of SNe-driven winds ($A_{\text{SN}1}$) yields an unclear, but perhaps minor positive correlation with spread, while greater speed of SNe-driven winds ($A_{\text{SN}2}$) shows a strong negative correlation. This indicates that in SIMBA, overly efficient SNe feedback reduces the spreading of gas to large scales, which can be explained by driving a net reduction of energy injection by

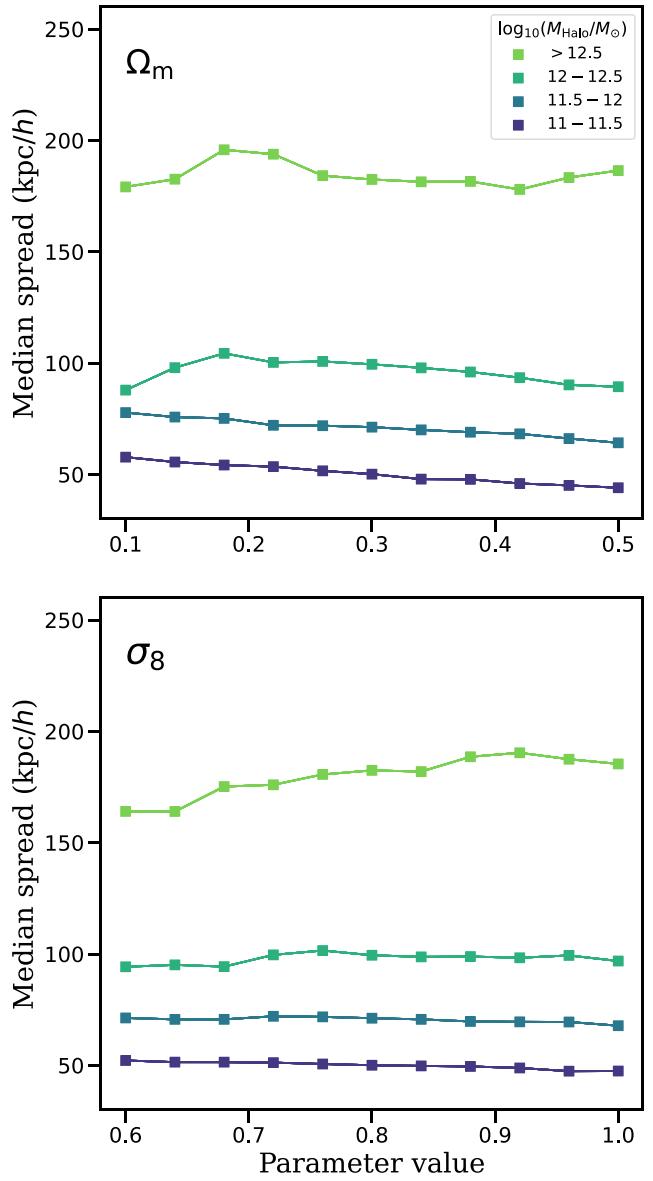


Figure 3. Median spread of dark matter particles inside of halos of selected mass ranges as a function of Ω_m (top) and σ_8 (bottom) in each of the N -body simulations at $z = 0$. Colour represents the halo mass range as described by the legend in the top panel. Binning by halo mass removes most of the dependence of spread on cosmological parameters and shows a clear positive correlation between halo mass and median spread.

AGN feedback due to the evacuation of gas from central areas (see Section 4). We note here that while these distances are comparable to the size of the simulated box in CAMELS, Borrow et al. (2020) found that the maximum gas spread distance in the fiducial $100 \text{ Mpc } h^{-1}$ SIMBA simulation was within half the CAMELS boxsize and thus should not affect results.

Fig. 5 compares the spread of dark matter and gas in different simulations in the form of cumulative distributions which more clearly show the amount of mass spreading up to a given distance. We compare the fiducial 1P simulation to the highest and lowest AGN jet speed models as well as the full spread distribution of the CV set. As expected, gas spreads significantly farther than dark matter (20 per cent of gas spreads farther than $1.8 \text{ Mpc } h^{-1}$ compared to $0.26 \text{ Mpc } h^{-1}$ for dark matter), and the spreading of gas increases

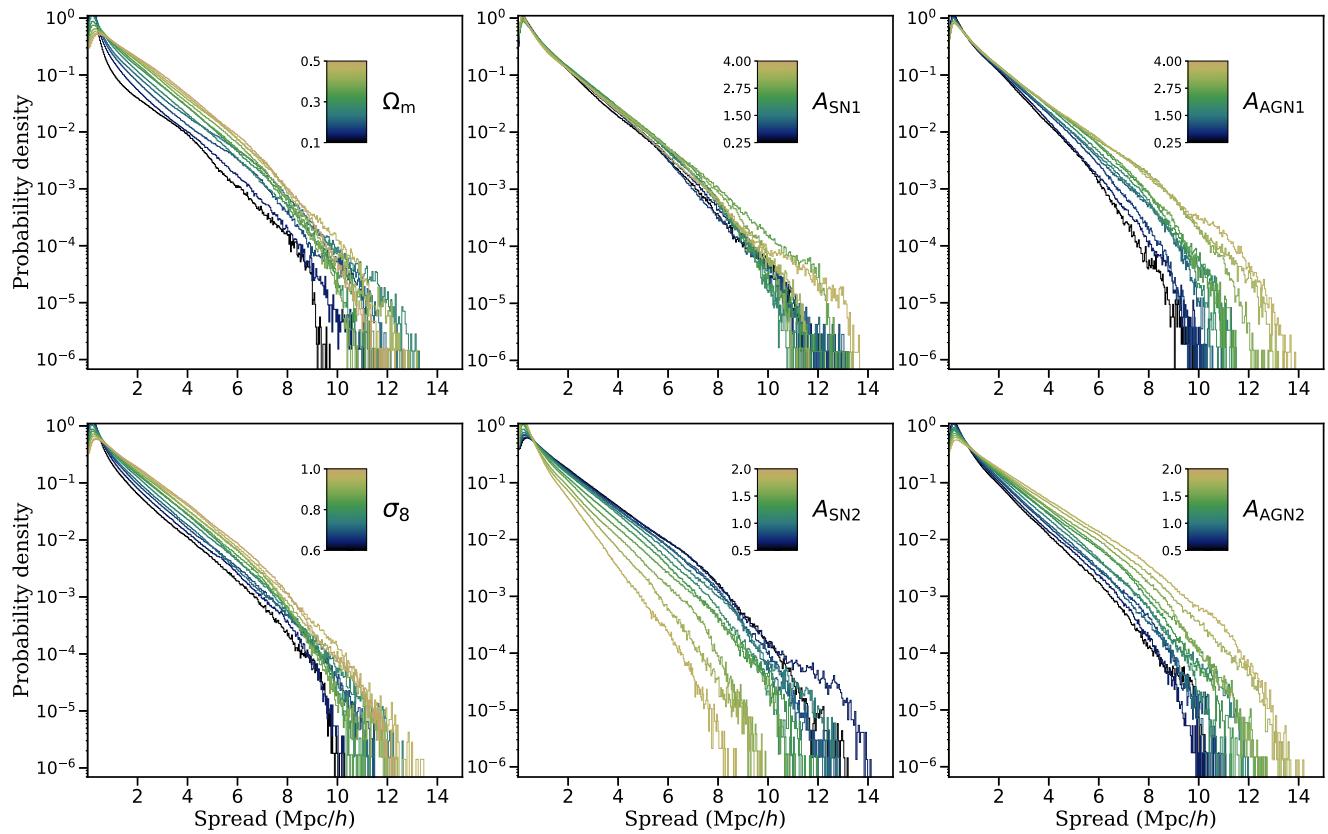


Figure 4. Distribution of spread distances for all gas particles in the six parameter SIMBA 1P set at $z = 0$. Each panel contains distributions corresponding to one simulation run in which all other parameters are held constant at fiducial values. Colour bars indicate parameter value for each simulation. For both cosmological parameters and AGN parameters, spread is clearly positively correlated, while it shows a minor positive correlation with $A_{\text{SN}1}$ (mass loading) and a strong negative correlation with $A_{\text{SN}2}$ (wind speed).

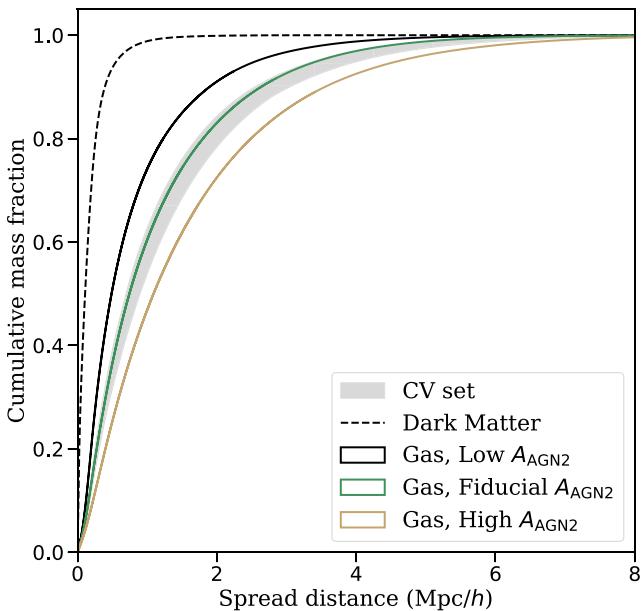


Figure 5. Cumulative mass distribution as a function of spread distance at $z = 0$ for all dark matter (dashed) and gas (green) particles in the fiducial SIMBA 1P simulation, for all gas particles in the highest (brown) and lowest (black) $A_{\text{AGN}2}$ (jet speed) simulations, and for all gas particles in the full SIMBA CV set (the grey shaded area spans the full CV set distribution). In all cases, gas spreads much further than dark matter and increases with AGN jet speed.

with faster jet speeds (20 per cent of gas spreads farther than $1.23 \text{ Mpc } h^{-1}$ with the lowest jet speed as compared to $2.5 \text{ Mpc } h^{-1}$ with the highest jet speed). The CV set shows the extent to which varying initial conditions can affect the large-scale spreading of material, which can be attributed to differences in the halo mass function and, in particular, the abundance of the most massive halos hosting SMBHs with powerful AGN jet feedback. In the CV set simulation with the least amount of gas spread, 20 per cent of gas spreads farther than $1.7 \text{ Mpc } h^{-1}$, while in the simulation with the highest gas spread, 20 per cent of gas spreads farther than $2.1 \text{ Mpc } h^{-1}$.

For gas, we expect the majority of spreading to occur around haloes, where galaxies act as sources of stellar and AGN feedback powering the ejection of gas to large distances. However, the spreading of gas in haloes becomes complex due to competing effects: feedback pushing gas outward and the ability for gas to radiate away energy (radiative cooling) and fall to lower radii in the gravitational potential well of dark matter haloes. Fig. 6 depicts this dichotomy by quantifying the spread of gas initially located inside of the Lagrangian regions of $z = 0$ halos at the initial conditions. For each halo at $z = 0$, we define its Lagrangian region by tracking the corresponding dark matter particles back to their location at the initial conditions. Gas particles are then defined to be in a Lagrangian region at the initial conditions if their nearest dark matter particle neighbour ends up in a halo at $z = 0$ (Borrow et al. 2020). All of the selected gas particles shown in Fig. 6 were initially located inside of a Lagrangian region, meaning that their nearest dark matter neighbour particle is inside of a halo at $z = 0$. We compute separately the median spread distance for Lagrangian region gas that ends up inside of halos (red

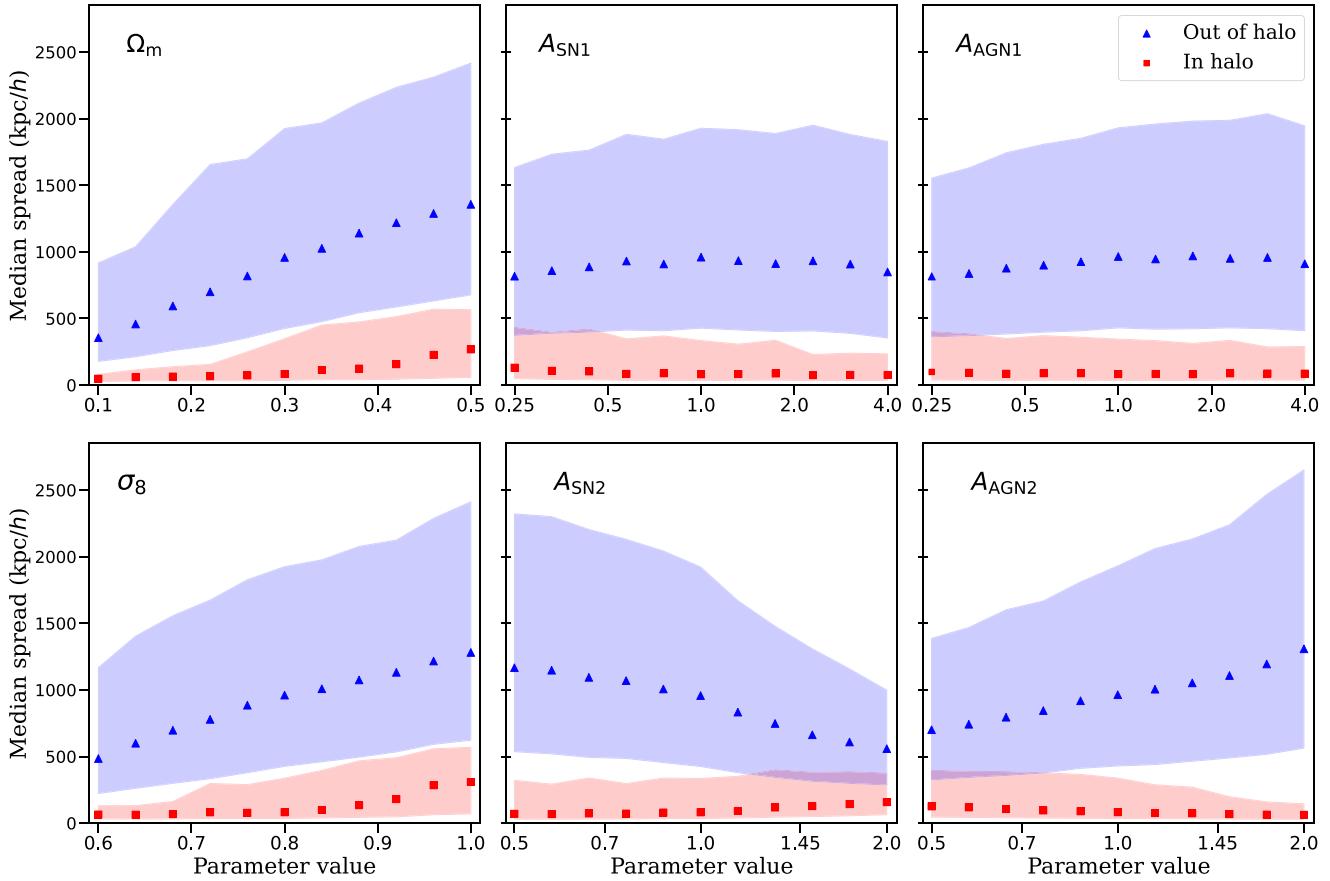


Figure 6. Cosmological spread of gas selected from halo Lagrangian regions at the initial conditions (i.e. gas particles whose nearest initial dark matter particle neighbour ends up inside of a halo at $z = 0$) as a function of cosmological and feedback parameters in SIMBA. Median spreads are shown separately for gas that ends up inside of the corresponding halo (squares, red shading) or outside of any halo (triangles, blue shading) at the end of the simulation ($z = 0$). Each panel corresponds to one individual parameter variation, as indicated. Shaded regions denote the 25th (lower) and 75th (upper) percentile of spread. The median spread of gas ejected from haloes is significantly larger with a stronger dependence on feedback parameters than for gas that remains inside of haloes.

squares) or outside of halos (blue triangles) at $z = 0$, and we indicate the 25th to 75th percentile range as shaded regions. Gas particles from Lagrangian regions that end up outside of halos spread much farther than the gas particles that remain in a halo regardless of parameter variations, as expected. Additionally, the spread of gas outside of haloes has a stronger dependence on parameter variations than gas that remains inside of halos. The dichotomy of baryonic spread seen here shows that Lagrangian regions contain baryons that will spread very little (dense gas converting into stars in the cores of dark matter halos) and very far (gas ejected in high-speed AGN jets).

Fig. 7 illustrates the spatial distribution of gas at $z = 0$ that has spread by different amounts within a given simulated volume, comparing simulations with different AGN jet speed in SIMBA. Each panel represents a 2D mass projection of 20 per cent of the gas particles in the simulation, and thus each panel contains the same amount of mass. Each row corresponds to one simulation, and each column denotes the percentile of spread of the particles being plotted. The $A_{\text{AGN}2}$ parameter (jet speed) is increased from top to bottom with values $A_{\text{AGN}2} \approx [0.5, 0.66, 1.0, 1.52, 2.0]$. In the low-spread panels (first column), gas particles trace both the densest regions at the centres of haloes as well as a diffuse component far from the feedback generated by the most massive halos. As the spread increases, gas particles trace regions around massive haloes and filaments at increasingly large distances from halo centres. The

contrast in the relative distributions of gas for different percentile ranges of spread is enhanced when increasing the AGN jet speed.

3.3 Impact of baryon spread on the matter power spectrum

Using the hydrodynamic and N -body simulation pairs in CAMELS, we can investigate the impact of baryonic effects on the total matter power spectrum as a function of feedback strength, cosmic variance, and cosmology. Fig. 8 shows the ratio of hydrodynamic and N -body power spectra for the six parameter SIMBA 1P set simulations that vary feedback parameters. If the distribution of matter was exactly the same in both simulations, the ratio of these power spectra would be $P_{\text{hydro}}/P_{\text{Nbody}} = 1$ at all values of k . Therefore, deviance from a ratio equaling 1.0 indicates baryonic impact on the matter power spectrum. At the smallest spatial scales, the ratio increases far above 1.0 due to radiative cooling of baryons and the formation of stars. On larger spatial scales, the ratio decreases below 1.0 due to feedback processes redistributing baryons and (to a much lesser extent) dark matter owing to back-reaction (Daalen et al. 2011; Chisari et al. 2019). To investigate the origin of these effects, we colour-code the $P_{\text{hydro}}/P_{\text{Nbody}}$ lines based on a normalized measure of the amount of baryonic spread for each hydrodynamic simulation. Motivated by results from van Daalen et al. (2020) showing that haloes with masses $M_{\text{halo}} > 10^{14} M_{\odot}$ dominate the power on scales $k > 1 h \text{ Mpc}^{-1}$, we

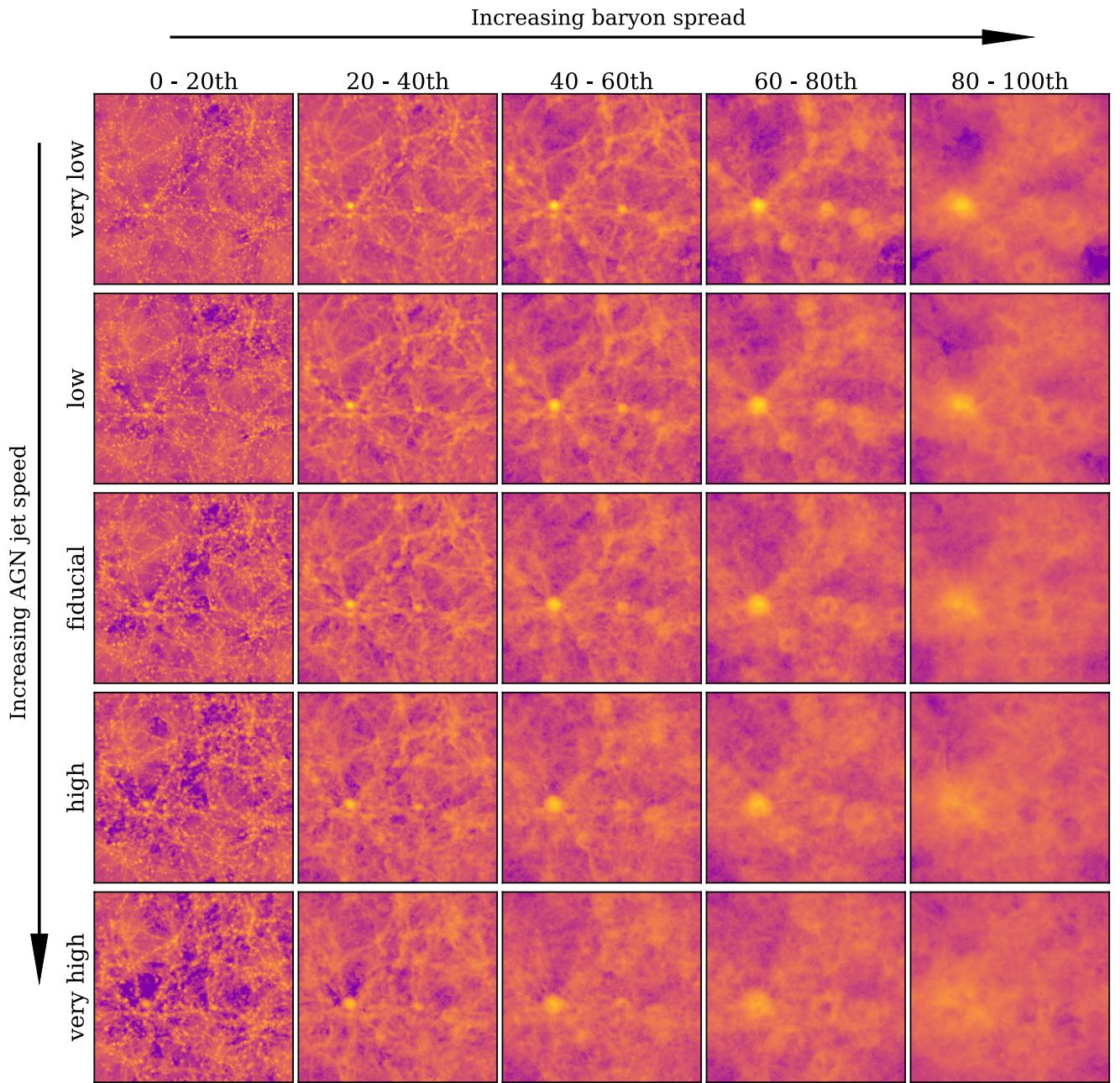


Figure 7. Spatial distribution of gas in a $(25 \text{ Mpc } h^{-1})^3$ volume as a function of AGN jet speed (A_{AGN2} ; rows) and relative amount of spread at $z = 0$ (columns) in SIMBA. Each panel in a given row is a 2D gas mass projection at $z = 0$ containing 20 per cent of the baryonic content of the simulation, ranked by amount of spread from left (lowest spread percentile range) to right (highest spread percentile range). From top to bottom, simulations with progressively higher A_{AGN2} parameter values are shown. Colour scale represents the mass density in the 2D projection and is logarithmic and identical in all panels. Particles of low spread percentile tend to be inside halos and filaments, while greatly spread particles appear to reside in bubbles around halos and filaments.

choose only gas particles that are in one of the five most massive Lagrangian regions in each simulation (due to the small box sizes in CAMELS, these ‘most massive’ regions correspond to haloes with masses between 10^{13} M_\odot and 10^{14} M_\odot , and our results are insensitive to the exact number of regions considered). Additionally, we normalize the spread of each gas particle by dividing by the spread of its dark matter neighbour, partially mitigating the effect of varying cosmological parameters on the halo mass function and thus the (gravitational) dynamic spreading of matter. Lastly, we compute the median of this normalized baryonic spread for all selected gas

particles to compare CAMELS simulations with different model parameters. For the six parameter SIMBA 1P set simulations varying feedback parameters (Fig. 8), we see a clear relationship between the impact of feedback on the clustering of matter and the spreading of gas particles across a range of scales, with greater spread correlating with greater reduction of power. This relationship continues to hold even at small scales ($k \sim 30 h \text{ Mpc}^{-1}$).

We next explore this relationship as a function of cosmic variance. Fig. 9 is similar to Fig. 8 but now showing $P_{\text{hydro}}/P_{\text{Nbody}}$ as a function of k for the SIMBA CV set. Given the small $(25 \text{ Mpc } h^{-1})^3$

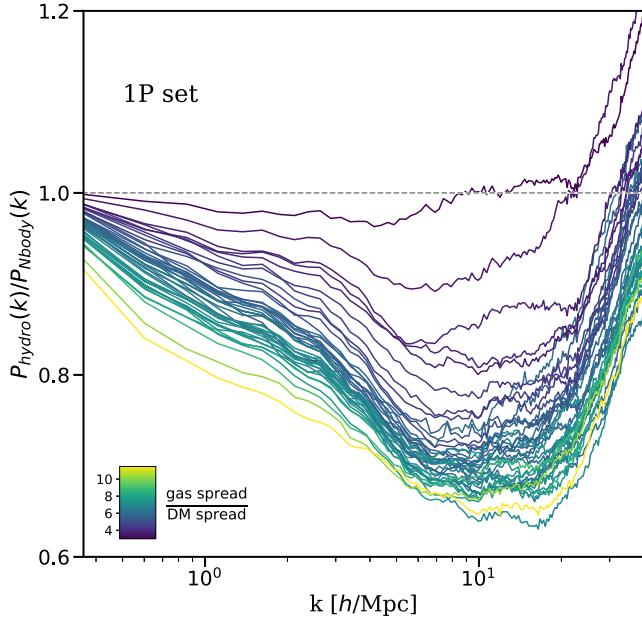


Figure 8. Ratio of hydrodynamic and N -body simulation power spectra for the 44 simulations from the six parameter SIMBA 1P set that vary feedback parameters (no cosmology variations), at $z = 0$. Colour scale corresponds to the median of gas particle spread divided by the initial neighbouring dark matter particle spread for gas selected to have a dark matter neighbour in one of the five most massive haloes in the final snapshot. There exists a clear correlation between suppression of power on all scales and median normalized gas spread.

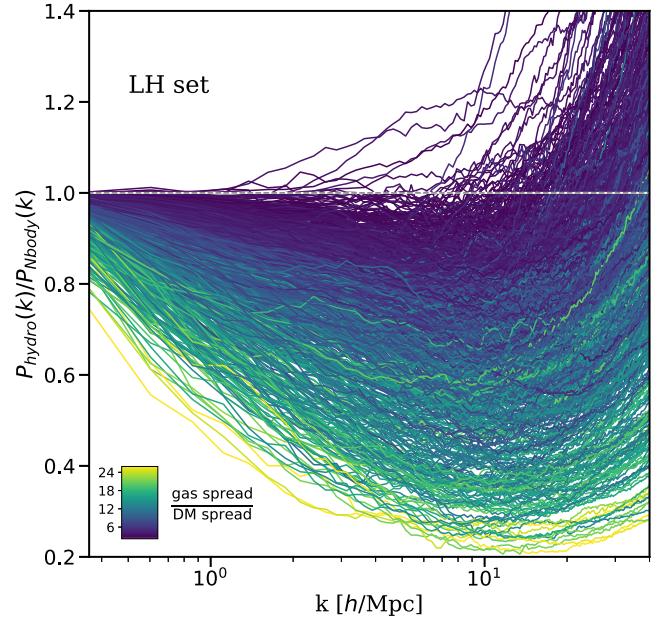


Figure 10. Same as Fig. 8 but for the 1000 simulation pairs in the SIMBA LH set, implementing different cosmological and feedback parameters along with varying initial conditions. Despite the large scatter, gas spread is a good predictor of the suppression of matter clustering due to baryonic physics.

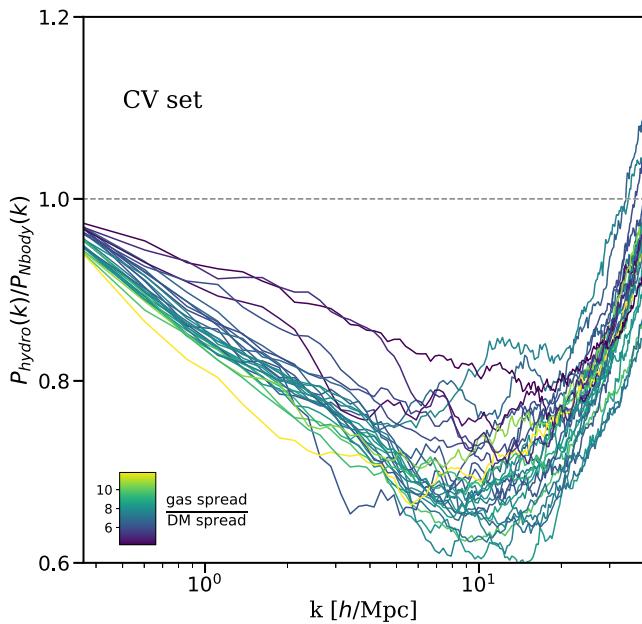


Figure 9. Same as Fig. 8 but for the 27 simulation pairs of the SIMBA CV set, using fiducial simulation parameters and varied initial conditions. The strong correlation between suppression of power and baryonic spread holds on large scales ($k \lesssim 2 h \text{ Mpc}^{-1}$) in spite of cosmic variance but is lost on smaller scales.

simulated volumes in CAMELS, cosmic variance alone represents a significant amount of scatter in the amount of power suppression even for fiducial feedback parameters (Villaescusa-Navarro et al. 2021c; Delgado et al. 2023; Ni et al. 2023). Interestingly, the median gas spread remains a good predictor of the impact of feedback on the total matter power spectrum, capturing most of the variation of $P_{\text{hydro}}/P_{\text{Nbody}}$ due to cosmic variance on scales $k \lesssim 2 h \text{ Mpc}^{-1}$. Notably, this result holds even without normalizing the spread. As expected, the correlation between median gas spread and $P_{\text{hydro}}/P_{\text{Nbody}}$ worsens on smaller scales as the inherent stochasticity of galaxy formation becomes a more dominant effect (Genel et al. 2019; Keller et al. 2019; Borrow et al. 2023). Fig. 10 repeats this exercise for the full SIMBA LH set, where each of the 1000 simulations has different cosmological and astrophysical parameters in addition to different initial conditions. In this case, there is significantly more scatter, but the relationship between total matter power spectrum suppression and the amount of baryon spreading continues to hold over a range of scales.

To examine this trend in further depth, we plot in Fig. 11 the fractional difference in the power spectra against the normalized spread for different values of k , where we show $\Delta P/P \equiv (P_{\text{hydro}} - P_{\text{Nbody}})/P_{\text{Nbody}}$ for both the six parameter SIMBA 1P set (only feedback variations, squares) and CV set (circles) at four different values of k . The relationship between $\Delta P/P$ and baryonic spread is nearly linear at low values of k , while at higher k values it takes on a more complicated form. Additionally, the data points from the CV set tend to have larger scatter, which increases with k as expected from Fig. 9. We use PySR symbolic regression (Crammer 2020) with the six parameter SIMBA 1P set simulations that vary feedback parameters and find the following functional form to model the fractional difference in the total matter power spectrum due to baryonic effects as a function of both k and the amount of spread,

$$-\Delta P/P = a_1 \times (k \times (a_2 \times S - a_3))^{1/4} - a_4, \quad (2)$$

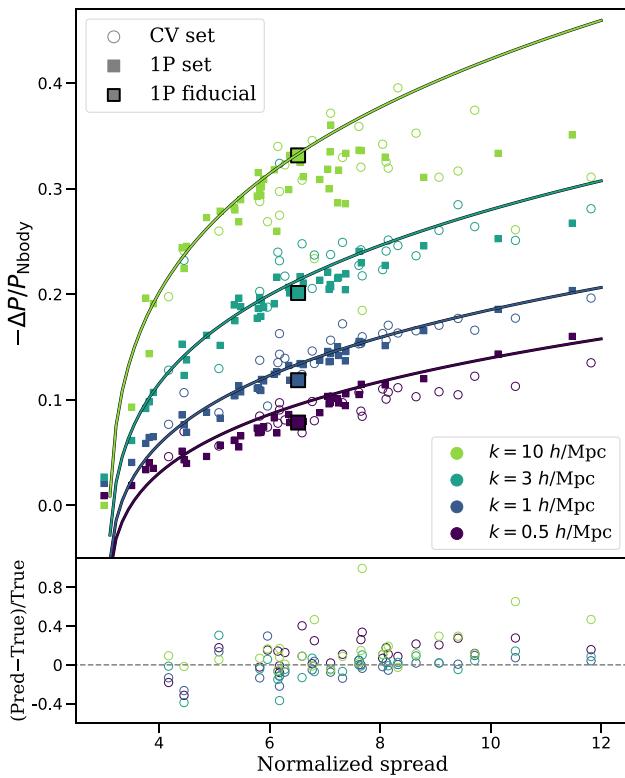


Figure 11. Fractional difference in power spectrum at $z = 0$ as a function of the normalized spread (as described in Fig. 8) at different values of k for both the six parameter 1P set (only feedback variations, squares) and CV set (circles) in SIMBA (top panel). The simulation with fiducial parameters from the 1P set is plotted with a larger, outlined square for each k value. Best-fitting lines found via symbolic regression for the 1P simulations are plotted at each of these k values. While only trained on the six parameter SIMBA 1P set simulations that varied feedback parameters, symbolic regression roughly captures the trends seen in the CV set as well. Errors for CV set predictions are shown in the bottom panel.

where S is the median normalized gas spread metric and $a_1 = 0.25$, $a_2 = 0.35$, $a_3 = 1.09$, and $a_4 = 0.14$ are the best-fitting coefficients for the preferred functional form found by PySR. The errors from CV set predictions are shown in the bottom panel. Encouragingly, this simple expression roughly captures the dependence of $\Delta P/P$ on baryonic spread as a function of k despite cosmic variance effects.

3.4 Comparison to other CAMELS suites

Here, we compare gas spreading between SIMBA, IllustrisTNG, and ASTRID. Fig. 12 shows the cumulative mass distribution of gas as a function of spread distance for each fiducial model (solid lines) and for the full range of parameter variations as given by either the LH set (for SIMBA and ASTRID) or the SB28 set (for IllustrisTNG). Clearly, the fiducial SIMBA simulation spreads gas the furthest (~ 40 per cent of gas in the simulated volume spreads further than $1 \text{ Mpc } h^{-1}$ from its neighbouring dark matter), and the variations in the LH set predict a wide range of gas spread (anywhere from 1–75 per cent of gas spreading further than $1 \text{ Mpc } h^{-1}$). The fiducial ASTRID simulation appears to generally spread gas the least (~ 7 per cent spreading further than $1 \text{ Mpc } h^{-1}$), but also predicts a wide variation in the LH set (0.2–63 per cent for $> 1 \text{ Mpc } h^{-1}$). The gas spread in the fiducial IllustrisTNG simulation is on a similar level to ASTRID (~ 11 per cent spreading further than $1 \text{ Mpc } h^{-1}$). The range given by the SB28 set shows the smallest minimum and maximum spread (0.01–57 per cent), but spans the largest logarithmic range. The increased gas spread seen in SIMBA relative to IllustrisTNG is in qualitative agreement with results found in Ayromlou, Nelson & Pillepich (2022), where the ‘closure radius’ (defined as the characteristic radius from a halo within which the enclosed baryon fraction equals the cosmic baryon fraction) is significantly larger in SIMBA relative to IllustrisTNG and EAGLE, indicating a greater redistribution of baryons.

We extend our investigation of the relationship between baryon spread and large-scale suppression of the matter power spectrum now as a function of galaxy formation model. Fig. 13 is an extension of Fig. 8 that now includes simulations from the ASTRID 1P set and the full 28 parameter 1P sets for IllustrisTNG and SIMBA that do not vary cosmological parameters (180 simulations in total). Remarkably, we find a very clear correlation between the suppression of power on

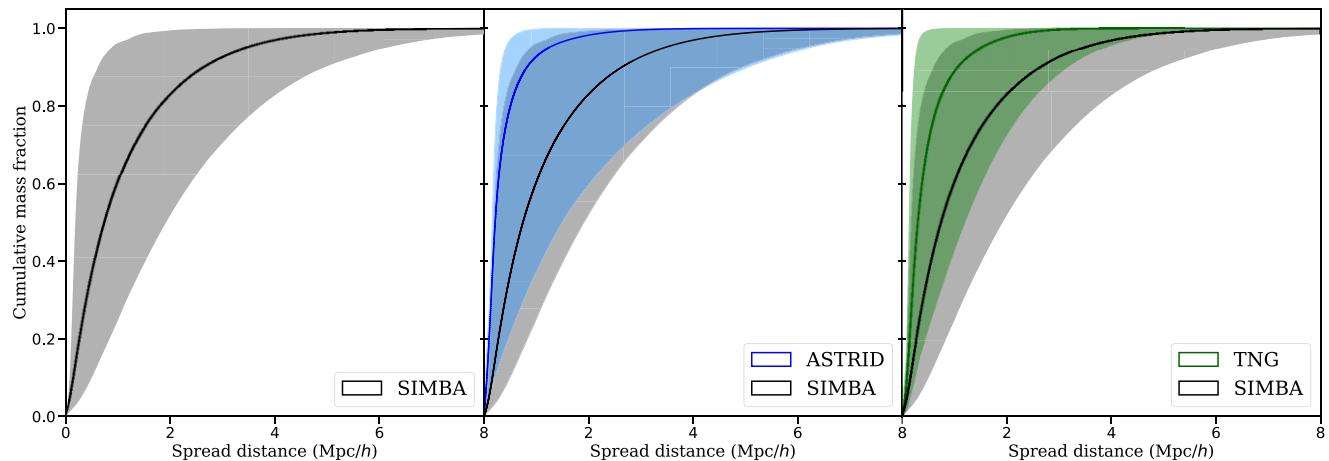


Figure 12. Cumulative mass distribution as a function of spread distance at $z = 0$ for gas particles in the fiducial 1P simulation (solid lines) and the full LH set (SB28 for IllustrisTNG; shaded region) for SIMBA (grey; all panels), ASTRID (blue; middle panel), and IllustrisTNG (green; right panel). The fiducial SIMBA model spreads gas significantly further than ASTRID and IllustrisTNG, although the full range of gas spreading is comparable in the LH sets of SIMBA and ASTRID.

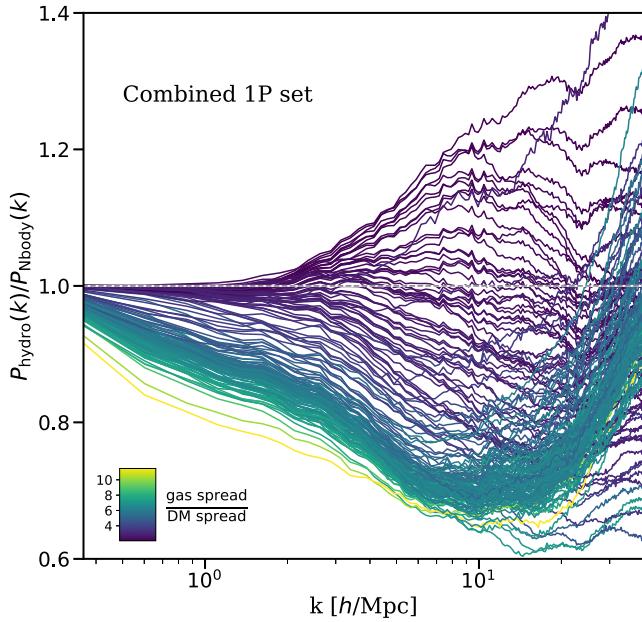


Figure 13. Same as Fig. 8, but now including simulations from the ASTRID 1P set, and 28 parameter 1P sets for IllustrisTNG and SIMBA that do not vary cosmological parameters. The correlation between suppression of power and gas spread is quite clear even across different galaxy formation models and variations of up to 23 astrophysical parameters.

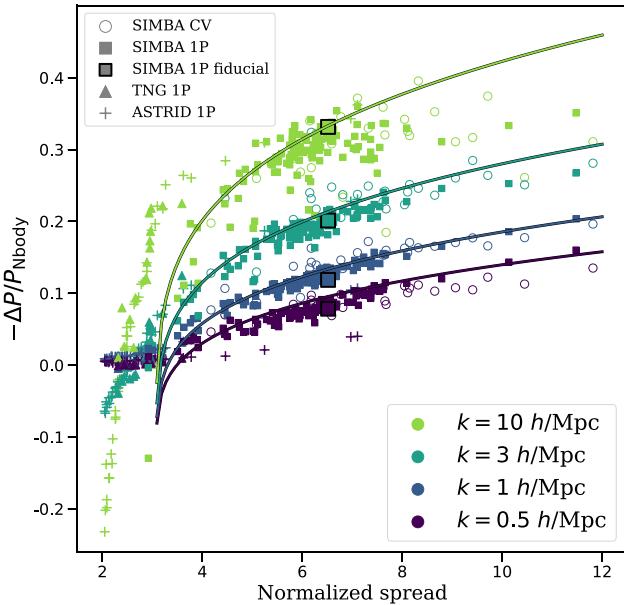


Figure 14. Same as Fig. 11, but now including simulations from the ASTRID 1P set, and 28 parameter 1P sets for IllustrisTNG and SIMBA that do not vary cosmological parameters. The same analytic function from symbolic regression is plotted (trained on six parameter SIMBA 1P simulations), which appears to be successful in reproducing the full SIMBA 1P results but does not capture the trends seen for IllustrisTNG and ASTRID at low spread values.

scales $k \lesssim 10 h \text{ Mpc}^{-1}$ and the gas spread metric regardless of galaxy formation model and variations of up to 23 astrophysical parameters. Once again, we investigate this trend quantitatively in Fig. 14, where we show $\Delta P/P$ as a function of gas spread for different values of k , as in Fig. 11 but now including also the ASTRID 1P set and

the full 28 parameter 1P sets for IllustrisTNG and SIMBA. Here we reproduce the analytic function found with symbolic regression using only the six parameter SIMBA 1P simulations, which also seems to succeed in fitting the broader parameter variations in the full SIMBA 1P set. However, it is clear that this function does not quite match the ASTRID and IllustrisTNG simulations at all k , nor at small spreads. At small k , ASTRID suppresses power notably less than SIMBA simulations with comparable gas spread, and at large k , both ASTRID and IllustrisTNG show far more suppression than SIMBA when gas is not spreading far. Many ASTRID simulations spread gas by small amounts and in fact show an increase in power relative to N -body as the gas spread declines. There indeed appears to be a relationship between baryonic impact on the matter power spectrum and the spreading of baryons, but this trend can significantly depend on the galaxy formation model.

4 DISCUSSION

The decoupling of baryons from dark matter on cosmological scales represents a key signature of astrophysical feedback processes. As gas does not simply follow the gravitational pull of dark matter, N -body simulations do not tell the whole story. Systematic comparisons between N -body and hydrodynamic simulations allow for a controlled analysis of the role that feedback plays in the distribution of matter. It was shown in Borrow et al. (2020) that baryonic matter can spread great distances away from the initial neighbouring dark matter distribution in the SIMBA simulation. Here, we have used CAMELS to extend this analysis to a wide range of variations in cosmological and subgrid parameters in different plausible galaxy formation models, with the goal of encompassing the actual baryonic spread in the real Universe. Our results highlight the extent to which baryonic matter can cycle in and out of galaxies, be ejected from the circumgalactic medium, or even transferred to other haloes (Davé, Finlator & Oppenheimer 2012; Christensen et al. 2016; Anglés-Alcázar et al. 2017b; Hafen et al. 2019, 2020; Mitchell, Schaye & Bower 2020; Wright et al. 2020; Ayromlou et al. 2022).

In the fiducial SIMBA model, 40 per cent of the gas mass in the entire simulated cosmological volume spreads further than 1 Mpc, which is beyond the virial radius of all haloes in the simulation. Increasing AGN jet speed (the $A_{\text{AGN}2}$ parameter in CAMELS) from lowest to highest increases this percentage from 25 per cent up to 55 per cent. This large-scale spreading of baryons represents an important consideration for halo models and baryonification methods (e.g. Seljak 2000; Semboloni et al. 2013; Mead et al. 2015; Schneider & Teyssier 2015; Schneider et al. 2019; Weiss et al. 2019). Models that consider only a redistribution of matter relative to N -body simulations within the scale of individual halos can easily miss the large-scale effects of baryons. Even with the weakest AGN feedback in SIMBA, more than 25 per cent of gas from halo Lagrangian regions (which should otherwise accrete onto haloes) end up spreading far out of haloes. While some models do take this ejection of gas into account (e.g. Schneider & Teyssier 2015; Mead et al. 2020), we have shown here that different subgrid parameters and models can spread varying amounts of gas to a wide range of distances, highlighting the challenge of modeling the ejected component. Additionally, we explore the spreading of gas outside of Lagrangian regions (which never belonged to any haloes) in Appendix B. Fig. B2 shows the cumulative mass distribution of this gas as a function of spread distance for varied AGN jet speeds in SIMBA (similar to Fig. 5), with transparent lines showing the spread of all gas in each simulation. Gas outside of Lagrangian regions (which accounts for ~ 70 per cent of the total gas mass in the fiducial SIMBA simulation) spreads

significantly (~ 40 percent of gas outside of Lagrangian regions spreads farther than 1 Mpc) but is not accounted for in halo-based models.

We have shown that dark matter itself spreads (relative to its initial neighbouring dark matter distribution) by the largest amount within and around massive haloes, as they are the largest sources of gravitational potential. Dark matter spreading increases when haloes get larger, as chaotic dynamics can make initial particle neighbour trajectories diverge on scales comparable to the splashback radius (Adhikari, Dalal & Chamberlain 2014; Diemer & Kravtsov 2014; More, Diemer & Kravtsov 2015; Mansfield, Kravtsov & Diemer 2017). We see larger haloes forming when increasing the values of Ω_m and σ_8 (as expected, see Press & Schechter 1974; Villaescusa-Navarro et al. 2021c), but otherwise large-scale dark matter spreading seems to be roughly independent of these cosmological parameter variations. One exception is the slight decrease in dark matter spread in haloes of equal mass as Ω_m (and σ_8 for lower-mass haloes) increases (Fig. 3), which may be explained by variations in halo concentration. Higher values of Ω_m and σ_8 increase halo concentration at fixed virial mass (Dooley et al. 2014), which would result in haloes with more mass concentrated in the central region and thus reduced amount of dark matter spreading within them. Gas largely follows the same trend as Ω_m and σ_8 increase but experience additional interactions. Generally, larger haloes may increase spread by generating stronger feedback (due to the presence of more massive central black holes) and having chaotic particle trajectories over larger scales, but also host baryons that collapse to higher densities in the central galaxy and end up spreading very little.

The dependence of gas spread on feedback parameters shows complex non-linear effects. A clear result in Fig. 4 is that increasing the strength of AGN feedback in SIMBA (momentum flux and jet speed) significantly increases gas spread, while increasing the stellar feedback efficiency either yields mixed results (mass loading factor) or a significant decrease in gas spread (wind speed). One possible explanation for this non-intuitive result is that there is significant non-linear interaction between stellar and AGN feedback, as seen in previous works (Booth & Schaye 2013; van Daalen et al. 2020; Nicola et al. 2022; Delgado et al. 2023). Analyses of high resolution FIRE zoom-in simulations show that strong stellar feedback can limit early black hole growth by continually ejecting material away from the nuclear region (Anglés-Alcázar et al. 2017c; Ćatmabacak et al. 2022; Byrne et al. 2023; see also Dubois et al. 2015; Bower et al. 2017; Habouzit, Volonteri & Dubois 2017; Lapiner, Dekel & Dubois 2021), which can therefore reduce the impact of AGN feedback. Indeed, Ni et al. (2023) showed that increasing the SNe wind speed parameter (A_{SN2}) in SIMBA greatly decreased the quantity of massive black holes. Borrow et al. (2020) found that gas particles tagged as having directly interacted with AGN jets in SIMBA were spread significantly further than particles that only directly interacted with stellar feedback. Furthermore, by contrasting with a ‘No-Jet’ SIMBA simulation, it was shown that gas does not spread to large distances without AGN jets and instead spreads on the same level as dark matter. Given these effects, our results suggest that increasing the speed of galactic winds suppresses the overall output of AGN jets and therefore their ability to redistribute gas over large scales.

Massive dark matter haloes are responsible for both very small and extremely large baryonic spreads owing to the competing effects of radiative cooling (allowing gas to collapse down to halo centres) and strong AGN feedback (ejecting gas to large scales). This dichotomy in the fate of gas can be clearly seen for gas particles that belong to halo Lagrangian regions at the initial conditions (Fig. 6). Gas particles that remain inside of their parent halos at $z = 0$ spread very little,

with minimal dependence on feedback parameters, while Lagrangian region gas outside of parent halos at $z = 0$ show significantly larger and feedback-dependent spreads. Investigating the large-scale spatial distribution of gas as a function of the amount of spread provides further support for this picture (Fig. 7), where the least spread gas is constrained to halo centres while the most spread gas is spread out around large haloes and filaments (in agreement with Borrow et al. 2020). As the jet speed increases, the least spread gas is even more tightly constrained to halo centres, and the most spread gas is even more diffusely spread out over a large fraction of the simulated volume in CAMELS. This depiction is in agreement with our finding in Fig. 6 that the in-halo gas spread variation (25th–75th percentile shaded region) decreases while the out-of-halo gas spread increases with higher AGN jet speed.

We have shown that the amount of baryonic spread in simulations is closely related to the overall impact of feedback on the total matter power spectrum. Previous authors have investigated the impact of baryons on the matter power spectrum and found that it can be significantly ‘contaminated’ by non-linear baryonic effects relative to dark matter-only simulations (Daalen et al. 2011; Chisari et al. 2019; van Daalen et al. 2020; Delgado et al. 2023; Pandey et al. 2023). This contamination typically results in a reduction of power on large scales in hydrodynamic simulations as stellar feedback and, in particular, AGN feedback redistribute gas far from haloes they would otherwise reside in or around. With a large extended set of ~ 200 simulations with identical initial conditions and varying up to 23 astrophysical parameters in the SIMBA, IllustrisTNG, and ASTRID models (Ni et al. 2023), we have shown that there is a tight correlation between the suppression of power on scales $k \lesssim 10 h \text{ Mpc}^{-1}$ and the large-scale baryon spread (Fig. 13). Generally, simulations that spread baryons further relative to their initial neighbouring dark matter distribution show a greater suppression of power on large scales, regardless of the specific galaxy formation model and feedback parameter variations.

Due to the small simulated volumes in CAMELS, cosmic variance can have significant effects on many measured quantities. While all of the SIMBA CV set simulations implement identical feedback parameters, different initial conditions may result in a different population of haloes for which the same feedback model can have widely different effects. In particular, our results highlight the extent to which cosmic variance in a $(25 \text{ Mpc} h^{-1})^3$ volume can play a role in the large-scale spreading of baryons, with some SIMBA CV simulations showing a median gas spread twice that of other simulations with identical parameters. Previous works in CAMELS have partially mitigated the limitation of small simulated volumes by finding good predictors of cosmic variance. Nicola et al. (2022) reduced the effect of cosmic variance on neural networks trained to constrain cosmological and astrophysical parameters from electron density power spectra by incorporating a parameter encoding the distribution of halo masses for each input simulation. Thiele et al. (2022) quantified the constraining power of spectral distortion measurements for baryonic feedback models and reduced sample variance in CAMELS by deriving a correction factor based on scaling the measured halo mass functions in CAMELS to that of reference large volume simulations. Delgado et al. (2023) trained a random forest regressor to predict the impact of baryonic effects on the matter power spectrum given halo baryon fractions and increased its predictive power significantly by including a form of the halo mass function as input feature, partially mitigating cosmic variance effects. Interestingly, the normalized gas spread metric is an excellent predictor of the effects of cosmic variance on the inferred impact of feedback on the matter power spectrum, where the amount of power

suppression in SIMBA CV simulations with identical feedback parameters is tightly correlated with baryon spreading up to scales $k \lesssim 2 h \text{ Mpc}^{-1}$ (Fig. 9).

The way feedback processes are implemented, and thus the impact they have on the distribution of matter, can vary greatly between hydrodynamic simulation models. Chisari et al. (2019) compared the impact of baryons on the matter power spectrum at $z = 0$ in the fiducial models of a handful of cosmological hydrodynamic simulations, including Horizon-AGN (Dubois et al. 2014), MassiveBlack-II (Khandai et al. 2015), OWLS (Schaye et al. 2010), cosmo-OWLS (Le Brun et al. 2014), EAGLE (Schaye et al. 2015), BAHAMAS (McCarthy et al. 2017, 2018), Illustris (Genel et al. 2014; Vogelsberger et al. 2014a,b), and IllustrisTNG (Springel et al. 2018). They found that the suppression of power relative to N -body simulations can range from 10–30 per cent in the different models at wave numbers from a few up to $20 h \text{ Mpc}^{-1}$. More recently, van Daalen et al. (2020) found similar results when quantifying the impact of baryonic physics on matter clustering for nearly 100 simulations from the OWLS, cosmo-OWLS, and BAHAMAS models that varied cosmological and feedback parameters. Our results further emphasize that the distribution of matter in cosmological hydrodynamic simulations strongly depends on both the simulation model and the strength of feedback. Fig. 12 highlights this fact, with the spreading of baryons varying widely throughout thousands of CAMELS simulations with different parameters, initial conditions, and galaxy formation models. This large library of simulations enables the development of machine learning algorithms that can quantify baryonic uncertainties and marginalize over them for cosmological parameter inference (Villaescusa-Navarro et al. 2021c; Perez et al. 2022; Ni et al. 2023), as well as devise observational probes that can help constrain baryonic physics. For example, Nicola et al. (2022) used CAMELS to investigate the electron density power spectrum (measurable through kinematic Sunyaev-Zel'dovich observations or Fast Radio Burst dispersion measures) as a means to break the baryon-cosmology degeneracy (see also Jo et al. 2023) and improve theoretical models of the impact of baryonic feedback on the matter power spectrum, and Pandey et al. (2023) used Dark Energy Survey weak lensing and *Atacama Cosmology Telescope* thermal Sunyaev-Zel'dovich effect measurements together with models trained on CAMELS to constrain the impact of feedback on matter clustering. Constraints such as these will be extremely valuable for extracting the maximum information out of upcoming cosmological surveys to further constrain the fundamental parameters of the Universe.

Previous authors have also noted a connection between lower halo baryon fraction f_b and increased baryonic impact on the matter power spectrum (van Daalen et al. 2020; Nicola et al. 2022; Delgado et al. 2023; Pandey et al. 2023). Our results connecting baryonic spread and suppression of power across a large library of models are in agreement with these findings. Simulations forming halo populations with lower f_b require more baryons to be ejected out of haloes, which results in greater baryonic spread. The baryon fraction of massive haloes ($M_{\text{halo}} \sim 10^{14} M_{\odot}$) is particularly well correlated with the suppression of the matter power spectrum, and van Daalen et al. (2020) formulated an empirical model to predict power suppression as a function of f_b which is accurate up to $k \lesssim 1 h \text{ Mpc}^{-1}$ across a number of galaxy formation models. However, Delgado et al. (2023) showed that this model cannot capture the f_b –power suppression relation in SIMBA, expanding this work to thousands of CAMELS simulations across the full halo mass range ($10^{10} M_{\odot} \leq M_{\text{halo}} \leq 10^{14} M_{\odot}$) to train a random forest regressor capable of predicting the baryonic impact on the matter power spectrum well into the non-linear regime. We have shown that a simple analytic function

found by symbolic regression can roughly capture the suppression of the matter power spectrum as a function of baryon spread also into the non-linear regime for SIMBA simulations varying 23 astrophysical parameters despite training on simulations varying only four feedback parameters (Fig. 11). However, the same model cannot accurately capture this relationship for IllustrisTNG and ASTRID, particularly in simulations with lower baryon spread compared to that represented in the SIMBA simulations (Fig. 14). Furthermore, ASTRID simulations with comparable baryon spread to those in SIMBA show less suppression at larger scales ($k > 1 h \text{ Mpc}^{-1}$), while at smaller scales ($k \sim 10 h \text{ Mpc}^{-1}$) small deviations in the normalized spread show large variations in power suppression. Such discrepancies indicate that while in general simulations that spread baryons further show a greater suppression of power, the exact relationship between baryon spread and power suppression does indeed depend on the galaxy formation model. These results highlight that differences in the predicted cosmic baryon cycle in simulations provide unique discriminating power between galaxy formation models (Oppenheimer et al. 2021; Crain & van de Voort 2023; Wright et al. 2024).

This also emphasizes the need to construct robust models relative to changes in the galaxy formation implementation, which is a common difficulty of many current models (for further discussion and examples of robust models, see Villaescusa-Navarro et al. 2021a, b, 2022a; de Santi et al. 2023; Echeverri et al. 2023; Shao et al. 2023a, b).

5 CONCLUSIONS

We have explored in detail the use of a recently proposed *cosmological spread metric* to describe the redistribution of dark and baryonic matter on large scales owing to gravitational dynamics and feedback from astrophysical sources, providing a framework to quantify and interpret the impact of baryonic effects on the total matter power spectrum with a single summary statistic. Our main results can be summarized as follows:

(i) Dark matter in N -body simulations spreads relative to the initial neighbouring matter distribution owing to chaotic gravitational dynamics, with the largest spread distances occurring in and around massive haloes. As expected, dark matter spreading increases with Ω_m and σ_8 following the formation of higher mass haloes in simulations.

(ii) On average, gas spreads much further than dark matter due to astrophysical feedback effects. Radiative cooling can allow gas to lose energy and fall to lower bound orbits at the centres of haloes, but gas impacted by feedback can be ejected to large distances. This dichotomy of gas cooling and feedback yields a large variation in spread distances for gas inside of halo Lagrangian regions at the initial conditions.

(iii) Increasing the efficiency of AGN feedback increases the spread of gas, but increasing the stellar feedback efficiency can decrease the spread of gas. This supports the notion that AGN feedback is the dominant component spreading baryons to large distances while stronger stellar feedback may inhibit black hole growth and therefore reduce the impact of AGN feedback.

(iv) The baryonic spread metric is a good predictor of the global impact of feedback on the large scale distribution of matter as described by the ratio of the matter power spectrum in hydrodynamic and N -body simulations, with larger baryonic spread driving stronger suppression of power on large scales.

(v) Using symbolic regression, we have found a simple analytic function that captures the matter power suppression as a function

of wave number and baryonic spread in simulations varying >20 astrophysical parameters in the SIMBA model, while extrapolating to IllustrisTNG and ASTRID remains a challenge.

The extent to which matter is redistributed by feedback processes is significant and can contribute to uncertainties in approximation methods that do not model baryonic physics explicitly, while predictions from cosmological hydrodynamic simulations can vary widely depending on the choice of feedback parameters and model implementation. Besides providing a clear physical interpretation for the impact of baryonic physics on the matter power spectrum, the simplicity of the spread metric makes it a useful summary statistic to characterize the global efficiency of feedback in galaxy formation simulations.

ACKNOWLEDGEMENTS

We thank Dylan Nelson and Joop Schaye for useful suggestions and the referee for a constructive report that helped improve the paper. The CAMELS simulations were performed on the supercomputing facilities of the Flatiron Institute, which is supported by the Simons Foundation. DAA acknowledges support by NSF grants AST-2009687 and AST-2108944, CXO grant TM2-23006X, Simons Foundation Award CCA-1018464, and Cottrell Scholar Award CS-CSA-2023-028 by the Research Corporation for Science Advancement. DN acknowledges support from the NSF AST-2206055 grant. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

DATA AVAILABILITY

The CAMELS simulations are publicly available (see Villaescusa-Navarro et al. 2023; Ni et al. 2023). More information and instructions for downloading the data can be found at <https://camels.readthedocs.io>.

REFERENCES

Abazajian K. N. et al., 2016, preprint ([arXiv:1610.02743](https://arxiv.org/abs/1610.02743))

Adhikari S., Dalal N., Chamberlain R. T., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 19

Anglés-Alcázar D., Özel F., Davé R., 2013, *ApJ*, 770, 5

Anglés-Alcázar D., Özel F., Davé R., Katz N., Kollmeier J. A., Oppenheimer B. D., 2015, *ApJ*, 800, 127

Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017a, *MNRAS*, 464, 2840

Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017b, *MNRAS*, 470, 4698

Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017c, *MNRAS*, 472, L109

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, *MNRAS*, 426, 2046

Ayromlou M., Nelson D., Pillepich A., 2022, *MNRAS* 524 5391

Baugh C. M., 2006, *Rep. Prog. Phys.*, 69, 3101

Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379

Beltz-Mohrmann G. D., Berlind A. A., 2021, *ApJ*, 921, 112

Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587

Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, *MNRAS*, 512, 3703

Bondi H., 1952, *MNRAS*, 112, 195

Booth C. M., Schaye J., 2013, *Sci. Rep.*, 3, 1738

Borrow J., Anglés-Alcázar D., Davé R., 2020, *MNRAS*, 491, 6102

Borrow J., Schaller M., Bahé Y. M., Schaye J., Ludlow A. D., Ploeckinger S., Nobels F. S. J., Altamura E., 2023, *MNRAS*, 526, 2441

Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, *MNRAS*, 465, 32

Byrne L., Faucher-Giguère C.-A., Stern J., Anglés-Alcázar D., Wellons S., Gurvich A. B., Hopkins P. F., 2023, *MNRAS*, 520, 722

Çatmabacak O., Feldmann R., Anglés-Alcázar D., Faucher-Giguère C.-A., Hopkins P. F., Kereš D., 2022, *MNRAS*, 511, 506

Chisari N. E. et al., 2018, *MNRAS*, 480, 3962

Chisari N. E. et al., 2019, *Open J. Astrophys.*, 2 4

Choi E., Ostriker J. P., Naab T., Johansson P. H., 2012, *ApJ*, 754, 125

Christensen C. R., Davé R., Governato F., Pontzen A., Brooks A., Munshi F., Quinn T., Wadsley J., 2016, *ApJ*, 824, 57

Christiansen J. F., Davé R., Sorini D., Anglés-Alcázar D., 2020, *MNRAS*, 499, 2617

Cicone C. et al., 2014, *A&A*, 562, A21

Cole S. et al., 2005, *MNRAS*, 362, 505

Crain R. A., van de Voort F., 2023, *ARA&A*, 61, 473

Cranmer M., PySR: Fast and Parallelized Symbolic Regression in Python/Julia. 2020, available at: <http://doi.org/10.5281/zenodo.4041459>

Croton D. J. et al., 2006, *MNRAS*, 365, 11

Cui W., Borgani S., Murante G., 2014, *MNRAS*, 441, 1769

Cui W. et al., 2016, *MNRAS*, 458, 4052

DESI Collaboration, 2016, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))

Davé R., Finlator K., Oppenheimer B. D., 2012, *MNRAS*, 421, 98

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 Santi N. S. M. et al., 2023, *ApJ*, 952, 69

Delgado A. M. et al., 2023, *MNRAS*, 526, 5306

Diemer B., Kravtsov A. V., 2014, *ApJ*, 789, 1

Dooley G. A., Griffen B. F., Zukin P., Ji A. P., Vogelsberger M., Hernquist L. E., Frebel A., 2014, *ApJ*, 786, 50

Dubois Y. et al., 2014, *MNRAS*, 444, 1453

Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., Teyssier R., 2015, *MNRAS*, 452, 1502

Echeverri N. et al., 2023, *ApJ* 954 125

Eisenstein D. J. et al., 2005, *ApJ*, 633, 560

Euclid Collaboration, 2022, *A&A*, 662, A112

Fabian A. C., 2012, *ARA&A*, 50, 455

Faucher-Giguère C.-A., 2020, *MNRAS*, 493, 1614

Feruglio C., Maiolino R., Piconcelli E., Menci N., Aussel H., Lamastra A., Fiore F., 2010, *A&A*, 518, L155

Genel S., Vogelsberger M., Nelson D., Sijacki D., Springel V., Hernquist L., 2013, *MNRAS*, 435, 1426

Genel S. et al., 2014, *MNRAS*, 445, 175

Genel S. et al., 2019, *ApJ*, 871, 21

Greene J. E., Zakamska N. L., Smith P. S., 2012, *ApJ*, 746, 86

Guo Q. et al., 2011, *MNRAS*, 413, 101

Habouzit M., Volonteri M., Dubois Y., 2017, *MNRAS*, 468, 3935

Hadzhiyska B., García-García C., Alonso D., Nicola A., Slosar A., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 20

Hafen Z. et al., 2019, *MNRAS*, 488, 1248

Hafen Z. et al., 2020, *MNRAS*, 494, 3581

Hikage C. et al., 2019, *PASJ*, 71, 43

Hildebrandt H. et al., 2017, *MNRAS*, 465, 1454

Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, *MNRAS*, 442, 2304

Hopkins P. F., 2015, *MNRAS*, 450, 53

Hopkins P. F., Quataert E., 2011, *MNRAS*, 415, 1027

Huang H.-J. et al., 2021, *MNRAS*, 502, 6010

Ivezic Ž. et al., 2019, *ApJ*, 873, 111

Jo Y. et al., 2023, *ApJ*, 944, 67

Katz N., Weinberg D. H., Hernquist L., 1996, *ApJS*, 105, 19

Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201

Keller B. W., Wadsley J. W., Wang L., Kruijssen J. M. D., 2019, *MNRAS*, 482, 2244

Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, *MNRAS*, 450, 1349

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, *ApJ*, 740, 102

Lapiner S., Dekel A., Dubois Y., 2021, *MNRAS*, 505, 172

Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, *MNRAS*, 441, 1270

Lu T., Haiman Z., 2021, *MNRAS*, 506, 3406

Lynds C. R., Sandage A. R., 1963, *ApJ*, 137, 1005

McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, *MNRAS*, 465, 2936

McCarthy I. G., Bird S., Schaye J., Harnois-Deraps J., Font A. S., van Waerbeke L., 2018, *MNRAS*, 476, 2999

Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, *MNRAS*, 283, 1388

Mansfield P., Kravtsov A. V., Diemer B., 2017, *ApJ*, 841, 34

Martin C. L., 1998, *ApJ*, 506, 222

Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *MNRAS*, 454, 1958

Mead A. J., Tröster T., Heymans C., Van Waerbeke L., McCarthy I. G., 2020, *A&A*, 641, A130

Merloni A. et al., 2012, preprint ([arXiv:1209.3114](https://arxiv.org/abs/1209.3114))

Mitchell P. D., Schaye J., Bower R. G., 2020, *MNRAS*, 497, 4495

More S., Diemer B., Kravtsov A. V., 2015, *ApJ*, 810, 36

Moser E. et al., 2022, *ApJ*, 933, 133

Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, *MNRAS*, 454, 2691

Naab T., Ostriker J. P., 2017, *ARA&A*, 55, 59

Ni Y. et al., 2022, *MNRAS*, 513, 670

Ni Y. et al., 2023, *ApJ* 959 136

Nicola A. et al., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 46

Oppenheimer B. D., Babul A., Bahé Y., Butsky I. S., McCarthy I. G., 2021, *Universe*, 7, 209

Osato K., Nagai D., 2023, *MNRAS*, 519, 2069

Osato K., Flender S., Nagai D., Shirasaki M., Yoshida N., 2018, *MNRAS*, 475, 532

Osato K., Shirasaki M., Miyatake H., Nagai D., Yoshida N., Oguri M., Takahashi R., 2020, *MNRAS*, 492, 4780

Pandey S. et al., 2023, *MNRAS* 525 1779–1794

Perez L. A., Genel S., Villaescusa-Navarro F., Somerville R. S., Gabrielpillai A., Anglés-Alcázar D., Wandelt B. D., Yung L. Y. A., 2023, *ApJ* 954 11

Pettini M., Shapley A. E., Steidel C. C., Cuby J.-G., Dickinson M., Moorwood A. F. M., Adelberger K. L., Giavalisco M., 2001, *ApJ*, 554, 981

Pillepich A. et al., 2018, *MNRAS*, 473, 4077

Planck Collaboration, 2020, *A&A*, 641, A6

Press W. H., Schechter P., 1974, *ApJ*, 187, 425

Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, *MNRAS*, 430, 2427

Reichardt C. L. et al., 2012, *ApJ*, 755, 70

Schaye J. et al., 2010, *MNRAS*, 402, 1536

Schaye J. et al., 2015, *MNRAS*, 446, 521

Schaye J. et al., 2023, *MNRAS*, 526, 4978

Schneider A., Teyssier R., 2015, *J. Cosmol. Astropart. Phys.*, 2015, 49

Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 20

Seljak U., 2000, *MNRAS*, 318, 203

Semboloni E., Hoekstra H., Schaye J., 2013, *MNRAS*, 434, 148

Shao H. et al., 2022, *ApJ*, 927, 85

Shao H. et al., 2023a, *ApJ*, 956, 149

Shao H. et al., 2023b, *ApJ*, 944, 27

Shaw L. D., Nagai D., Bhattacharya S., Lau E. T., 2010, *ApJ*, 725, 1452

Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, *MNRAS*, 380, 877

Smith B. D. et al., 2016, *MNRAS*, 466, 2217

Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51

Somerville R. S., Primack J. R., 1999, *MNRAS*, 310, 1087

Sorini D., Davé R., Cui W., Appleby S., 2022, *MNRAS*, 516, 883

Springel V., 2005, *MNRAS*, 364, 1105

Springel V., 2010, *MNRAS*, 401, 791

Springel V., Hernquist L., 2003, *MNRAS*, 339, 289

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726

Springel V. et al., 2005, *Nature*, 435, 629

Springel V. et al., 2018, *MNRAS*, 475, 676

Sturm E. et al., 2011, *ApJ*, 733, L16

Thiele L. et al., 2022, *Phys. Rev. D*, 105, 83505

Tillman M. T. et al., 2023, *ApJ*, 166, 228

van Daalen M. P., Schaye J., Booth C. M., Vecchia C. D., 2011, *MNRAS*, 415, 3649

van Daalen M. P., McCarthy I. G., Schaye J., 2020, *MNRAS*, 491, 2424

van den Bosch F. C. et al., 2007, *MNRAS*, 376, 841

Villanueva-Domingo P., Villaescusa-Navarro F., 2022, *ApJ*, 937, 115

Villaescusa-Navarro F., Wandelt B. D., Anglés-Alcázar D., Genel S., Zorrilla Mantilla J. M., Ho S., Spergel D. N., 2020, preprint ([arXiv:2011.05992](https://arxiv.org/abs/2011.05992))

Villaescusa-Navarro F. et al., 2021a, preprint ([arXiv:2109.09747](https://arxiv.org/abs/2109.09747))

Villaescusa-Navarro F. et al., 2021b, preprint ([arXiv:2109.10360](https://arxiv.org/abs/2109.10360))

Villaescusa-Navarro F. et al., 2021c, *ApJ*, 915, 71

Villaescusa-Navarro P. et al., 2021d, preprint ([arXiv:2111.14874](https://arxiv.org/abs/2111.14874))

Villaescusa-Navarro F. et al., 2022a, *ApJ*, 929, 132

Villaescusa-Navarro P. et al., 2022b, *ApJ*, 935, 30

Villaescusa-Navarro F. et al., 2023, *ApJS*, 265, 54

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, 436, 3031

Vogelsberger M. et al., 2014a, *MNRAS*, 444, 1518

Vogelsberger M. et al., 2014b, *Nature*, 509, 177

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, *Nature Rev. Phys.*, 2, 42

Wadekar D. et al., 2023a, *Proc. Natl. Acad. Sci.*, 120, e2202074120

Wadekar D. et al., 2023b, *MNRAS*, 522, 2628

Weinberger R. et al., 2017, *MNRAS*, 465, 3291

Weiss A. J., Schneider A., Sgier R., Kacprzak T., Amara A., Refregier A., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 11

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, *MNRAS*, 399, 574

Wright R. J., Lagos C. d. P., Power C., Mitchell P. D., 2020, *MNRAS*, 498, 1668

Wright R. J., Somerville R. S., Lagos C. d. P., Schaller M., Davé R., Anglés-Alcázar D., Genel S., 2024, preprint ([arXiv:2402.08408](https://arxiv.org/abs/2402.08408))

APPENDIX A: DARK MATTER SPREAD IN SELECTED HALO MASS RANGES

The spreading of dark matter particles within haloes shows a strong dependence on halo mass, but seemingly little direct dependence on cosmological parameters (Fig. 3). In Fig. A1, we show the full spread distributions of dark matter particles in haloes of the same mass ranges used in Fig. 3 for the fiducial N -body 1P simulation. As expected, dark matter particles in more massive haloes spread farther from their initial neighbours. The most extreme cases of spread can occur in any mass range, corresponding to particles whose initial neighbour ended up very far from the halo.

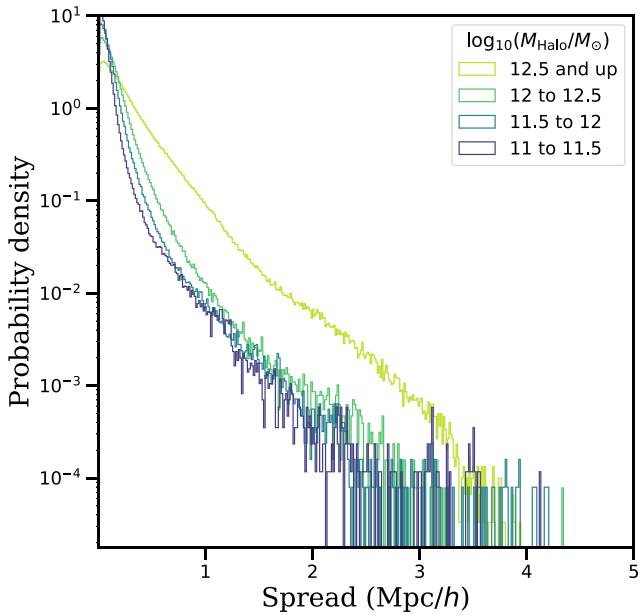


Figure A1. Distribution of spread distances for dark matter particles in haloes of selected mass ranges in the fiducial N -body simulation at $z = 0$. Larger haloes generally spread dark matter farther, but cases of extremely large spreads appear to be independent of halo mass.

APPENDIX B: GAS SPREAD IN AND OUT OF HALOES

We show in Fig. B1 the spreading of gas particles in haloes as a comparison to Fig. 2. We similarly select gas particles in the ten most massive haloes (while again excluding the most massive halo due to its outlying mass) and in the upper panel we plot the median spread of these particles against the average radius of the selected haloes. Gas that remains in haloes shows a very similar trend to dark matter, but spreads roughly twice the distance. In the bottom panel, we divide the spread of each gas particle by the spread of their initial dark matter neighbour to normalize out the gravitational spreading effect and see that the correlation with halo radius largely disappears.

We next show the spreading of matter outside of haloes in SIMBA. Baryonic approximation methods generally alter the distribution of matter on the scales of haloes, but some do account for matter ejected outside of haloes. However, halo-based models cannot account for the spreading of material that never accreted onto haloes. Fig. B2 shows the spreading of particles outside of Lagrangian regions (i.e. gas whose dark matter neighbour remained outside of a halo at $z = 0$). We show here that the spreading of gas outside of Lagrangian regions (solid lines) remains significant, and is only slightly smaller than all gas in the box (dotted lines). Thus, the effects of feedback on baryons outside of haloes represent a significant source of uncertainty for models that ignore this component.

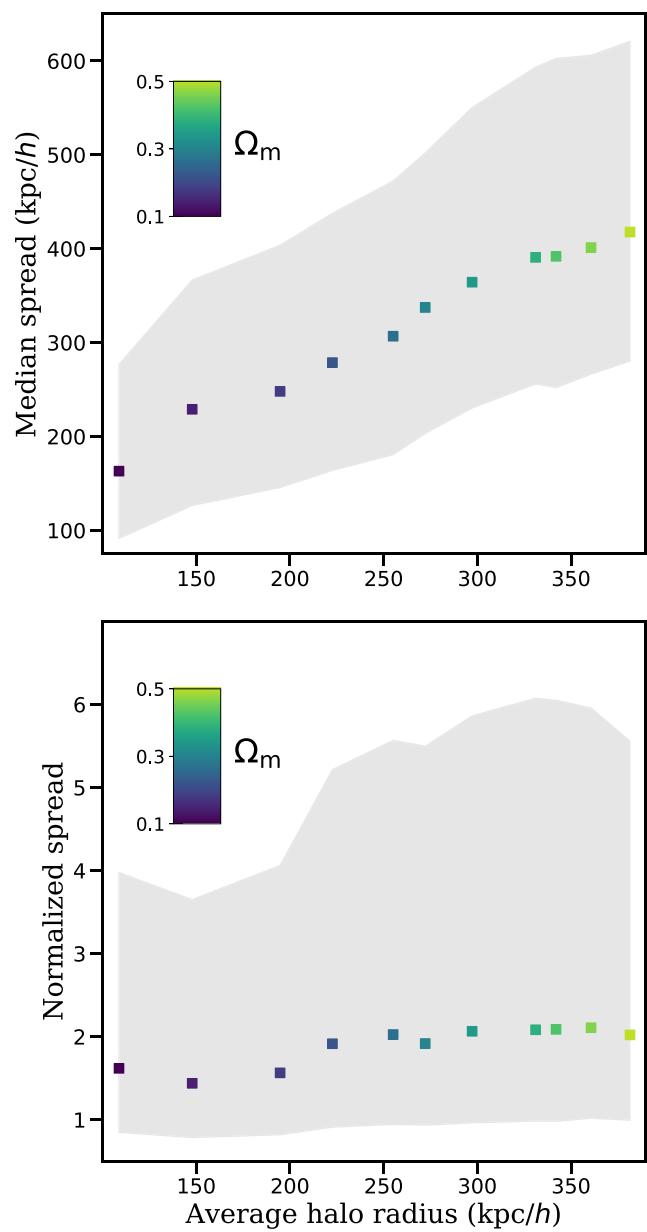


Figure B1. Median spread of gas particles in the 10 most massive haloes (excluding the most massive halo, top panel) and the median of the spread of these particles divided by the spread of their initial dark matter neighbour (bottom panel) as a function of the average virial radius among these haloes at $z = 0$ in each of the SIMBA 1P set simulations varying Ω_m . The colour bar indicates the value of Ω_m for each simulation. As Ω_m increases, haloes get larger and spreading increases. However, normalizing to the spread of dark matter neighbours largely suppresses the trend.

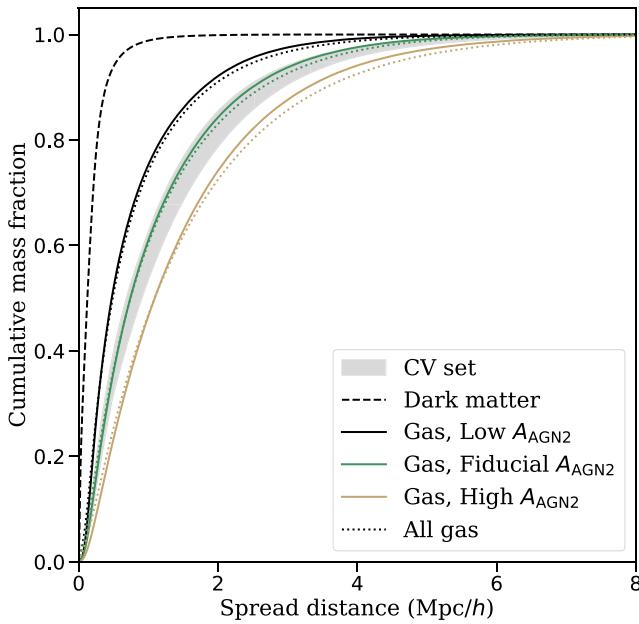


Figure B2. Cumulative mass distribution as a function of spread distance at $z = 0$ for dark matter (dashed) and gas (green) particles outside of any Lagrangian region in the fiducial SIMBA 1P simulation, the highest (brown) and lowest (black) $A_{\text{AGN}2}$ (jet speed) simulations, and the full SIMBA CV set (the grey shaded area spans the full CV set distribution). For comparison, the dotted lines represent all particles in the box (the same lines as in Fig. 5). In all cases, particles outside of Lagrangian regions spread comparable to the particles in the box, which represents a significant effect that baryonic approximation methods do not account for.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.