

Predicting the impact of feedback on matter clustering with machine learning in CAMELS

Ana Maria Delgado,¹★ Daniel Anglés-Alcázar,^{2,3} Leander Thiele[Ⓜ],⁴ Shivam Pandey,^{5,6} Kai Lehman,^{7,8} Rachel S. Somerville,³ Michelle Ntampaka,^{9,10} Shy Genel,³ Francisco Villaescusa-Navarro³ and Lars Hernquist¹

¹Center for Astrophysics, Harvard and Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

²Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269, USA

³Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

⁴Department of Physics, Princeton University, Jadwin Hall, Princeton, NJ 08544, USA

⁵Department of Physics, Columbia University, New York, NY 10027, USA

⁶Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

⁷Institute for Astronomy, University of Hawai'i, 2680 Woodlawn drive, Honolulu, HI 96822, USA

⁸Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 München, Germany

⁹Data Science Mission Office, Space Telescope Science Institute, Baltimore, MD 21218, USA

¹⁰Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA

Accepted 2023 September 22. Received 2023 September 15; in original form 2023 January 5

ABSTRACT

Extracting information from the total matter power spectrum with the precision needed for upcoming cosmological surveys requires unraveling the complex effects of galaxy formation processes on the distribution of matter. We investigate the impact of baryonic physics on matter clustering at $z = 0$ using a library of power spectra from the Cosmology and Astrophysics with Machine Learning Simulations project, containing thousands of $(25 h^{-1} \text{Mpc})^3$ volume realizations with varying cosmology, initial random field, stellar and active galactic nucleus (AGN) feedback strength and subgrid model implementation methods. We show that baryonic physics affects matter clustering on scales $k \gtrsim 0.4 h \text{Mpc}^{-1}$ and the magnitude of this effect is dependent on the details of the galaxy formation implementation and variations of cosmological and astrophysical parameters. Increasing AGN feedback strength decreases halo baryon fractions and yields stronger suppression of power relative to N -body simulations, while stronger stellar feedback often results in weaker effects by suppressing black hole growth and therefore the impact of AGN feedback. We find a broad correlation between mean baryon fraction of massive haloes ($M_{200c} > 10^{13.5} M_{\odot}$) and suppression of matter clustering but with significant scatter compared to previous work owing to wider exploration of feedback parameters and cosmic variance effects. We show that a random forest regressor trained on the baryon content and abundance of haloes across the full mass range $10^{10} \leq M_{\text{halo}}/M_{\odot} < 10^{15}$ can predict the effect of galaxy formation on the matter power spectrum on scales $k = 1.0\text{--}20.0 h \text{Mpc}^{-1}$.

Key words: methods: numerical – galaxies: haloes – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

The field of cosmology has many exciting endeavors to look forward to within the next decade. With the arrival of enormous photometric and spectroscopic galaxy redshift survey missions such as DESI (DESI Collaboration 2016), the *Nancy Roman Space Telescope* (Spergel et al. 2015), Euclid (Laureijs et al. 2011) and the Vera Rubin Observatory (LSST Science Collaboration 2009), the community will have the opportunity to tackle many ambitious goals, such as mapping the distribution of matter and the large-scale structure of the Universe, measuring cosmological parameters to percent-level precision, and constraining the sum of neutrino masses. An important step in fully realizing the statistical power of these

upcoming surveys is to model the matter power spectrum and other summary statistics to ~ 1 percent precision down to scales as small as $k = 10 h \text{Mpc}^{-1}$ (Huterer & Takada 2005; Laureijs 2009; Hearin, Zentner & Ma 2012). However, previous studies have shown that complex galaxy formation processes involving feedback from massive stars and active galactic nuclei (AGNs) can suppress power relative to dark matter-only simulations out to large scales (van Daalen et al. 2011; Chisari et al. 2018; van Daalen, McCarthy & Schaye 2020; Gebhardt et al. 2023). Galactic winds driven by supernovae and AGN-driven outflows can eject a large amount of material from the centre of galaxies out to large distances (Anglés-Alcázar et al. 2017b; Borrow, Anglés-Alcázar & Davé 2020; Hafen et al. 2020; Wright et al. 2020; Mitchell & Schaye 2022; Sorini et al. 2022; Ayromlou, Nelson & Pillepich 2023) and the resulting suppression of power by feedback creates significant biases when attempting to constrain cosmological

★ E-mail: ana_maria.delgado@cfa.harvard.edu

parameters (Semboloni et al. 2011; Zentner et al. 2013; Chisari et al. 2019).

Several approaches to addressing the suppression of matter clustering caused by baryonic physics have been devised. Cosmological hydrodynamic simulations provide the most direct method to understand the impact of baryonic effects on the distribution and clustering of matter (Vogelsberger et al. 2014b; Tenneti et al. 2015; Hellwing et al. 2016; Springel et al. 2018; Chisari et al. 2019). Modern cosmological large-volume simulations such as Horizon-AGN (Dubois et al. 2014), Eagle (Schaye et al. 2015), IllustrisTNG (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b), and SIMBA (Davé et al. 2019) produce galaxies that broadly match observations in properties such as the stellar mass function and the bimodality in galaxy colours. Comparing the power spectrum of hydrodynamic simulations with those of their phase-matched, collisionless N -body, dark matter-only simulations allows us to measure how baryonic feedback suppresses the clustering of matter. However, many key feedback processes remain poorly understood and most current models require extensive tuning of free parameters to match observations, limiting their predictive power (Somerville & Davé 2015). Higher resolution cosmological ‘zoom-in’ simulations can reduce subgrid model uncertainties (e.g. Agertz & Kravtsov 2016; Hopkins et al. 2018; Anglés-Alcázar et al. 2021), but at the expense of modelling volumes that are too small for many cosmological applications.

More flexible approaches to address the impact of baryonic physics using analytic models include modifying the ‘halo model’ (Seljak 2000; Semboloni, Hoekstra & Schaye 2013; Fedeli 2014; Mead et al. 2015) using observational constraints and simulation results as the basis for parameterizing the transfer of power produced by the presence of baryons (Mohammed & Seljak 2014; Schneider & Teyssier 2015), and mitigating the presence of baryons altogether by marginalizing over the parameters of effective models (Semboloni et al. 2011) or over the principle components in linear combinations of observables that are most strongly affected by baryonic effects (Eifler et al. 2015; Kitching et al. 2016). However, the success of these techniques relies heavily on the flexibility of the models to capture the true underlying distribution of matter (McCarthy et al. 2017) and they are limited by assumptions about halo bias relative to the linear density field, smooth halo profiles neglecting substructure, and uncertainties in the spatial and redshift dependence of baryonic effects (Chisari et al. 2019). Alternatively, power spectra produced by a large number of cosmological simulations with varying cosmologies and feedback parameters can be used to inform semi-analytic models attempting to mitigate the effects of baryons, characterize the theoretical uncertainties in galaxy formation, and marginalize over feedback effects.

van Daalen et al. (2011) employed a suite of 50 cosmological hydrodynamic simulations from the OWLS project (Schaye et al. 2010) to study the effects of different baryonic processes on the matter power spectrum over a range of scales. More recently, van Daalen et al. (2020, henceforth vDMS) included additional simulations from the cosmo-OWLS (Le Brun et al. 2014) and BAHAMAS (McCarthy et al. 2017) projects to produce a library of 92 matter power spectra from simulations with varying subgrid models and feedback strengths. Relating the effects of galaxy formation physics to the suppression of power, vDMS proposed that it is possible to predict the fractional impact of baryons on the clustering of matter, $P_{\text{hydro}}/P_{\text{DM}}$, given only the mean baryon fraction of massive haloes ($M_{\text{halo}} \sim 10^{14} M_{\odot}$), where P_{hydro} and P_{DM} are the matter power spectra from hydrodynamic simulations and their corresponding N -body simulations, respectively. Importantly, the empirical vDMS relation

between baryon fraction and power suppression is satisfied by a variety of simulations with different galaxy formation implementations, including the Horizon-AGN, EAGLE, and IllustrisTNG simulations, which opens the possibility to accurately correct dark matter only power spectra based on observational constraints on gas fractions in massive haloes. However, this relation is valid only on large scales, $k \leq 1 h \text{ Mpc}^{-1}$, and the still limited number of different feedback implementations and cosmologies represented in the vDMS library of matter power spectra may not be representative of a broader range of plausible galaxy formation models.

In this work, we use 2000 + cosmological hydrodynamic simulations and their corresponding collisionless (N -body) simulations from the Cosmology and Astrophysics with Machine Learning Simulations (CAMELS¹) project (Villaescusa-Navarro et al. 2021c) to examine the impact of baryonic physics on matter clustering using the largest library of power spectra available including variations of cosmological and feedback parameters. In recent related work using CAMELS, Nicola et al. (2022) trained a neural network on thousands of electron density autopower spectra from large scales down to $k = 10 h \text{ Mpc}^{-1}$, breaking the baryon–cosmology degeneracy and providing tight constraints on the total matter density Ω_m and the mean baryon fraction in intermediate-mass haloes while marginalizing over uncertainties in galaxy formation physics implementations. Here, we significantly expand upon the work of vDMS and investigate how supernova and AGN feedback affect the mean baryon fraction across a range of halo masses ($10^{10} \leq M_{\text{halo}}/M_{\odot} < 10^{15}$) and the resulting impact on the matter power spectrum. Furthermore, we take advantage of the design of CAMELS for machine learning and train a random forest (RF) regressor to predict the relative difference between the matter clustering in hydrodynamical and N -body simulations on scales $k = 1.0\text{--}20 h \text{ Mpc}^{-1}$ given the mean baryon fraction of haloes across a broad range of halo masses. We thus demonstrate that we are able to extract valuable information from lower mass haloes and predict the suppression of power all the way to the highly non-linear regime. The work presented here is complementary to Pandey et al. (2023), which show that information about the impact of baryonic effects on the matter power spectrum can be extracted using the tSZ signals from low-mass haloes, and include related results utilizing the suite of CAMELS produced with the Astrid simulation.

The layout of this paper is as follows: In Section 2, we describe the simulations and halo selection, define our variables, and describe our machine-learning methods. In Sections 3 and 4, we present our results. Finally, in Section 5, we provide a summary and discussion of our work.

2 METHODS

2.1 Simulations: CAMELS

The CAMELS project (Villaescusa-Navarro et al. 2021c) contains thousands of state-of-the-art (magneto-)hydrodynamic simulations and their corresponding N -body simulations. In this work, we focus on the simulation suites produced with the IllustrisTNG (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b) and SIMBA (Davé et al. 2019) galaxy formation models that are part of the CAMELS public data release² (Villaescusa-Navarro et al. 2023a), providing

¹<https://www.camel-simulations.org>

²<https://camels.readthedocs.io>

a total of >4000 realizations with parameter variations, with >1000 hydrodynamical and >1000 N -body simulations for each of the two independent feedback model implementations. Each simulation is a periodic box of length $L_{\text{box}} = 25 h^{-1} \text{Mpc}$ containing 256^3 resolution elements with mass resolution of $6.49 \times 10^7 (\Omega_m - \Omega_b)/0.251 h^{-1} \text{M}_\odot$ for dark matter and $1.27 \times 10^7 h^{-1} \text{M}_\odot$ for baryons. This is the same resolution as the original SIMBA simulation and similar to that of the original TNG300-1 simulation of IllustrisTNG.

The initial conditions of CAMELS simulations were generated at $z = 127$ using second-order Lagrangian perturbation theory assuming that the initial power spectra of dark matter and gas are the same and equal to that of total matter. Each of the CAMELS simulations contains 34 snapshots from redshifts $z = 6$ down to $z = 0$; in this work, we focus on $z = 0$. In addition to the initial random phases, each simulation is specified by two cosmological parameters and four astrophysical (feedback) parameters which are varied across the individual realizations. In the case of cosmological parameters, we vary:

- (i) Ω_m : the fraction of the universe made up of ordinary and dark matter varies in the range $\Omega_m \in [0.1, 0.5]$ while keeping $\Omega_b = 0.049$ constant.
- (ii) σ_8 : the variance of the spatial fluctuations of total matter on $8 \text{ Mpc } h^{-1}$ scales is varied in the range $\sigma_8 \in [0.6, 1.0]$.

In the case of astrophysical parameters, the fiducial values are defined by the stellar and AGN feedback models of the corresponding original IllustrisTNG and SIMBA simulations. The fiducial astrophysical parameters are assigned a value $A = 1.0$ and then varied across realizations by multiplying by an amplitude factor A in order to increase/decrease the amount of feedback. However, we emphasize that the stellar and AGN feedback prescriptions differ substantially between IllustrisTNG and SIMBA and the corresponding parameter variations in CAMELS have a different definition in each model, which we briefly describe below.

2.1.1 IllustrisTNG

The IllustrisTNG model (also referred to as ‘TNG’; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b) is implemented in the AREPO hydrodynamics code (Springel 2010; Weinberger, Springel & Pakmor 2020), which utilizes a hybrid tree/particle-mesh scheme to solve for gravitational interactions and an unstructured, moving mesh to solve the equations of hydrodynamics. Compared to the galaxy formation model of its predecessor *Illustris* (Genel et al. 2014; Vogelsberger et al. 2014a, b), the galaxy formation model in IllustrisTNG has updated implementations of AGN feedback (Weinberger et al. 2017) and galactic winds (Pillepich et al. 2018a), and incorporates magnetic fields (Pakmor, Marinacci & Springel 2014).

The stellar feedback parameter variations in the CAMELS–TNG simulations introduce A_{SN1} to control the total energy injection rate in galactic winds per unit star formation ($A_{\text{SN1}} \in [0.25, 4.0]$) and A_{SN2} to vary the galactic wind speed ($A_{\text{SN2}} \in [0.5, 2.0]$). The AGN feedback parameter variations pertain to the low-accretion, kinetic-mode black hole feedback, where A_{AGN1} varies the feedback energy per unit black hole accretion rate ($A_{\text{AGN1}} \in [0.25, 4.0]$) and A_{AGN2} varies the burstiness and effective ejection speed ($A_{\text{AGN2}} \in [0.5, 2.0]$).

2.1.2 SIMBA

The SIMBA galaxy formation model (Davé et al. 2019) is implemented in the GIZMO meshless finite mass hydrodynamics code (Hopkins 2015, 2017). Relative to its predecessor MUFASA (Davé, Thompson & Hopkins 2016), SIMBA includes a black hole model based on gravitational torque accretion and two-mode kinetic feedback (Anglés-Alcázar et al. 2017a), galactic winds with mass-loading and velocity scalings derived from the FIRE zoom-in simulations (Muratov et al. 2015; Anglés-Alcázar et al. 2017b), and a model for the creation and destruction of dust (Li, Narayanan & Davé 2019).

The stellar feedback parameter variations in the CAMELS–SIMBA simulations introduce A_{SN1} to control the mass-loading factor of galactic winds and A_{SN2} to control the wind speed. The AGN feedback parameter variations introduce A_{AGN1} to change the total momentum flux of either quasar-mode winds or radio-mode jets, while A_{AGN2} controls the maximum velocity of gas ejected by jets. These parameters are varied over the same range as in IllustrisTNG, with $A_{\text{SN1}} = A_{\text{SN2}} = A_{\text{AGN1}} = A_{\text{AGN2}} = 1$ corresponding to the fiducial model. As described in Villaescusa-Navarro et al. (2021c), the range of feedback parameters explored in CAMELS was chosen to roughly produce factor-of-two variations of injected feedback energy relative to the fiducial models, as a compromise between investigating a wide range of feedback effects while still considering physically plausible models. We also stress that despite using the same range of parameter variations in IllustrisTNG and SIMBA, the resulting effects are model dependent (as shown below and previous works) and reflect their specific implementation.

2.1.3 Simulation sets in CAMELS

We take advantage of the following simulation sets for each of the IllustrisTNG and SIMBA suites in CAMELS:

- (i) Latin hypercube (‘LH’) set: 1000 realizations, each containing different initial conditions and different values of the six aforementioned parameters. The LH set is the main training set in this work.
- (ii) 1 parameter (‘1P’) set: 66 realizations using the same initial conditions and further divided into six subsets of 11 realizations where only the value of one parameter is varied while the other five parameters are held constant. In this work, we make use of the 1P sets to study how a single cosmological or feedback parameter can affect halo baryon fractions and the suppression of the matter power spectrum.
- (iii) Cosmic variance (‘CV’) set: 27 realizations with different initial conditions while the fiducial values of all six parameters are held constant. The CV set is used to evaluate the impact of CV on any of the quantities that we measure from the simulations.

We refer the reader to Villaescusa-Navarro et al. (2021c) for further details about CAMELS, the parameter variations, and the simulation sets available.

2.2 Halo selection

We identify haloes in CAMELS using the AMIGA Halo Finder (AHF; Knollmann & Knebe 2011). AHF uses an adaptive mesh to locate halo centres, calculate the gravitational potential of the halo and iteratively remove unbound particles (particles whose velocities are greater than the escape velocity at a given radius) from within the boundary of the halo. We refer the reader to Knollmann & Knebe

(2011) for a full description and implementation of AHF. We select haloes with masses $M_{\text{halo}} \geq 10^{10} M_{\odot}$ using a virial radius definition of 200c (i.e. 200 times the critical density of the Universe).

2.3 Matter power spectra and halo baryon fractions

We use a library of >4000 total matter power spectra from CAMELS (Villaescusa-Navarro et al. 2023b). For each simulation, the matter power spectrum is computed by assigning particle masses (dark matter, gas, stars, and black holes) to a regular grid with 512^3 voxels using a cloud-in-cell (CIC) assignment scheme. The grid is then Fourier transformed and the power spectrum is computed by averaging over k -bins with an equal width to the fundamental frequency, $k_F = 2\pi/L$, where $L = 25h^{-1}\text{Mpc}$. We then compute the relative difference between the total matter power spectrum of hydrodynamical and phase-matched N -body simulations, which we refer to as the ‘suppression of matter power spectrum’ and define as

$$\frac{\Delta P}{P_{\text{DM}}} = \frac{P_{\text{hydro}} - P_{\text{DM}}}{P_{\text{DM}}}, \quad (1)$$

where P_{DM} is the matter power spectrum of the N -body simulation and P_{hydro} is that of its corresponding hydrodynamical simulation.

We compute the baryon fraction of a given halo as

$$f_{\text{bar}} = \frac{M_{\text{star}} + M_{\text{gas}}}{M_{\text{halo}}}, \quad (2)$$

where M_{star} and M_{gas} are the total stellar mass and gas mass of the halo and M_{halo} is the virial mass of the halo corresponding to R_{200c} . We further calculate the mean baryon fraction within a given halo mass range in each simulation as

$$\bar{f}_{\text{bar}} = \frac{1}{n} \sum_{i=1}^n f_{\text{bar}_i} \frac{\Omega_b}{\Omega_m}, \quad (3)$$

where f_{bar} is defined in equation (2), subscript i is the i th halo and n the total number of haloes in a given mass range, and following vDMS we normalize by Ω_b/Ω_m in order to account for the differences in cosmology for different simulations.

2.4 Machine learning

A supervised machine-learning algorithm trains a model by providing a subset of data, referred to as the training set, including input variables (henceforth called ‘features’) and output variables (henceforth called ‘target’). The goal is for the algorithm to use the training set to learn the relation between the features and the target. The trained model is then used to predict the target for a different subset of features referred to as the test set.

In this work, we use the RF regressor algorithm from the publicly available package `Scikit-Learn` (Pedregosa et al. 2011). An RF is an ensemble ML method that can be used for both classification and regression problems. The algorithm works by constructing a ‘forest’ from a user specified number of decision trees and using the mean of the predictions from those trees as output. This method has three key advantages: (1) little hyperparameter tuning is required, (2) it is computationally efficient, and (3) its ensemble characteristic lessens over fitting. Furthermore, the RF algorithm provides us with some interpretability by way of the ‘feature importance’ attribute, with a ranking of features based on their frequency used as a predictor variable by each tree.

We use the following metrics for scoring the predictive performance of the RF:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4)$$

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad (5)$$

where y_i are the given target values, \hat{y}_i are the RF predicted target values, and \bar{y} is the mean of y_i . The R^2 score provides the proportion of the target variable that is predictable by the given features. Because the R^2 outputs a score between 0.0 and 1.0, it provides comparable information about performance when comparing various experiments. The RMSE scores, on the other hand, are based on the target value range. Therefore, in order to account for the range in target values across multiple experiments, we normalize our RMSE scores by the interquartile range (IQR):

$$\text{IQR} = Q_3 - Q_1, \quad (6)$$

where Q_3 is the third quartile (75th percentile) of a given set and Q_1 is the first quartile (25th percentile) of the set. In other words, we normalize the RMSE by the middle 50 per cent dispersion of the target values as RMSE/IQR .

2.4.1 Features and targets

Using the thousands of realizations in the CAMELS LH simulation sets, we train an RF regressor to predict the suppression of the matter power spectrum $\Delta P/P_{\text{DM}}$ as a function of the mean baryon fraction at a range of scales. We construct the following features for each realization:

- (i) $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$: the mean baryon fraction of high-mass haloes, those with masses $> 10^{13.5} M_{\odot}$ in each simulation.
- (ii) $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$: an array containing mean baryon fraction in 10 bins of halo masses within the range $[10^{10} - 10^{15}] M_{\odot}$.
- (iii) N_{halo} : the number of haloes within a halo mass range.
- (iv) N_{halo}^j : an array with the same shape as $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ containing the number of haloes per mass bin.

Our target is the $\Delta P/P_{\text{DM}}$ values for each realization at four different k -values: $k = [1, 5, 10, 20] h \text{ Mpc}^{-1}$.

2.4.2 Robustness of RF predictions

One inherent benefit of CAMELS is that we are able to test the effects of feedback model implementation by way of its TNG and SIMBA simulations sets. We create an 80/20 per cent train/test split of the LH simulations and perform the following experiments using either TNG or SIMBA:

- (i) Train on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ to predict $\Delta P/P_{\text{DM}}$ at $k = [1, 5, 10, 20] h \text{ Mpc}^{-1}$. We perform this experiment as an extension of vDMS, where in this work we explore a larger halo mass range and probe the non-linear regime.
- (ii) Train on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ at $k = [1, 5, 10, 20] h \text{ Mpc}^{-1}$.
- (iii) Repeat the above two experiments with additional features related to CV: N_{halo} or N_{halo}^j .

We are further able to test the robustness of our algorithm and determine how well our the RF can marginalize over subgrid physics model by performing ‘two-model’ experiments. In these experiments, we train on the entire LH set of one of the feedback implementation and test on the entire LH set of the other, i.e. training

on SIMBA and testing on TNG and vice versa. In these experiments, we use the same hyperparameters that produced the best results from the previous ‘single-model’ experiments.

3 IMPACT OF COSMOLOGICAL AND FEEDBACK PARAMETER VARIATIONS

In this section, we examine the impact of cosmological and baryonic feedback parameter variations on the matter power spectrum and the baryon fraction of haloes of different masses, exploring also the connection between the suppression of the total matter power spectrum and the mean baryon fraction of massive haloes. We perform this analysis for both the TNG and SIMBA galaxy formation models.

3.1 Matter power spectra

We use the total matter power spectra from the 1P simulations to examine how cosmological and feedback parameters affect the clustering of matter at various scales. For each run in the 1P set, described in Section 2.1, we measure the fractional impact of baryons on the total matter power spectrum, $P_{\text{hydro}}/P_{\text{DM}}$. If baryonic physics has no effect on matter clustering, $P_{\text{hydro}}/P_{\text{DM}}$ should be of order unity. However, if baryonic physics suppresses the clustering of matter compared to dark matter only simulations, usually by way of feedback ejecting gas out to large distances, $P_{\text{hydro}}/P_{\text{DM}}$ should fall below unity on a range of scales.

Fig. 1 shows $P_{\text{hydro}}(k)/P_{\text{DM}}(k)$ as a function of wavenumber k for the 1P simulations, where each spectrum is colour coded by the value of each parameter variation. We notice two overall trends that are roughly independent of cosmological or feedback parameters. The first is the general ‘scoop’ shape of $P_{\text{hydro}}/P_{\text{DM}}$, which is consistent with previous works (van Daalen et al. 2011; Hellwing et al. 2016; Chisari et al. 2018; Peters et al. 2018; Springel et al. 2018; vDMS). This shape conveys agreement between matter clustering in hydrodynamical and dark matter-only simulations on large scales ($k \lesssim 0.1 h \text{ Mpc}^{-1}$) while at intermediate scales there is suppression of power by baryonic feedback ($P_{\text{hydro}}/P_{\text{DM}} < 1$) and at small scales ($k \gtrsim 40 h \text{ Mpc}^{-1}$) there is enhanced, as opposed to suppressed, clustering relative to dark matter owing to gas dissipative processes ($P_{\text{hydro}}/P_{\text{DM}} > 1$). The second overall trend is that the SIMBA galaxy formation model (solid lines) tends to suppress power on intermediate scales more strongly compared to the TNG galaxy formation model (dashed lines) while driving a steeper increase in small-scale clustering ($k \gtrsim 30 h \text{ Mpc}^{-1}$).

We now analyse in more detail how each parameter affects matter clustering by comparing $P_{\text{hydro}}(k)/P_{\text{DM}}(k)$ between the fiducial models of TNG and SIMBA (shown in red) and that of the individual parameter variations:

(i) *Cosmological parameters*: The top two panels in Fig. 1 show the sensitivity of the total matter power spectrum to Ω_m and σ_8 for a fixed galaxy formation model. We see a strong dependence of $P_{\text{hydro}}/P_{\text{DM}}$ on the value of Ω_m both in TNG and SIMBA. As Ω_m decreases (at fixed Ω_b), there is a greater suppression of power on intermediate scales. This can be understood as a consequence of baryons contributing a higher fraction of the total matter content making feedback more efficient at pushing gas out of haloes and distributing matter on larger scales, in agreement with the analysis of large-scale baryon spread in Gebhardt et al. (2023). In contrast, we identify weaker trends for σ_8 , with significant scatter.

(ii) *Supernova feedback parameters*: The middle two panels in Fig. 1 show the impact of changing the stellar feedback parameters A_{SN1} and A_{SN2} , respectively, which control the mass loading and velocity of galactic winds, on matter clustering. Both panels show somewhat counterintuitive effects of stellar feedback. Increasing A_{SN1} in SIMBA reduces (rather than enhances) the suppression of power on small scales ($k \gtrsim 10 h \text{ Mpc}^{-1}$) and increasing A_{SN2} further increases $P_{\text{hydro}}/P_{\text{DM}}$ over the full range of scales. This can be understood as a consequence of the non-linear interplay between stellar and AGN feedback, where stronger stellar feedback suppresses black hole growth and results in weaker effective impact of AGN feedback on matter clustering (van Daalen et al. 2011; Pandey et al. 2023). The TNG model shows rather different trends, with reduced suppression of power on scales $k \lesssim 10\text{--}20 h \text{ Mpc}^{-1}$ but enhanced suppression of power on smaller scales when increasing A_{SN1} and A_{SN2} . These results are consistent with the analysis of electron power spectra in CAMELS by Nicola et al. (2022), which highlights the sensitivity of predicted baryonic effects on galaxy formation implementation.

(iii) *AGN feedback parameters*: The bottom two panels of Fig. 1 show the impact of varying AGN feedback efficiency on matter clustering. In this case, there are clear systematic trends for stronger suppression of power when increasing both A_{AGN1} and A_{AGN2} for both galaxy formation models (TNG and SIMBA). The sensitivity of $P_{\text{hydro}}/P_{\text{DM}}$ to A_{AGN1} is weaker given its range of variation, with no more than 10 per cent difference relative to the fiducial model. In contrast, the matter power spectrum in SIMBA displays a strong sensitivity to the AGN jet speed, A_{AGN2} , with strong suppression of power across scales, reaching $P_{\text{hydro}}/P_{\text{DM}} \sim 0.6$ at $k \sim 10 h \text{ Mpc}^{-1}$ with jets twice as fast relative to the fiducial model. This results are also consistent with previous findings for electron power spectra Nicola et al. (2022) and the impact of large-scale jets on cosmological baryon spread (Gebhardt et al. 2023).

3.2 Halo baryon fraction

In Fig. 2, we use again the specialized CAMELS 1P simulation sets to analyse the impact of individual cosmological and feedback parameter variations on the average halo baryon fraction as a function of halo mass, $\bar{f}_{\text{bar}}(M_{\text{halo}})$, where we consider logarithmically spaced halo mass bins in the range $10^{10}\text{--}10^{13} M_{\odot}$. We notice two main trends roughly independent of cosmological or feedback parameters when comparing the fiducial realizations (indicated in red) for the TNG (dashed lines) and SIMBA (solid lines) models. The first being that peak of the halo baryon fraction occurs at $\sim 10^{12} M_{\odot}$. We notice a drop in mean baryon fraction as haloes exceed this mass range, when powerful feedback process can expel material out of the halo. We note, however, that at very high-mass haloes, we expect feedback to be less efficient at expelling material and for there to be another rise in mean baryon fraction. The second main trend is that SIMBA has overall lower $\bar{f}_{\text{bar}}(M_{\text{halo}})$ compared to TNG, with the fiducial models reaching their peak at $\bar{f}_{\text{bar}} \sim 0.5$ for SIMBA and $\bar{f}_{\text{bar}} \sim 0.7$ for TNG.

We now analyse in more detail how each CAMELS parameter variation affects halo baryon fractions, keeping in mind that the definitions of feedback parameters are not the same for TNG and SIMBA:

(i) *Cosmological parameters*: The top two panels of Fig. 2 show the sensitivity of $\bar{f}_{\text{bar}}(M_{\text{halo}})$ to our cosmological parameters Ω_m and σ_8 . Halo baryon fractions appear to be more sensitive to cosmology in SIMBA compared to TNG. However, both galaxy formation models

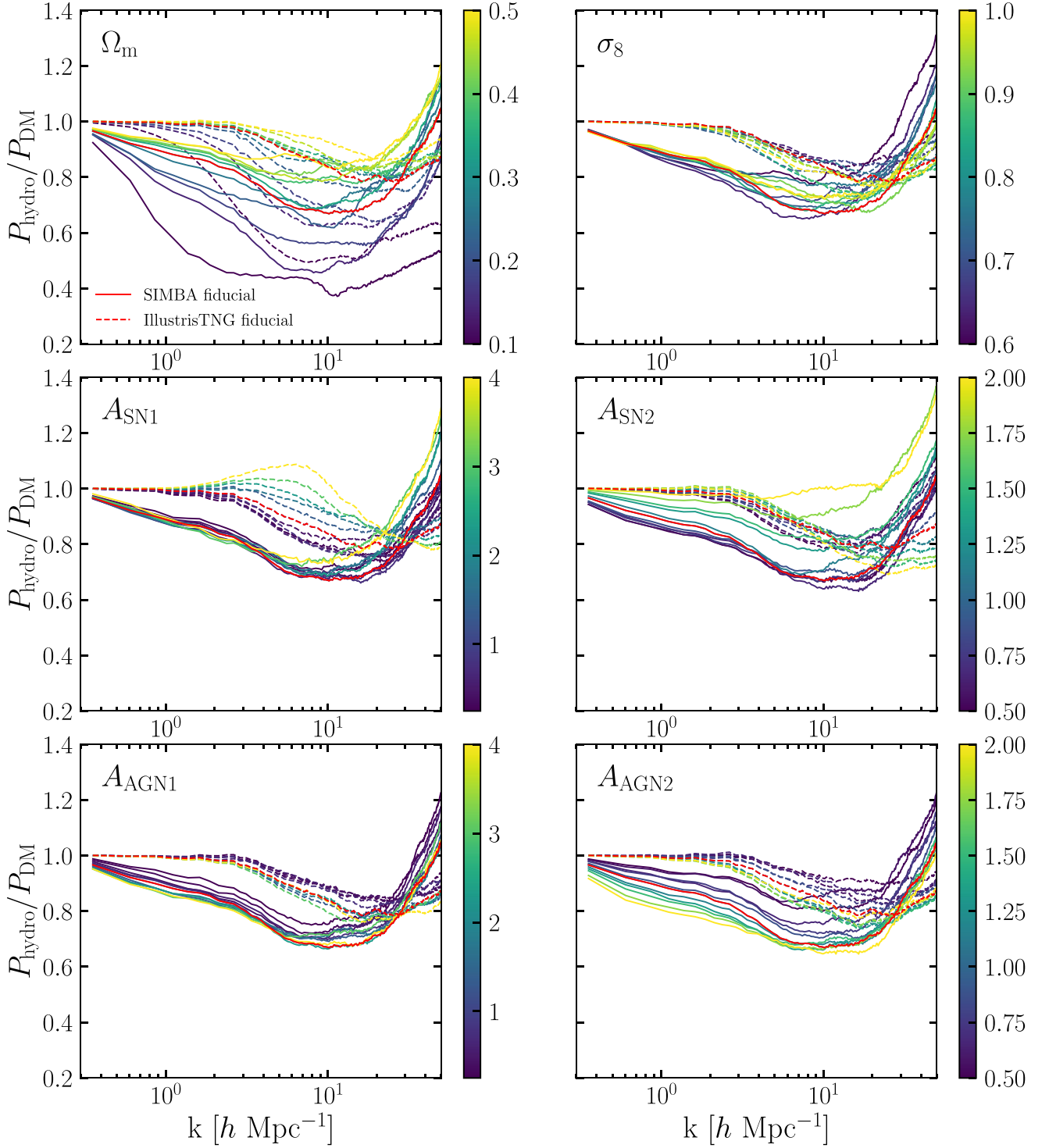


Figure 1. The effect of baryonic physics on matter clustering for different cosmological and feedback parameter variations. Each panel shows the ratio of total matter power spectrum in hydrodynamic simulations to that of the corresponding dark matter-only simulations ($P_{\text{hydro}}/P_{\text{DM}}$) as a function of wave number k when varying a single parameter in the CAMELS 1P sets. Lines of different colours indicate the value of each parameter variation, and red lines indicate the fiducial model for TNG (dashed lines) and SIMBA (solid lines). Variations in feedback model, as well as in feedback amplitude, result in variation in total matter clustering.

predict qualitatively similar trends, with lower $\bar{f}_{\text{bar}}(M_{\text{halo}})$ when increasing Ω_m and σ_8 across a range of halo masses. This trend may seem trivial for Ω_m since we hold the value of Ω_b constant in all CAMELS simulations, implying that the average cosmic baryon

fraction decreases with higher Ω_m and so should the corresponding halo baryon fractions. However, $\bar{f}_{\text{bar}}(M_{\text{halo}})$ is normalized by Ω_b/Ω_m for each simulation (equation 3), removing the trivial effect of varying Ω_m at fixed Ω_b . The impact of increasing Ω_m on $\bar{f}_{\text{bar}}(M_{\text{halo}})$ is thus a

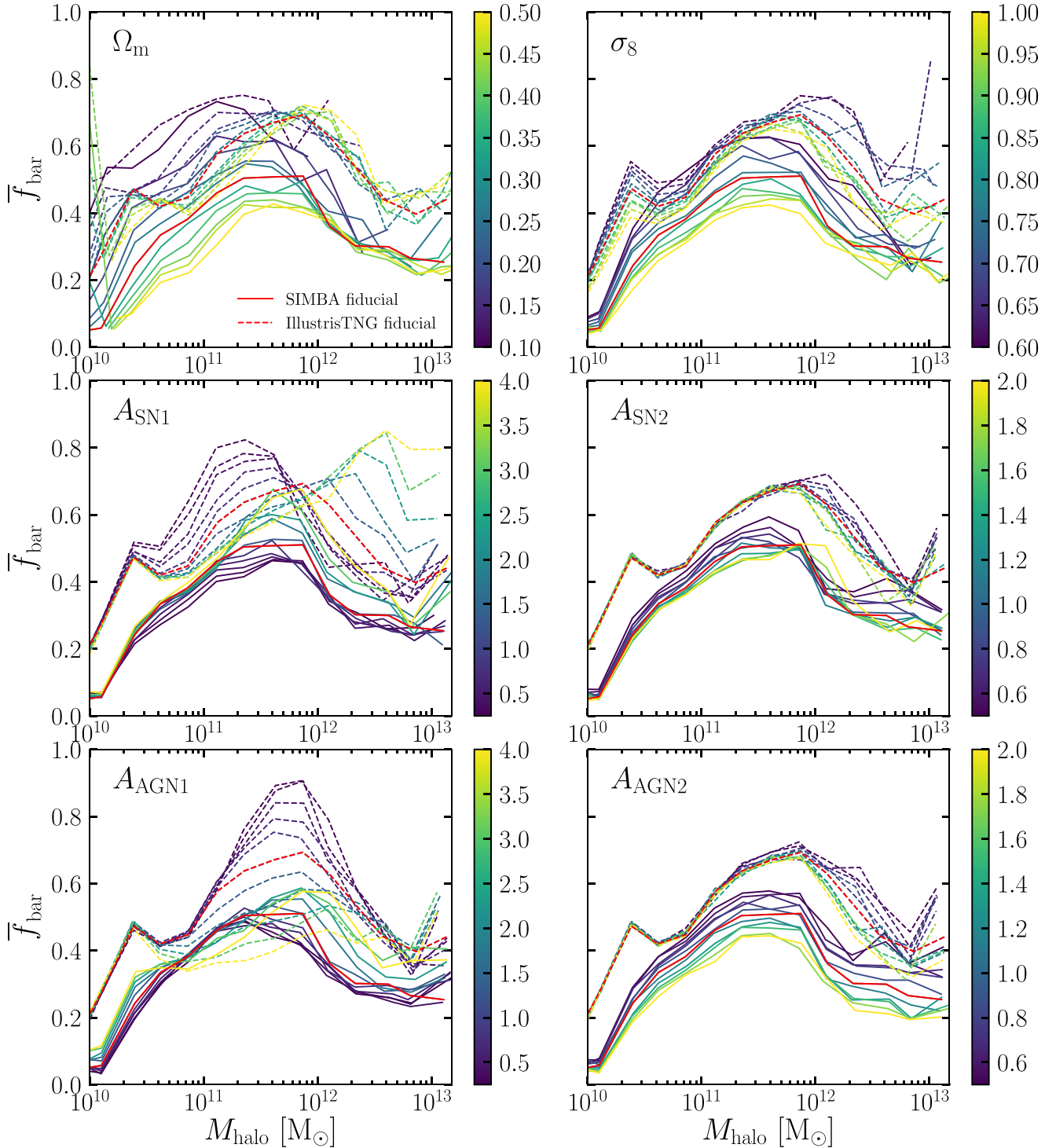


Figure 2. The effect of each parameter variation on the mean halo baryon fraction as a function of halo mass, $\bar{f}_{\text{bar}}(M_{\text{halo}})$. As in Fig. 1, the colour bar for each panel indicates the value of the corresponding parameter variation while all other parameters are held constant. The red lines indicate the fiducial parameters for TNG (dashed lines) and SIMBA (solid lines), with their peak baryon fraction occurring at $\lesssim 10^{12} M_{\odot}$ in both fiducial models. The \bar{f}_{bar} values and their halo mass dependence differ substantially between galaxy formation implementations and model parameter variations.

reflection of the effective efficiency of feedback when changing the amount of baryons relative to the dark matter gravitational potential, and this effect seems more prominent in lower mass haloes for both TNG and SIMBA. Interestingly, the baryon fraction decreases systematically at all halo masses when increasing σ_8 , while the

suppression of power does not seem to follow a clear trend with σ_8 .

(ii) *Supernova feedback parameters:* The middle two panels in Fig. 2 show the impact of changing A_{SN1} and A_{SN2} on halo baryon fractions. As for the power spectra, varying the mass loading of

galactic winds (A_{SN1}) has a different effect in each galaxy formation model. We might intuitively expect that as A_{SN1} increases, more gas would be ejected out of galaxies resulting in lower $\bar{f}_{\text{bar}}(M_{\text{halo}})$. However, we only see this behavior in TNG for haloes with mass $M_{\text{halo}} \lesssim 10^{12} M_{\odot}$, while the baryon fraction of higher mass haloes increases with A_{SN1} owing to the suppression of AGN feedback. This reversed trend with A_{SN1} can explain its impact on $P_{\text{hydro}}/P_{\text{DM}}$ for TNG, with an enhancement of power on intermediate scales coinciding with the highest gas fraction in high-mass haloes with $A_{\text{SN1}} = 4$. Meanwhile, SIMBA shows systematically higher baryon fractions when increasing A_{SN1} across the full halo mass range, indicating a different non-linear coupling of stellar and AGN feedback compared to TNG, which can explain the suppression of power seen in Fig. 1. On the other hand, increasing the speed of galactic winds (A_{SN2}) results in systematically lower $\bar{f}_{\text{bar}}(M_{\text{halo}})$ values for both SIMBA and TNG, but in this case lower baryon fractions correlate with less suppression of matter clustering in SIMBA on all scales.

(iii) *AGN feedback parameters*: The bottom two panels of Fig. 2 show the sensitivity of $\bar{f}_{\text{bar}}(M_{\text{halo}})$ to changes in AGN feedback efficiency. Halo baryon fractions are significantly reduced by increasing the kinetic mode black hole feedback efficiency A_{AGN1} in TNG in the intermediate halo mass range $M_{\text{halo}} = 10^{11} - 10^{12.5} M_{\odot}$, while the burstiness parameter A_{AGN2} has a stronger effect reducing $\bar{f}_{\text{bar}}(M_{\text{halo}})$ in higher mass haloes. In both cases, the decrease in halo baryon fraction with higher AGN feedback efficiency correlates with stronger suppression of matter clustering. Similarly, increasing the AGN jet speed in SIMBA (A_{AGN2}) drives an overall reduction of halo baryon fractions and increased suppression of matter clustering on all scales shown in Fig. 1, corresponding to more efficient spread of baryons on large scales relative to the TNG model (Gebhardt et al. 2023; Tillman et al. 2023). However, the effect of increasing the momentum flux A_{AGN1} in SIMBA seems more complex, driving an increase in baryon fraction in high-mass haloes (possibly due to black hole self-regulation) but stronger suppression in the matter power spectrum, particularly at low k -values.

3.3 Suppression of matter power spectrum as a function of baryon fraction

Using a suite of matter power spectra from hydrodynamical and dark matter only simulations, vDMS found a tight relation between the suppression of the matter power spectrum ($\Delta P/P_{\text{DM}}$; defined in Section 2.3) in the linear regime and the average baryon fraction (\bar{f}_{bar}) of high-mass haloes ($\sim 10^{14} M_{\odot}$). In this study, a direct comparison to vDMS is not possible due to the small size of the individual CAMELS realizations and relatively small number of high-mass haloes. We therefore extend what was done in vDMS and use the LH simulation sets of CAMELS, described in Section 2.1, to investigate how $\Delta P/P_{\text{DM}}$ is affected by cosmological and baryonic feedback parameters over a broader range of model variations.

Figs 3 and 4 show $\Delta P/P_{\text{DM}} \equiv (P_{\text{hydro}} - P_{\text{DM}})/P_{\text{DM}}$ evaluated at $k = 1.0 h \text{ Mpc}^{-1}$ as a function of \bar{f}_{bar} for the TNG and SIMBA 1P and LH sets, respectively. We note that due to the small simulated volumes in CAMELS there are not enough haloes of mass $\sim 10^{14} M_{\odot}$ to replicate the results in vDMS and we therefore evaluate \bar{f}_{bar} for haloes with mass $> 10^{13.5} M_{\odot}$. Each panel reproduces the same data points depicting the LH sets (small dots), overlaid by the 1P set (large circles) corresponding to the labelled parameter in that panel. The data are colour coded by the parameter value. We remind the reader that each of the six parameters are simultaneously varied in the LH sets, while only one parameter is varied in the 1P sets. We thus examine how individual cosmological and feedback

parameters affect the relation between $\Delta P/P_{\text{DM}}$ and \bar{f}_{bar} , which we compare to the fitting function derived by vDMS (their equation 5) for baryon fractions calculated using the 200c virial definition, which we henceforth refer to as the ‘vDMS model’ and indicate by the blue solid line and grey-shaded region. Lastly, we overlay the results for the CV sets of TNG and SIMBA (described in Section 2.1) as the red squares in the top centre panel of each figure in order to examine the effect of CV on this relation. We emphasize that this represents a lower limit of CV as the 27 CV realizations probe only a small volume. We have indicated where the fiducial realization (cyan star) lies within the spread. It is interesting to note that there is significant spread due to CV in both $\Delta P/P_{\text{DM}}$ and \bar{f}_{bar} in SIMBA compared to the fiducial run.

Our CAMELS results in Figs 3 and 4 reveal a good qualitative agreement with the general trend found in vDMS: the suppression of the matter power spectrum increases as the average baryon fraction in massive haloes decreases. We note that we have kept the y-axis limits the same in both figures for a more clear comparison of TNG and SIMBA, however, there are several SIMBA data points in Fig. 3 that fall below the visible y-axis. While the SIMBA LH set probes a range of $\Delta P/P_{\text{DM}}$ and \bar{f}_{bar} values significantly larger than the TNG LH set, as expected from Figs 1 and 2, both models roughly follow the vDMS trend, suggesting that \bar{f}_{bar} in massive haloes can be used to infer the redistribution of baryons over large scales regardless of galaxy formation model implementation. The location of the TNG fiducial run is in strong agreement with the vDMS model. The fiducial run for SIMBA, however, appears to fall outside of the vDMS fit line, and is further discussed in Section 5 of this paper. However, we find considerable spread in $\Delta P/P_{\text{DM}}$ at fixed \bar{f}_{bar} compared to vDMS, which can be attributed to the broader range of parameter variations explored in CAMELS:

(i) *Cosmological parameters*: The left two panels in Figs 3 and 4 explore the dependence of the $\Delta P/P_{\text{DM}} - \bar{f}_{\text{bar}}$ relation on Ω_m and σ_8 . We remind the reader that \bar{f}_{bar} is normalized by Ω_b / Ω_m . For both TNG and SIMBA, the LH sets show a trend of higher $\Delta P/P_{\text{DM}}$ (i.e. less suppression of power) at fixed \bar{f}_{bar} for higher values of Ω_m in high-mass haloes. This implies that the same impact on the total matter power spectrum (at $k = 1.0 h \text{ Mpc}^{-1}$) can be predicted by simulations that yield different halo baryon fractions, in this case as a consequence of the different response of feedback to changes in Ω_m at fixed Ω_b . We further note that the LH data points which stray furthest from the vDMS model correspond to the lower end values of the Ω_m variation. The SIMBA 1P set more clearly shows the trend of decreased suppression of the matter power spectrum as a function of mean baryon fraction with increased value of Ω_m in high-mass haloes.

There is also a visible, albeit less pronounced, trend for σ_8 , where $\Delta P/P_{\text{DM}}$ becomes more negative (i.e. stronger suppression of power) at fixed \bar{f}_{bar} for higher values of σ_8 . Overall, the non-linear response of the fiducial galaxy formation model to variations in cosmology appears to explain a significant fraction of the scatter in the vDMS relation seen for SIMBA and TNG.

(ii) *Supernova feedback parameters*: The middle panels of Figs 3 and 4 explore the dependence of the vDMS relation on systematic variations of the mass loading factor and speed of galactic winds driven by stellar feedback (parameters A_{SN1} and A_{SN2} , respectively). In TNG, there is indication for simulations clustering around $\Delta P/P_{\text{DM}} \sim 0$ and $\bar{f}_{\text{bar}} \sim 1$ for higher values of A_{SN1} and A_{SN2} , corresponding to weaker overall impact of feedback owing to the suppression of black hole growth and therefore AGN feedback. We find qualitatively similar trends in SIMBA for variations in A_{SN2} ,

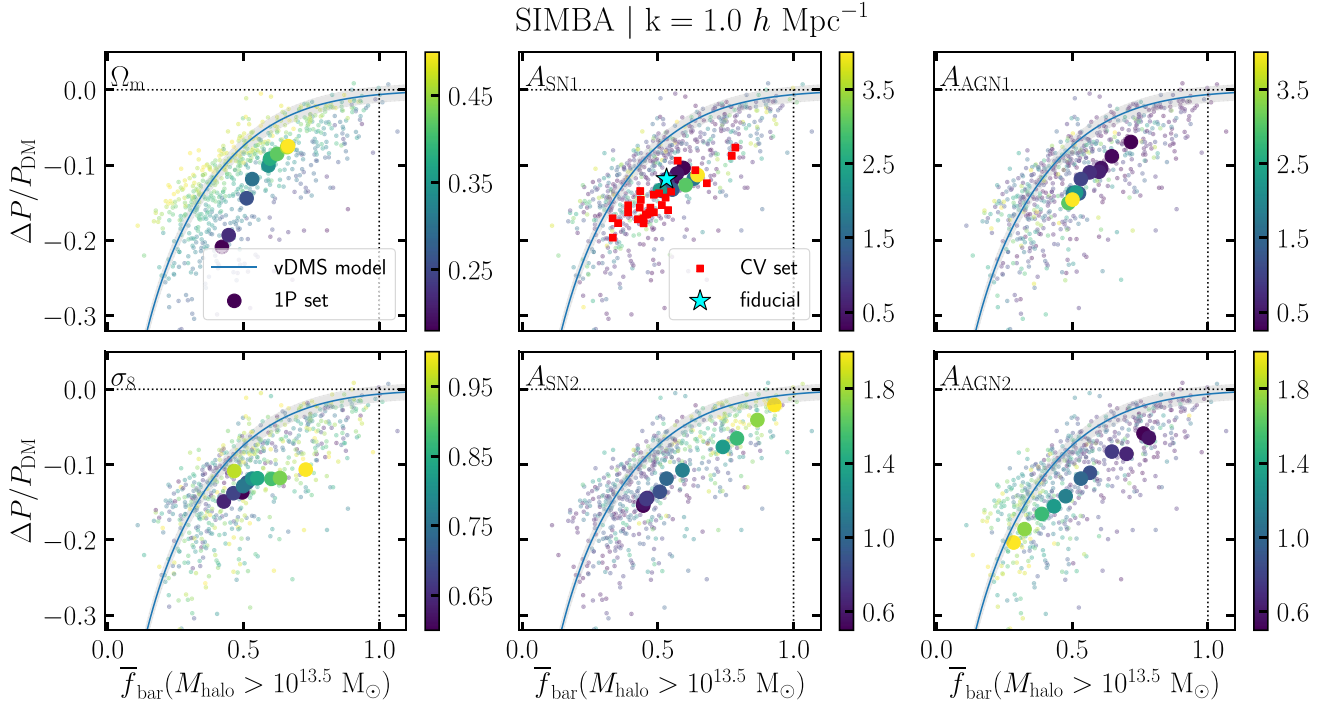


Figure 3. Suppression of the matter power spectrum, $\Delta P/P_{\text{DM}}$, as a function of mean baryon fraction of high-mass haloes, $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$. The blue line is the fitting function of vDMS for a halo definition of 200c (200 times the critical density of the Universe), with the grey-shaded region indicating 1 per cent variation in $\Delta P/P_{\text{DM}}$. Each panel shows the 1P set overlaid on the LH set, colour coded by the value of each of the six parameters. We remind the reader that all six parameters are varied simultaneously in the LH set, while only one parameter is varied in the 1P set. The middle top panel shows additional results from the CV simulation set, where all six parameters are constant and only the initial conditions are varied. The fiducial realization is indicated by the cyan star. We find that $\Delta P/P_{\text{DM}}$ increases for higher values of \bar{f}_{bar} , meaning that there is less suppression of the matter power spectrum in simulations where feedback is less effective at removing gas from haloes. Data points fall generally along the vDMS model, with large scatter owing to broad parameter variations and CV.

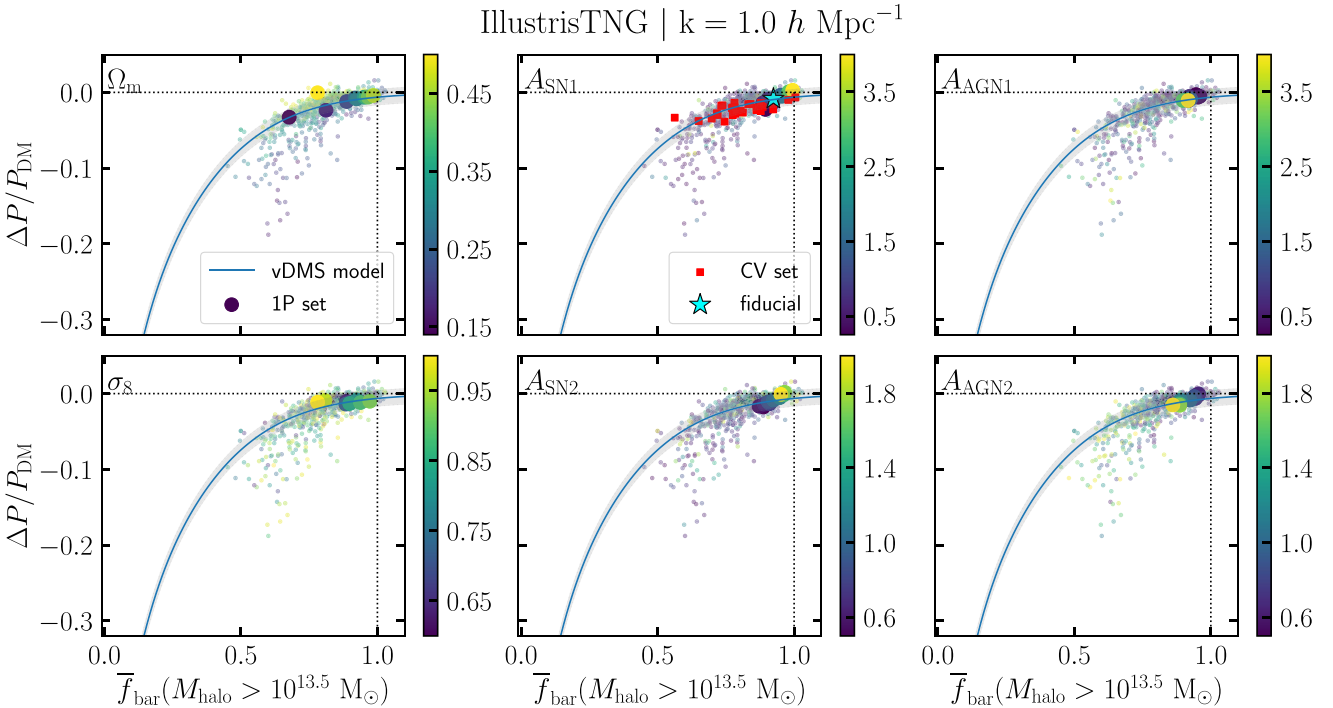


Figure 4. Same as Fig. 3 but for the TNG LH and 1P simulation sets. The TNG sets produce a smaller range of variation in $\Delta P/P_{\text{DM}}$ and \bar{f}_{bar} compared to the SIMBA sets and are in closer agreement with the vDMS model than SIMBA. The location of our TNG fiducial realization (cyan star) along the vDMS model is consistent with the location of the original IllustrisTNG simulations in the vDMS study.

with a clear trend in the 1P set showing weaker suppression of the power spectrum and higher \bar{f}_{bar} with the increase of galactic wind speed due to stellar feedback. Decreasing the strength of stellar feedback parameters (indicated by dark purple points) tends to yield more negative $\Delta P/P_{\text{DM}}$ values and correspondingly lower \bar{f}_{bar} (i.e. stronger impact). This displaces realizations roughly along the vDMS relation but with increasing scatter.

(iii) *AGN feedback parameters*: The right two panels of Figs 3 and 4 show the impact of AGN feedback parameter variations in the vDMS relation. In this case, the large range of both $\Delta P/P_{\text{DM}}$ and \bar{f}_{bar} values in SIMBA allows for a clear depiction of the stronger dependence in AGN parameters along the vDMS relation. The SIMBA 1P set, in particular, shows that stronger AGN feedback leads to both lower values of $\Delta P/P_{\text{DM}}$ and lower \bar{f}_{bar} , indicating more efficient evacuation of gas from haloes and stronger suppression of total matter clustering. The trend is also present in TNG, more so in $A_{\text{AGN}2}$ than in $A_{\text{AGN}1}$ (as expected from Figs 1 and 2), albeit less clear due to the tight assembling of the TNG data points along a smaller range of the vDMS model.

(iv) *Cosmic variance*: In order to examine how CV affects the predicted variation of $\Delta P/P_{\text{DM}}$ as a function of \bar{f}_{bar} , the top middle panel of Figs 3 and 4 overlay the results from the CAMELS CV sets corresponding to 27 realizations of the fiducial TNG and SIMBA models using different initial conditions (red triangles). For TNG, the CV set yields roughly similar range in average baryon fraction of massive haloes as the entire LH set, indicating that stochastic variations owing to the small CAMELS volumes and correspondingly low number of massive haloes play an important role. None the less, the TNG CV simulations roughly follow the vDMS relation. The SIMBA CV set also yields a wide range of $\Delta P/P_{\text{DM}}$ and \bar{f}_{bar} values, but in this case suggesting a systematic offset relative to the vDMS model.

We can quantify the impact of CV on the predicted suppression of matter clustering as the root mean square variation in $\Delta P/P_{\text{DM}}$ relative to the mean:

$$\delta_{\text{cv}} \equiv \frac{\sigma_{\text{cv}}}{|\overline{p_{\text{cv}}}|}, \quad (7)$$

with $p_{\text{cv}} \equiv \Delta P/P_{\text{DM}}$ for the CV set and

$$\sigma_{\text{cv}}^2 = \frac{1}{n} \sum_{i=1}^n (p_{\text{cv}}^i - \overline{p_{\text{cv}}})^2, \quad (8)$$

where $n = 27$ realizations and $\overline{p_{\text{cv}}}$ represents the average of p_{cv} over the CV set. Evaluating equation (7) for $k = 1.0 h \text{ Mpc}^{-1}$ gives $\delta_{\text{cv}} = 0.192$ for SIMBA, and $\delta_{\text{cv}} = 0.357$ for TNG, indicating that there is considerable variation due to CV alone.

The considerable spread of CAMELS predictions relative to the vDMS model shown in this section provides motivation for the machine-learning experiments described in Section 2.4. Given the larger data set in CAMELS with broader variations in feedback and cosmology compared to previous libraries of power spectra, it is possible that the vDMS model relating halo baryon fraction and suppression of matter clustering is not general enough to include every plausible feedback model. For example, we later examine the original SIMBA model against the vDMS relation and find that SIMBA does not fall within 1 per cent of the vDMS fit, as shown in Fig. 12. However, it is also possible that having smaller volumes which are significantly affected by CV as compared to the data set in vDMS, along with the lack of haloes of mass $\sim 10^{14} M_{\odot}$, may explain the disagreement between our results and the vDMS model. These results motivate us to explore the relation between $\Delta P/P_{\text{DM}}$ and halo

baryon fraction with a machine-learning approach, where we can extract information from a broader halo mass range to improve the accuracy of predictions for the impact of baryonic physics on the total matter power spectrum.

4 ESTIMATING THE IMPACT OF FEEDBACK ON MATTER CLUSTERING WITH MACHINE LEARNING

A major goal of this work is to show that machine learning can be used to extract information from the full range of halo masses in order to estimate the suppression of the matter power spectrum by baryonic processes all the way to the non-linear regime. In this section, we discuss the results of training an RF regressor to estimate the impact of feedback on the clustering of matter using the LH simulation sets in CAMELS, which vary simultaneously cosmological and feedback parameters (Section 2.1). The general setup of our experiments is described in Section 2.4.

4.1 Extracting information across the halo mass range with RF regression

Fig. 5 shows the results from training an RF regressor on different input features to estimate the suppression of power $\Delta P/P_{\text{DM}}$ at $k = 1.0 h \text{ Mpc}^{-1}$. We begin by training an RF with $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ as the only training feature, in analogy with the information used by the vDMS fitting function. Haloes of mass $\sim 10^{13.5} M_{\odot}$ are only available for ~ 700 out of 1000 LH realizations for each of TNG and SIMBA, limiting the size of the training set. In this first experiment, the RF is only able to predict ~ 60 and ~ 50 per cent of the variation of $\Delta P/P_{\text{DM}}$ in TNG and SIMBA, respectively, with the predicted versus true values of $\Delta P/P_{\text{DM}}$ shown by the green data points in the left panels of Fig. 5. Next, we add the number of high-mass haloes N_{halo} corresponding to the measured $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ as an additional input feature, with results indicated by the orange data points. In this case, the RF predictions improved by ~ 5 per cent in TNG and ~ 20 per cent for SIMBA. We then incorporate information from haloes across the full mass range by introducing the baryon fraction $\bar{f}_{\text{bar}}(M_{\text{halo}}^i)$ and the corresponding number of haloes N_{halo}^i within each halo mass bin (see Section 2.4), with results shown by the blue data points. In this case, we can use the full LH sets of CAMELS for training and testing since we are not limited by the availability of high-mass haloes. With these additional features using information from a range of halo masses, the RF predicted ~ 75 per cent of the variation in $\Delta P/P_{\text{DM}}$ at $k = 1.0 h \text{ Mpc}^{-1}$ in TNG and ~ 82 per cent SIMBA, significantly improving upon the original results.

As stated in Section 2.4, one advantage of the RF is that it provides some level of interpretability by means of the ‘feature importance’ attribute. The right panels of Fig. 5 display the relative importance assigned to each feature by the trained RF on the test set for TNG (top panel) and SIMBA (bottom panel). We stress that this ranking is based on the frequency with which the features are used by each tree in order to predict the target variable, and it is possible for more than one feature to hold similar, correlated information but be ranked differently by the RF. Interestingly, the RF ranked \bar{f}_{bar} in haloes with mass $M_{\text{halo}} = 10^{10.5} - 10^{11.0} M_{\odot}$ as the most important feature to predict the suppression of power $\Delta P/P_{\text{DM}}$ at $k = 1.0 h \text{ Mpc}^{-1}$ in TNG. Furthermore, we see that the RF ranked several features across the mass ranges as important predictors of $\Delta P/P_{\text{DM}}$ in SIMBA. These feature importance results reveal that the RF was able to extract valuable information across a range of halo masses.

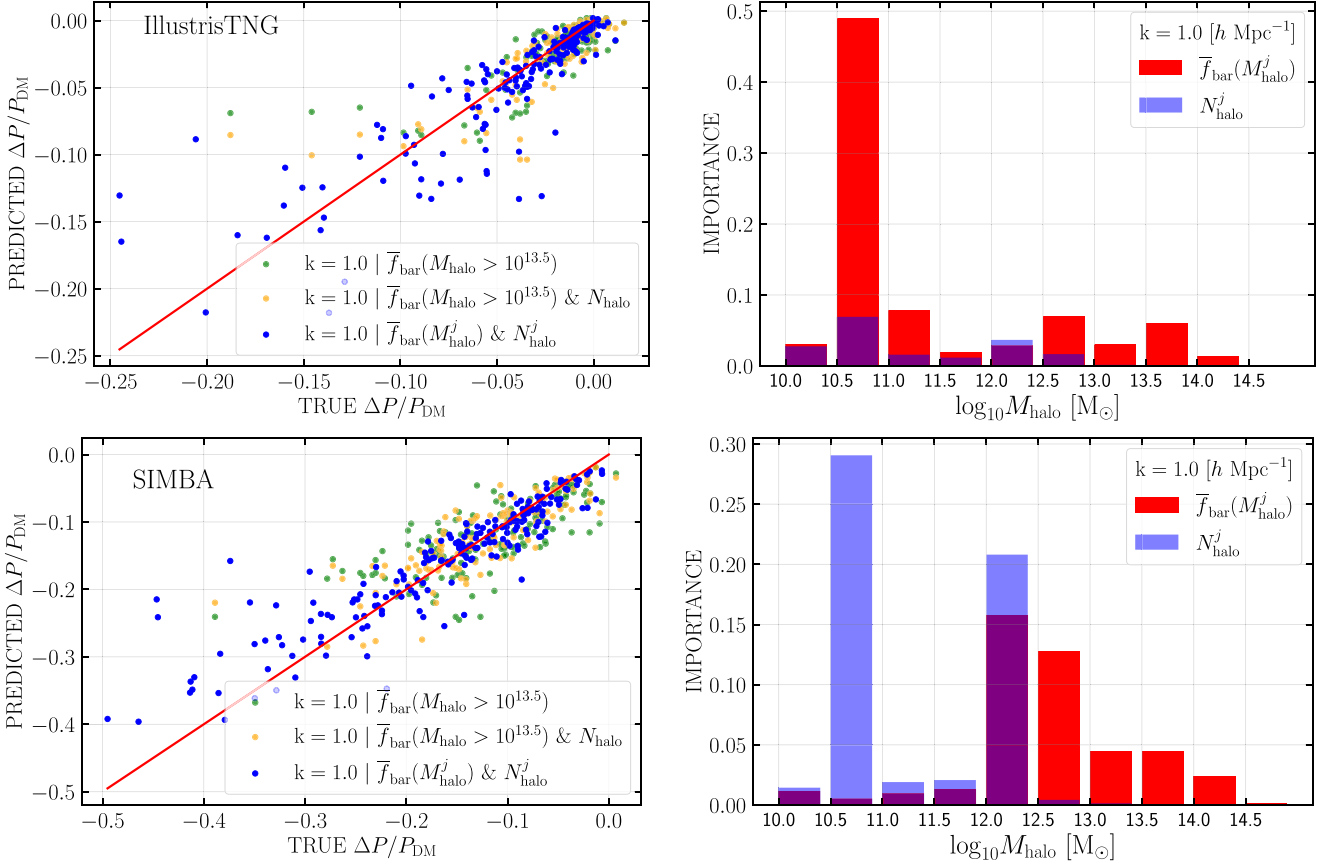


Figure 5. Results from different RF experiments. We create an 80/20 train/test split of the LH data sets to predict $\Delta P/P_{\text{DM}}$ at $k = 1.0$ for TNG (top panel) and SIMBA (bottom panel). The left panels show the predicted target values compared to the true target values as given by the test sets in CAMELS, where the red line indicates a perfect one-to-one relation. The green data points correspond to predictions by an RF trained only on the baryon fraction of massive haloes, $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$, the orange data points show results for an RF trained on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ as well as the number N_{halo} of massive haloes, while results from training an RF on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j for halo mass bins j spanning the full mass range are shown in blue. The right panels show the feature importances corresponding to the blue data points in the left panel (training on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j), indicating the relative rank ordering of importance (from 0.0 to 1.0) given to each training feature by the RF. Predictions improve by providing training data across the full range of halo masses.

4.2 RF predictions in the highly non-linear regime

In the previous subsection, we have established that training the RF on features from a range of halo masses improves the predictions for $\Delta P/P_{\text{DM}}$ in the linear regime compared to using $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ as a training feature alone. We now exploit the same methodology to extend our predictions into the highly non-linear regime by repeating our experiments at a range of scales, predicting $\Delta P/P_{\text{DM}}$ at $k = [1, 5, 10, 20] h \text{ Mpc}^{-1}$. Fig. 6 provides a summary of performance scores for these experiments using the LH simulation sets for TNG (squares) and SIMBA (circles). We also performed additional experiments using the two LH data sets combined, which we refer to as ‘TNG + SIMBA’ (diamonds). Descriptions of the scoring metrics can be found in Section 2.4. The top panels in Fig. 6 show the R^2 scores and the bottom two panels show the RMSE scores normalized by the IQR. We normalize the RMSE by the IQR in order to account for the variation in the range of $\Delta P/P_{\text{DM}}$, which depends on the k value; the range of $\Delta P/P_{\text{DM}}$ increases as we move to non-linear regimes. In addition to presenting results at a range of scales for each training set, Fig. 6 compares the predictions based on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ alone (right panels) versus providing the baryon fraction $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ in different halo mass bins (left panels). Results based on baryon fractions alone are shown in orange while results

that also incorporate the corresponding number of haloes N_{halo} are shown in blue.

Training on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j improved the prediction of $\Delta P/P_{\text{DM}}$ for both TNG and SIMBA on all scales $k = 1.0\text{--}20 h \text{ Mpc}^{-1}$ as measured by the R^2 and RMSE/IQR scores (Fig. 6). For TNG, we achieved the highest R^2 score of all experiments at $k = 5 h \text{ Mpc}^{-1}$, with $R^2 = 0.923$ (consistent with the lowest RMSE/IQR score). In other words, the RF was able to account for approximately 92 percent of the variation in the suppression of the matter power spectrum due to feedback using $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j as training features. This represents ~ 40 per cent improvement over training on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ at $k = 1.0 h \text{ Mpc}^{-1}$ and $\gtrsim 20$ per cent improvement over training on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j at $k = 1.0 h \text{ Mpc}^{-1}$. The trained RF also shows very good performance down to smaller scales, with $R^2 = 0.85\text{--}0.9$ at $k = 10\text{--}20 h \text{ Mpc}^{-1}$ when training simultaneously on the baryon fraction and number of haloes in different mass bins. Similar results are obtained for SIMBA, also performing better at $k = 5\text{--}20 h \text{ Mpc}^{-1}$ compared to larger scales, with R^2 scores slightly lower than TNG: $R^2 \approx 0.8\text{--}0.85$ at $k = 5\text{--}20 h \text{ Mpc}^{-1}$. Remarkably, while the vDMS model can only predict $\Delta P/P_{\text{DM}}$ for a given average baryon fraction of massive haloes on large scales $k < 1 h \text{ Mpc}^{-1}$, our RF regressor performs better on scales where the impact of feedback on the matter power spectrum

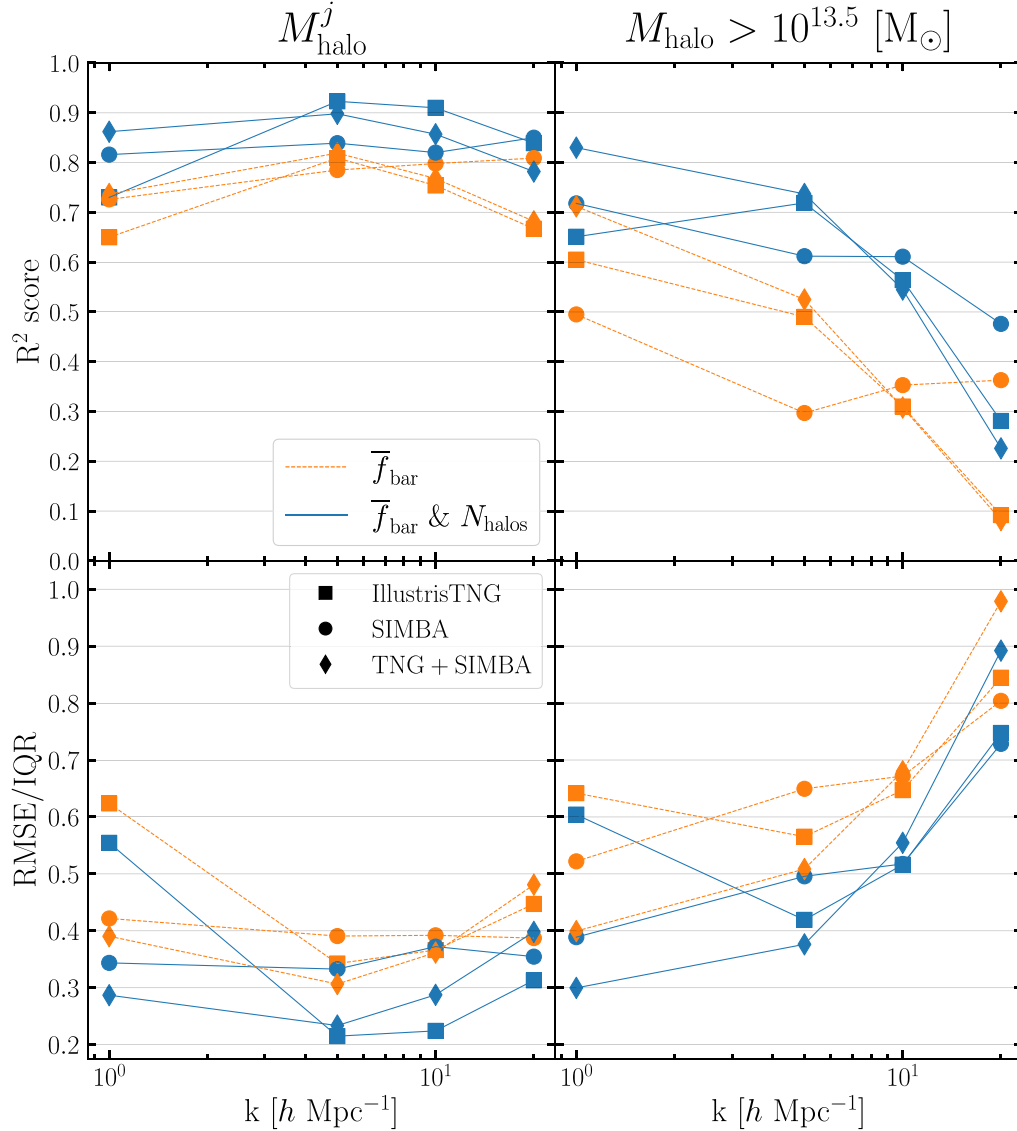


Figure 6. Performance scores for all RF experiments predicting $\Delta P/P_{\text{DM}}$ at four different scales, $k = [1, 5, 10, 20] h \text{ Mpc}^{-1}$. The orange dashed lines connect data points showing scores for an RF trained on average baryon fractions \bar{f}_{bar} , while the blue solid lines connect those where both \bar{f}_{bar} and the number of haloes N_{halo} are used as training features. We show results for three different data sets: the TNG LH set (squares), the SIMBA LH set (circles), and the two LH sets combined, labelled ‘TNG + SIMBA’ (diamonds). Left panels correspond to training on features from a range of halo masses, $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$, while the right panels correspond to training on high-mass haloes only, $M_{\text{halo}} > 10^{13.5} M_{\odot}$, as described in Section 2.4. We report R^2 scores (top panel) and RMSE scores normalized by the IQR of the respective data set (bottom panel); higher R^2 scores and lower RMSE/IQR scores convey an improved performance. We obtain higher scores when training on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo} for a range of halo masses at all scales compared to training on high-mass haloes only, and the highest scores occur in the highly non-linear regime at $k \sim 5\text{--}10 h \text{ Mpc}^{-1}$, meaning that important information can be extracted from a range of halo masses in the non-linear regime.

becomes the highest. The strongest suppression of power occurs at roughly $k = 10 h \text{ Mpc}^{-1}$ in SIMBA and $k = 20 h \text{ Mpc}^{-1}$ in TNG for their fiducial models, and the RF is able to account for $\sim 80\text{--}85$ per cent of the $\Delta P/P_{\text{DM}}$ variation on these scales.

Fig. 7 illustrates in more detail the predicted results versus true values of $\Delta P/P_{\text{DM}}$ when training an RF regressor on different input features at $k = 5 h \text{ Mpc}^{-1}$ for the TNG (top panel) and SIMBA (bottom panel) LH sets. The highest R^2 score was obtained at $k = 5 h \text{ Mpc}^{-1}$ for TNG using $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j as training features, which corresponds to the tighter distribution of blue data points along the one-to-one line of perfect prediction in the top left panel, with reduced scatter compared to the prediction based on high-mass haloes

alone (orange data points). We find similar trends for SIMBA, with an apparent increase in scatter relative to TNG as expected from the lower R^2 scores. Interestingly, the feature importance analysis (right panels) indicates that the number of haloes in the low-mass range $M_{\text{halo}} = 10^{10.5}\text{--}10^{11} M_{\odot}$ is contributing significantly to improve the $\Delta P/P_{\text{DM}}$ predictions.

4.3 Interpretation of feature importances

Comparing the feature importances shown in Figs 5 and 7 for $k = 1.0$ and $5 h \text{ Mpc}^{-1}$, respectively, it appears that the most informative input features vary with scale. We further investigate the physical

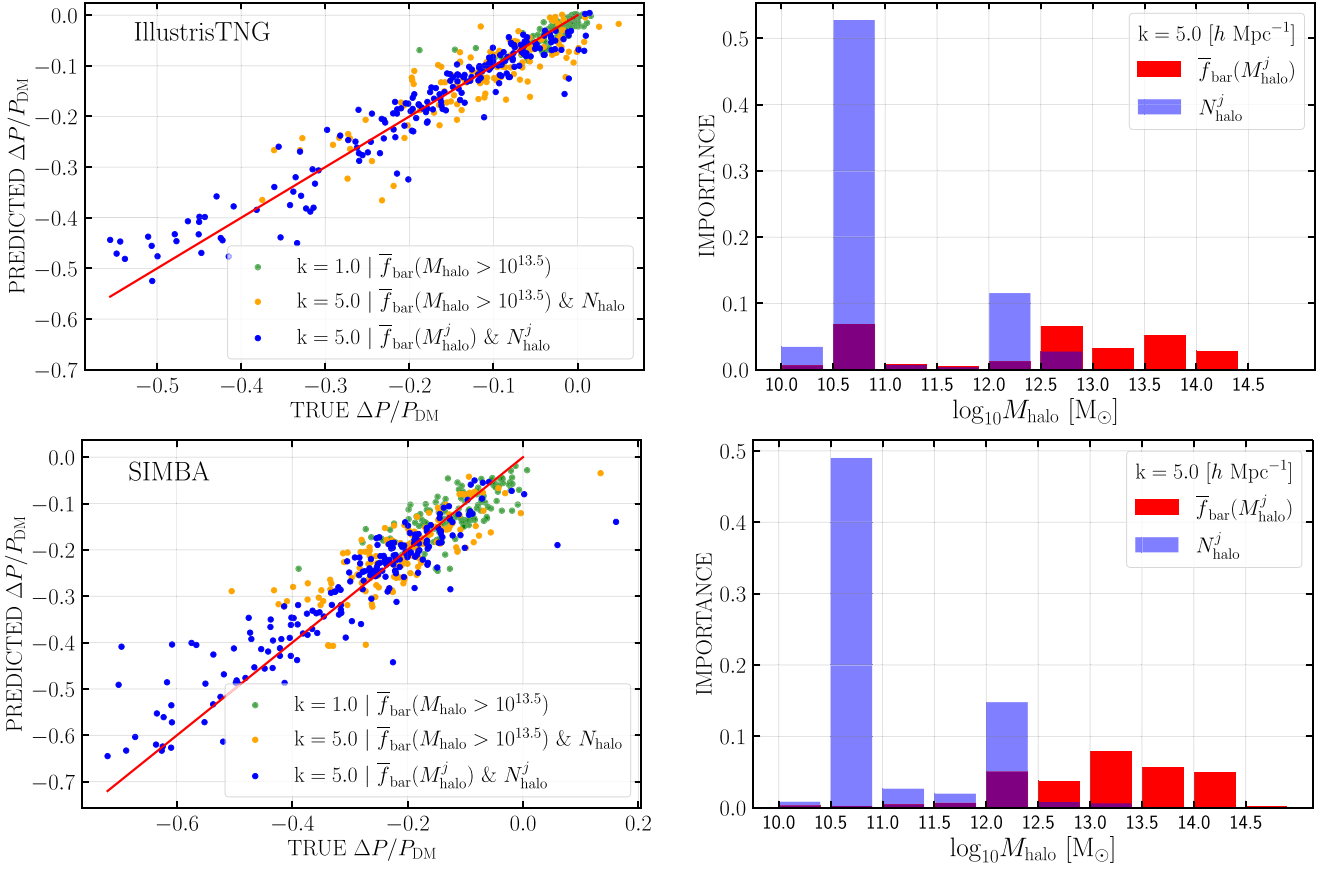


Figure 7. Similar to Fig. 5 but for results at $k = 5.0 h \text{ Mpc}^{-1}$. The left panels show the predicted versus true $\Delta P/P_{\text{DM}}$ values for RF regressors trained on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ and N_{halo} (orange) and trained on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j (blue), both at $k = 5 h \text{ Mpc}^{-1}$. For comparison, we also show the results using only $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ at $k = 0.5 h \text{ Mpc}^{-1}$ as in Fig. 5 (green). The right panels show the feature importances when training on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j . Going to non-linear regimes the range of the target value $\Delta P/P_{\text{DM}}$ increases, enabling more accurate predictions. The highest ranked feature at $k = 5.0 h \text{ Mpc}^{-1}$ is N_{halo}^j at $M_{\text{halo}} = 10^{10.5} - 10^{11} M_{\odot}$, meaning that the number of low-mass haloes is highly informative for estimating the power at 1-Mpc length scales.

properties that inform the prediction of $\Delta P/P_{\text{DM}}$ by the RF in Fig. 8, where we provide a summary of the relative importance of the $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ features (diamonds) and N_{halo}^j features (squares) corresponding to different halo mass bins when predicting $\Delta P/P_{\text{DM}}$ at a given scale k . The ranking of features for each k -value is indicated by the colour scale, which shows the log of the fractional importance assigned by the RF. The most important ranked feature at each k -value is further outlined in red. While the RF appears to be learning from the baryon fractions and abundances of haloes across the halo mass range, there are some interesting trends that are worth noting. The least informative features for both TNG and SIMBA correspond to the most massive haloes ($M_{\text{halo}} \sim 10^{14} M_{\odot}$), which is in contrast to earlier work identifying the baryon fraction in groups and clusters as a primary predictor of $\Delta P/P_{\text{DM}}$ on scales $k < 1 h \text{ Mpc}^{-1}$ (vDMS). Unsurprisingly, the small simulated volumes in CAMELS contain a small number of massive haloes, which are thus not optimal as predictors of $\Delta P/P_{\text{DM}}$. In contrast, the RF assigns significant importance to $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j in low- to intermediate-mass haloes in the range $M_{\text{halo}} \sim 10^{10.5} - 10^{12.5} M_{\odot}$ for all k -values analysed.

Interestingly, the number of haloes N_{halo}^j in the mass range $M_{\text{halo}} \sim 10^{10.5} - 10^{11} M_{\odot}$ is among the top features identified by the RF across different scales (see also Fig. 7). We explore further the significance of this feature in Fig. 9, where we show $\Delta P/P_{\text{DM}}$ at $k = 5 h \text{ Mpc}^{-1}$ as a function of N_{halo}^j in this halo mass range for

the LH sets of TNG (left panel) and SIMBA (right panel). We find that there is a clear correlation between $\Delta P/P_{\text{DM}}$ and the number of low-mass haloes for both galaxy formation models, as expected given that N_{halo}^j is identified by the RF as one of the most predictive features. Physically, a plausible explanation for this correlation is that N_{halo}^j for low-mass haloes is a strong tracer of Ω_m , as indicated by the colour scale, and Ω_m itself is one of the main parameters driving large variations in $\Delta P/P_{\text{DM}}$ at all k -values in CAMELS (see Figs 1, 4, and 3). This is consistent with Pandey et al. (2023), which find a simple model is also able to capture information about $\Delta P/P_{\text{DM}}$ given \bar{f}_{bar} in low-mass haloes and Ω_m . While increasing the value of Ω_m increases systematically the number of haloes at all masses (e.g. Villaescusa-Navarro et al. 2021c), we note that the correlation between $\Delta P/P_{\text{DM}}$ and N_{halo}^j worsens for higher halo mass bins, which are thus often assigned lower feature importance by the RF. This can be explained by the number of haloes in higher mass bins being more sensitive to CV and therefore not as good predictors of cosmology for small simulated volumes. On the other hand, the lowest halo mass bin considered here, $M_{\text{halo}} < 10^{10.5} M_{\odot}$, becomes unresolved in terms of the minimum number of dark matter particles per halo for the higher Ω_m values in CAMELS, as depicted in fig. 8 of Villaescusa-Navarro et al. (2021c). The haloes in the lowest mass bin in this study have less than 200 dark matter particles. We therefore conclude that N_{halo}^j in the halo mass bin $M_{\text{halo}} \sim 10^{10.5} - 10^{11} M_{\odot}$ is

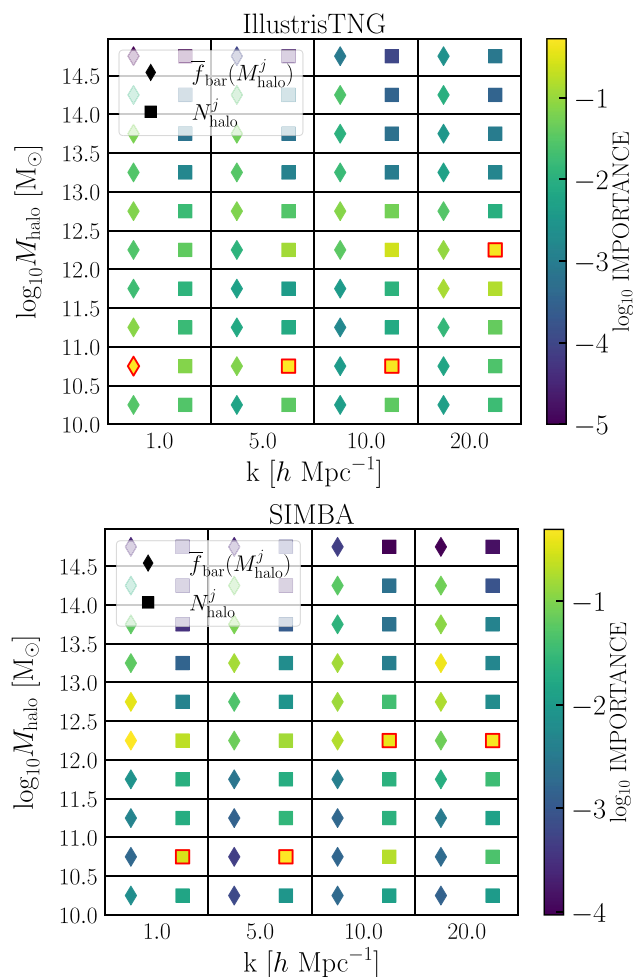


Figure 8. Feature importance summary for RF models trained on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo} to predict $\Delta P/P_{\text{DM}}$ at $k = [0.5, 1, 5, 10, 20] h \text{ Mpc}^{-1}$ for the TNG (top panel) and SIMBA (bottom panel) models. For each halo mass bin and k -value, the relative importance of the corresponding $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ (diamonds) and N_{halo} (squares) features is indicated by the colour scale. The highest ranked feature for each k -value is outlined in red. There is no clear trend of importance in features across scales, suggesting feature importance results are specific to each scale.

the least sensitive to CV and it is a strong predictor of $\Delta P/P_{\text{DM}}$ in part because N_{halo}^j informs the RF about the variation of Ω_m in a large number of well-resolved haloes.

In order to further examine the significance of tracers of Ω_m in low-mass haloes as predictors of $\Delta P/P_{\text{DM}}$ we look specifically at the dependance of the $\Delta P/P_{\text{DM}} - \bar{f}_{\text{bar}}$ relation on Ω_m in the halo mass range $M_{\text{halo}}^j = 10^{10.5} - 10^{11} M_{\odot}$. Results are shown in Fig. 10. Again, the small dots represent the LH sets while the large circles are 1P sets. We see an inverted trend compared to the vDMS relation which holds for high-mass haloes and there is still a clear dependence on Ω_m . These behaviours (increased suppression with higher \bar{f}_{bar} and lower Ω_m) are consistent with results in the Ω_m panels of both Fig. 1, where we see stronger suppression of the matter power spectrum with higher values of Ω_m , and of Fig. 2, where we see that \bar{f}_{bar} increases as Ω_m decreases in low-mass haloes. We again remind the reader that our computed \bar{f}_{bar} are normalized by Ω_b / Ω_m . Therefore, lowering the values of Ω_m at a fixed Ω_b results in a non-trivial effect on the suppression of power.

4.4 Marginalizing over galaxy formation physics

A key advantage of CAMELS over more standard cosmological simulations performed with a single fiducial galaxy formation model is the ability to train machine-learning algorithms to learn fundamental properties of galaxies and the Universe while marginalizing over uncertainties in subgrid physics (e.g. Villaescusa-Navarro et al. 2021a, b; Nicola et al. 2022; Villanueva-Domingo et al. 2022; Shao et al. 2022a, b; Perez et al. 2023). When training an RF on the LH simulation set of either TNG or SIMBA to predict $\Delta P/P_{\text{DM}}$ given halo baryon fractions as input features, we are at the same time marginalizing over uncertainties in physical processes represented by the parameter variations introduced in a given galaxy formation model.

However, evaluating the robustness of the ML model to uncertainties in galaxy formation physics should also consider different implementations and not just variations of parameters within a given subgrid physics implementation. We thus perform a more stringent test of robustness by training the RF on the full LH set of one galaxy formation model (either SIMBA or TNG) and then testing on the full LH set of the other model. Fig. 11 shows the predicted versus true values of $\Delta P/P_{\text{DM}}$ at $k = 5 h \text{ Mpc}^{-1}$ when training an RF using $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo} from SIMBA and testing on TNG (top panel) and when training on TNG and testing on SIMBA (bottom panel). The top panel of Fig. 11 shows that the RF trained on SIMBA can explain ~ 80 per cent of the variation of $\Delta P/P_{\text{DM}}$ at $k = 5 h \text{ Mpc}^{-1}$ when tested on TNG, suggesting that the RF has found a relation between halo baryon fractions and suppression of matter clustering which is relatively robust to galaxy formation model implementation. On the other hand, the bottom panel of Fig. 11 shows that the RF trained on TNG is less robust when tested on SIMBA, and can only explain ~ 70 per cent of the variation in $\Delta P/P_{\text{DM}}$ predicted by the SIMBA model. In this case, we can see that the predicted $\Delta P/P_{\text{DM}}$ is clearly biased high (i.e. less negative) when training on TNG and predicting on SIMBA, implying that the inferred suppression of matter clustering is underpredicted given the halo baryon fractions in SIMBA and the connection to $\Delta P/P_{\text{DM}}$ learned from TNG. As expected, we see a bias in the opposite direction when training on SIMBA and predicting on TNG (top panel), overpredicting the suppression of power at $k = 5 h \text{ Mpc}^{-1}$.

As seen in Figs 3 and 4, the overall range of variation in $\Delta P/P_{\text{DM}}$ is significantly larger in the SIMBA LH set compared to the TNG LH set. Given that the RF cannot predict values outside of the range of the training data, this can explain why the RF trained on TNG is less robust relative to galaxy formation implementation and shows a stronger bias when tested on SIMBA. The biased estimation of $\Delta P/P_{\text{DM}}$ can thus be partially attributed to the RF learning the limits of the range of variation in the training data. None the less, our results suggest that the TNG and SIMBA models may predict different $\Delta P/P_{\text{DM}}$ even when implementing parameters that yield similar $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$, implying a non-unique relation between halo baryon fractions and impact on matter clustering.

4.5 Supplementary studies

In order to further investigate the performance of training the RF on halo abundances, we performed several supplementary studies comparing our fiducial models to new models which include cosmological parameters as training features. The results of these experiments are presented in Appendix A, where we conclude that our fiducial model, utilizing $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo} across the full range of halo masses, extracts important cosmological information

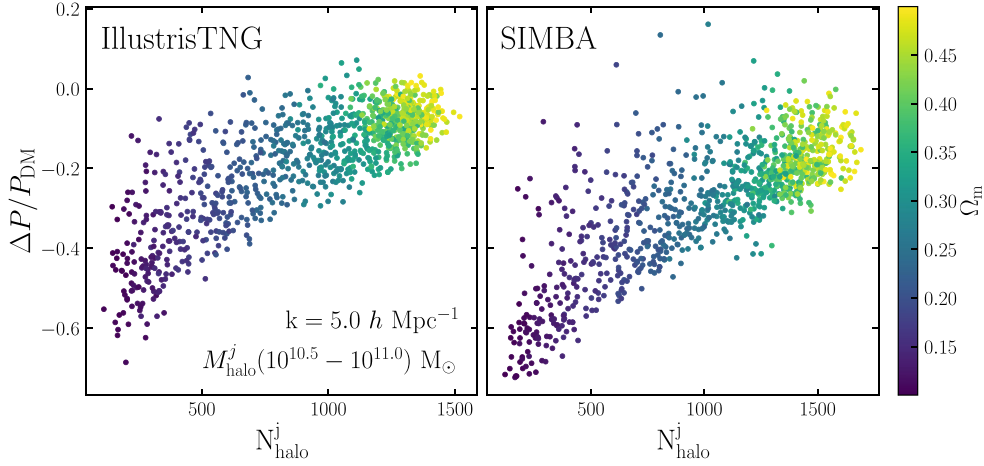


Figure 9. Correlation between power spectrum suppression $\Delta P/P_{\text{DM}}$ and the number N_{halo} of haloes in the mass range $M_{\text{halo}}^j = 10^{10.5}-10^{11} M_{\odot}$ at $k = 5 h \text{ Mpc}^{-1}$ for TNG (left panel) and SIMBA (right panel), colour coded by Ω_m . The number of low-mass haloes N_{halo} (highest ranked feature at $k = 5 h \text{ Mpc}^{-1}$) is a strong tracer of Ω_m in CAMELS and a good predictor of $\Delta P/P_{\text{DM}}$.

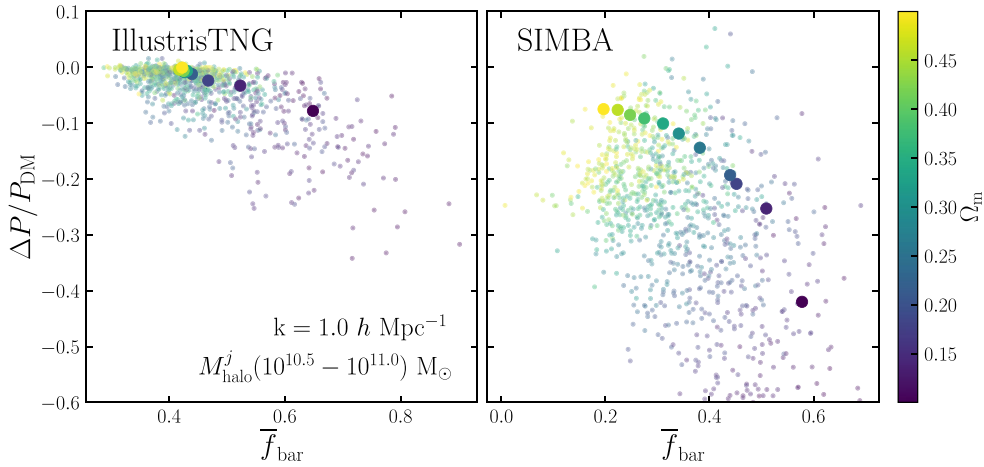


Figure 10. Similar to Figs 3 and 4 but here we look only at the dependence of the $\Delta P/P_{\text{DM}} - \bar{f}_{\text{bar}}$ relation on Ω_m for the $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ in the range $M_{\text{halo}}^j = 10^{10.5}-10^{11}$. Again, the small dots represent the LH sets while the large circles are 1P sets. We see an inverted trend compared to the vDMS relation which holds for high-mass haloes, but there is still a strong dependence on Ω_m .

as it is in close agreement (within 1 per cent) with a model trained on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and Ω_m . However, in the case of high-mass haloes, halo abundance is not a sufficient proxy for cosmological information. We also found that limiting the study to intermediate mass haloes resulted in comparable results to our fiducial model when the RF was also directly provided Ω_m as a training feature in addition to the \bar{f}_{bar} of intermediate mass haloes.

Based on a suite of simulations of substantially larger volumes, vDMS identified a tight relationship between mean baryon fraction in massive haloes and baryonic power spectrum suppression at $k < 1 h \text{ Mpc}^{-1}$. While some degree of correlation between these quantities is expected, as we find here, the extremely small scatter in the observed relationship was remarkable given the variety of galaxy formation models compared by vDMS. Since we observed some hints at a deviation from the vDMS relationship in CAMELS, particularly for the parameter variations based on the SIMBA model, we perform a direct comparison of the original $100 h^{-1} \text{ Mpc}$ SIMBA volume to the vDMS relation in Fig. 12. Here, we show the new SIMBA data points as circles, while the measurements considered in vDMS are shown as triangles. We have verified our pipeline by running it

on IllustrisTNG-300, achieving excellent agreement with the vDMS measurements. As can be seen, SIMBA constitutes a considerable outlier and does not fall within the 1 per cent interval around the vDMS fit. The only other simulation scattering that far is Illustris which, however, does not reproduce the observed baryon fraction. It is known that the feedback prescription in SIMBA is unique in its ability to re-distribute baryons across large scales (Borrow et al. 2020; Gebhardt et al. 2023), which could explain the observed deviation from the vDMS relation. However, SIMBA generally does not do worse in reproducing observational relationships than the other major hydrodynamic simulations. Thus, these results suggest the possibility that the vDMS relation only holds in a subspace of simulations while there exists at least one dimension along which deviations occur.

5 SUMMARY AND DISCUSSION

In this paper, we have investigated how baryonic physics affects the clustering of matter relative to N -body simulations and its relation to the baryonic content of haloes using thousands of cosmological hydrodynamic simulations from the CAMELS project (Villaescusa-

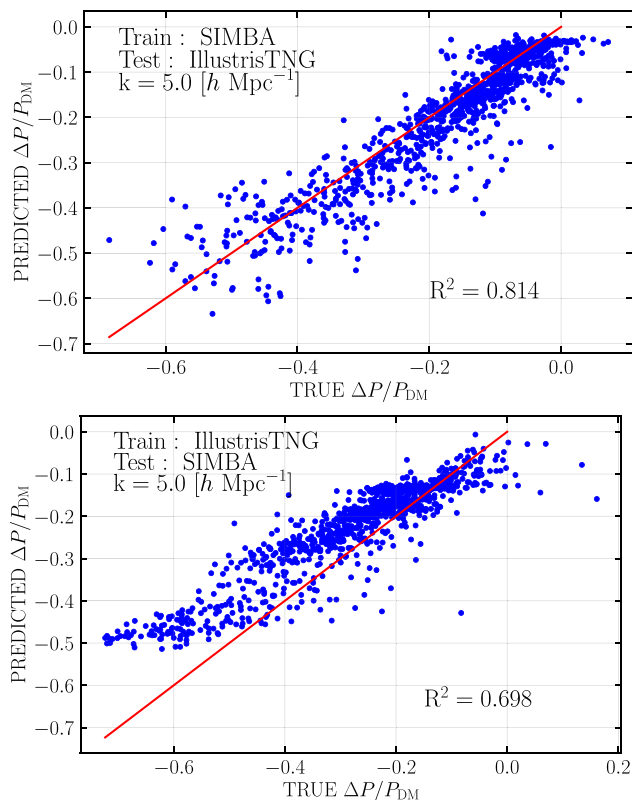


Figure 11. Similar to Fig. 7 but training an RF on the entire LH set of one galaxy formation model and testing on the entire LH set of the other for both TNG and SIMBA. We note higher prediction score when training on SIMBA and testing on TNG ($R^2 = 0.814$) compared to training on TNG and predicting on SIMBA ($R^2 = 0.698$), which can be attributed to SIMBA’s wider range of $\Delta P/P_{\text{DM}}$ values. The RF tends to underpredict $\Delta P/P_{\text{DM}}$ when trained on SIMBA and overpredict $\Delta P/P_{\text{DM}}$ when trained on TNG due to variations in feedback models which suggests that care must be made if applying this model to data where feedback is not precisely known.

Navarro et al. 2021c). In the first part of the paper, we examined how variations of individual cosmological parameters (Ω_m and σ_8) and feedback parameters (controlling the efficiency of large-scale outflows driven by SNe and AGN) impact the total matter power spectrum, the mean halo baryon fraction as a function of halo mass, and, motivated by vDMS, the connection between the suppression of clustering, $\Delta P/P_{\text{DM}}$, and the mean baryon fraction of massive haloes.

The small simulated volumes in CAMELS complicate a direct comparison between our results and vDMS, where in addition to CV affecting $\Delta P/P_{\text{DM}}$ we also lack haloes massive enough to evaluate the mean baryon fraction under the same conditions. Therefore, in the second part of the paper, we have presented a set of machine-learning experiments as an extension to what was done in vDMS, training an RF regressor on features including the mean baryon fraction and introducing the abundance of haloes across the mass range $10^{10} \leq M_{\text{halo}}/M_{\odot} h^{-1} < 10^{15}$ to predict $\Delta P/P_{\text{DM}}$ from linear to highly non-linear scales ($k = 1.0\text{--}20.0 h \text{ Mpc}^{-1}$). By utilizing halo abundance as a training feature, the RF learns about the CV present in the simulations without explicitly knowing the underlying cosmology.

Throughout the paper, we have made use of the CAMELS simulation suites performed with the TNG (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b) and SIMBA (Davé et al. 2019) models to understand the dependence of results and the robustness of the trained machine-

learning models to changes in the specific galaxy formation physics implementation.

Our main findings can be summarized as follows:

(i) In agreement with previous work (e.g. van Daalen et al. 2011; Chisari et al. 2018, 2019; Villaescusa-Navarro et al. 2021c), we find that baryonic physics can profoundly affect the total matter power spectrum all the way to scales $k < 0.5 h \text{ Mpc}^{-1}$, and the magnitude of this effect is highly dependent on the details of the galaxy formation implementation and variations of cosmological and astrophysical parameters.

(ii) The suppression of power, $|\Delta P/P_{\text{DM}}|$, increases systematically with decreasing Ω_m at fixed Ω_b , with baryons contributing a higher fraction of the total matter content and feedback more efficiently spreading matter over larger scales relative to N -body simulations. Varying σ_8 at fixed galaxy formation physics does not drive systematic variations in $\Delta P/P_{\text{DM}}$ when measured on the small $(25 h^{-1} \text{ Mpc})^3$ volumes simulated in CAMELS.

(iii) Increasing AGN feedback efficiency generally drives higher suppression of matter clustering, in agreement with previous work (e.g. vDMS; Nicola et al. 2022), with the strongest effect seen for high-speed jets in SIMBA which are able to spread a substantial amount of baryons over scales of several Mpc (Borrow et al. 2020; Gebhardt et al. 2023). The qualitative effect of stellar feedback on matter clustering is more dependent on galaxy formation model, which can either suppress or enhance power on different scales depending on the interplay between stellar and AGN feedback. Stronger stellar feedback often results in weaker overall suppression of matter clustering by suppressing black hole growth and therefore the effective efficiency of AGN feedback (van Daalen et al. 2011; Nicola et al. 2022; Gebhardt et al. 2023).

(iv) Halo baryon fractions \bar{f}_{bar} are very sensitive to galaxy formation model, with TNG producing systematically more baryon-rich haloes compared to SIMBA for a broad range of parameter variations. Higher AGN feedback efficiency generally decreases halo baryon fractions, but the extent of the effect and the affected halo mass range depend on model details. Increasing the strength of stellar feedback can either decrease or increase the baryon fraction depending on the non-linear coupling of stellar feedback and black hole growth. Halo baryon fractions are also very sensitive to changes in cosmology. Increasing Ω_m (at fixed Ω_b) or σ_8 systematically decreases the baryon fraction of haloes (normalized by Ω_b/Ω_m), indicating a non-trivial response of feedback to changes in the amount of baryons relative to dark matter and the growth history of haloes.

(v) We find a broad correlation between the amount of suppression of the matter power spectrum $\Delta P/P_{\text{DM}}$ and the baryon fraction of massive haloes \bar{f}_{bar} , indicating that the feedback mechanisms responsible for evacuating gas from massive haloes also dominate the impact of baryonic effects on matter clustering. These results are in broad agreement with vDMS, but the thousands of simulations in CAMELS produce significantly larger scatter in the $\Delta P/P_{\text{DM}} - \bar{f}_{\text{bar}}$ relation. CV alone can significantly affect the matter power spectrum on our $(25 h^{-1} \text{ Mpc})^3$ simulated volumes, but the complex trends seen for the impact of individual cosmological and feedback parameter variations on $\Delta P/P_{\text{DM}}$ and \bar{f}_{bar} suggest that the vDMS model predicting $\Delta P/P_{\text{DM}}$ given only \bar{f}_{bar} for massive haloes is not general enough to include every plausible feedback model.

(vi) Predicting the impact on matter clustering based only on the mean baryon fraction of massive haloes using the vDMS $\Delta P/P_{\text{DM}} - \bar{f}_{\text{bar}}$ relation is not possible given the broad range of galaxy formation models and the impact of CV in CAMELS. However, we have demonstrated that an RF regressor trained on CAMELS is able to

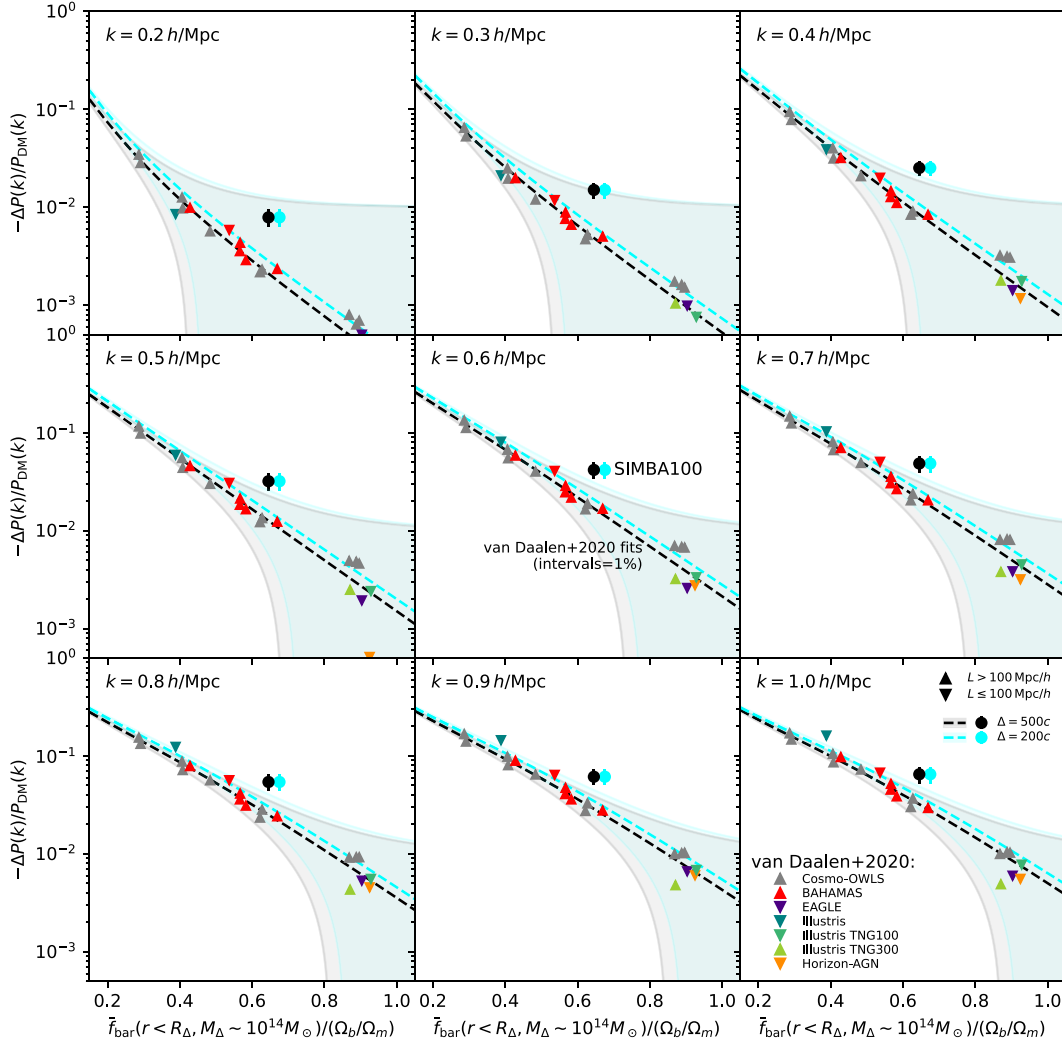


Figure 12. Comparison of the power spectrum suppression in the original SIMBA simulation with the results from vDMS. The SIMBA data points are circles, while the measurements from vDMS are reproduced as triangles. We also show the fitting functions from vDMS as dashed lines as well as 1 per cent range of variations. 200c and 500c mass definitions are shown in different colours. The vertical error bars have been estimated by splitting the $100 h^{-1} \text{Mpc}$ SIMBA simulation into eight subvolumes. Statistical error bars on the horizontal axis are small ($\sim 10^{-3}$ as estimated using jack-knife), but there are somewhat larger systematic errors from the halo finding ($\sim 10^{-2}$, as estimated by running the analysis with FOF and Rockstar). SIMBA appears to deviate from the best-fitting relation of vDMS.

extract information from haloes across the full mass range $10^{10} \leq M_{\text{halo}}/M_{\odot} < 10^{15}$ to estimate the suppression of the matter power spectrum on scales $k = 1.0\text{--}20 h \text{Mpc}^{-1}$. We are thus not only extracting information from low-mass haloes but also predicting $\Delta P/P_{\text{DM}}$ in the highly non-linear regime, significantly extending the range of scales $k < 1 h \text{Mpc}^{-1}$ where the vDMS model can be applied.

(vii) Using the mean halo baryon fraction and abundance in different halo mass bins as input features, the RF regressor was able to account for $\sim 80\text{--}85$ per cent of the $\Delta P/P_{\text{DM}}$ variation occurring on scales $k = 10\text{--}20 h \text{Mpc}^{-1}$ where the impact of feedback on the matter power spectrum becomes the highest. At $k = 5.0 h \text{Mpc}^{-1}$, our best model was able to explain ~ 92 per cent of the variance in the suppression of power due to feedback when training on the TNG model. However, the same model can only explain ~ 70 per cent of the variation in $\Delta P/P_{\text{DM}}$ and tends to underpredict the suppression of matter clustering when tested on the SIMBA simulations, indicating

that the RF is only moderately robust relative to changes in the underlying galaxy formation implementation. Training on SIMBA increases the robustness of the model owing to its larger range of variation in $\Delta P/P_{\text{DM}}$ compared to TNG, but in this case the RF tends to overpredict $|\Delta P/P_{\text{DM}}|$ when tested on TNG. These results suggest that the lack of a universal relation between halo baryon fractions and impact on matter clustering and emphasize the need to construct models that are robust against assumptions in baryonic physics (e.g. Villaescusa-Navarro et al. 2021b; Nicola et al. 2022; Shao et al. 2022a).

(viii) The original SIMBA volume constitutes a considerable outlier to the vDMS relation and does not fall within the 1 per cent interval around the vDMS fit.

(ix) Our fiducial model, utilizing $\bar{f}_{\text{bar}}(M_{\text{halo}}^i)$ and N_{halo}^i across the full range of halo masses, extracts important cosmological information as it is in close agreement (within 1 per cent) with a model trained on $\bar{f}_{\text{bar}}(M_{\text{halo}}^i)$ and Ω_m (Appendix A).

A unique advantage of CAMELS relative to previous work is that it performs simulations for different baryonic physics implementations and a broad range of cosmological and feedback parameter variations, providing a data set sufficiently large to train machine-learning algorithms for a variety of applications (Villaescusa-Navarro et al. 2021c). However, an important limitation of CAMELS is the small volume of each simulation realization, $L_{\text{box}} = 25 h^{-1} \text{Mpc}$, with important implications for this work. Given the small box sizes, the matter power spectrum is sensitive to the specific initial conditions in each realization, and the impact of baryonic effects further depends on stochastic processes related to feedback operating on a limited number of massive haloes. As a result, CV represents a challenge to infer the suppression of matter clustering $\Delta P/P_{\text{DM}}$ given only the baryon fraction of massive haloes. Previous works in CAMELS have devised strategies to correct for the noise introduced by CV. When training a neural network on electron density autopower spectra to predict Ω_m , Nicola et al. (2022) constructed a CV parameter based on the distribution of halo masses in each realization, improving the predictions significantly when introduced as an additional training feature. Thiele et al. (2022) used spectral distortion measurements to constrain baryonic feedback and applied a correction factor to the Compton- γ distortion by comparing expected values from a simple halo model evaluated for the halo mass function in each CAMELS simulation compared to that of a standard halo mass function. In our RF experiments, introducing the number of haloes in each mass bin, N_{halo} or N_{halo}^j , as input features (i.e. basically the halo mass function) improves the accuracy of the predictions significantly. The number of low-mass haloes was one of the most predictive features identified by the RF, which can be understood as a strong tracer of Ω_m (Fig. 9). Halo abundance further serves as a feature which provides the RF algorithm information about CV, in agreement to previous works. Furthermore, CV in CAMELS is lower in low-mass haloes than high-mass haloes, which may account for the predictive power of N_{halo}^j in low-mass haloes.

One final result that we found puzzling, was the inversion of the $\Delta P/P_{\text{DM}} - \bar{f}_{\text{bar}}$ relation and its dependence on Ω_m in low-mass haloes. It is unclear why a lower value in Ω_m at a fixed Ω_b would result in higher \bar{f}_{bar} and increased suppression in the total matter spectrum in low-mass haloes. The dependence on $\Delta P/P_{\text{DM}}$ on baryonic feedback in low-mass haloes will need to be further investigated in order to thoroughly explain these results. It will also be interesting to perform the RF training on a fixed, or more narrow range of Ω_m to study how well the RF can really learn the different effects of baryonic feedback, but we leave this to a future study.

ACKNOWLEDGEMENTS

We wish to thank the anonymous referee for insightful comments which aided in improving this work. We also thank Dylan Nelson who likewise provided valuable comments. DA-A 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.

DATA AVAILABILITY

The simulations used in this work are part of the CAMELS public data release (Villaescusa-Navarro et al. 2023b) and are available at <https://camels.readthedocs.io/en/latest/>.

REFERENCES

- Agertz O., Kravtsov A. V., 2016, *ApJ*, 824, 79
- 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. et al., 2021, *ApJ*, 917, 53
- Ayromlou M., Nelson D., Pillepich A., 2023, *MNRAS*, 524, 5391
- Borrow J., Anglés-Alcázar D., Davé R., 2020, *MNRAS*, 491, 6102
- Chisari N. E. et al., 2018, *MNRAS*, 480, 3962
- Chisari N. E. et al., 2019, *Open J. Astrophys.*, 2, 4
- DESI Collaboration, 2016, preprint (arXiv:1611.00036)
- 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
- Dubois Y. et al., 2014, *MNRAS*, 444, 1453
- Eifler T., Krause E., Dodelson S., Zentner A. R., Hearin A. P., Gnedin N. Y., 2015, *MNRAS*, 454, 2451
- Fedeli C., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 028
- Gebhardt M. et al., 2023, preprint (arXiv:2307.11832)
- Genel S. et al., 2014, *MNRAS*, 445, 175
- Hafen Z. et al., 2020, *MNRAS*, 494, 3581
- Hearin A. P., Zentner A. R., Ma Z., 2012, *J. Cosmol. Astropart. Phys.*, 2012, 034
- Hellwing W. A., Schaller M., Frenk C. S., Theuns T., Schaye J., Bower R. G., Crain R. A., 2016, *MNRAS*, 461, L11
- Hopkins P. F., 2015, *MNRAS*, 450, 53
- Hopkins P. F., 2017, preprint (arXiv:1712.01294)
- Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
- Huterer D., Takada M., 2005, *Astropart. Phys.*, 23, 369
- Kitching T. D., Verde L., Heavens A. F., Jimenez R., 2016, *MNRAS*, 459, 971
- Knollmann S. R., Knebe A., 2011, Astrophysics Source Code Library, recordascl:1102.009
- LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
- Laureijs R., 2009, preprint (arXiv:0912.0914)
- Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
- Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, *MNRAS*, 441, 1270
- Li Q., Narayanan D., Davé R., 2019, *MNRAS*, 490, 1425
- McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, *MNRAS*, 465, 2936
- Marinacci F. et al., 2018, *MNRAS*, 480, 5113
- Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *MNRAS*, 454, 1958
- Mitchell P. D., Schaye J., 2022, *MNRAS*, 511, 2600
- Mohammed I., Seljak U., 2014, *MNRAS*, 445, 3382
- Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, *MNRAS*, 454, 2691
- Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
- Nelson D. et al., 2018, *MNRAS*, 475, 624
- Nicola A. et al., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 046
- Pakmor R., Marinacci F., Springel V., 2014, *ApJ*, 783, L20
- Pandey S. et al., 2023, *MNRAS*, 525, 1779
- Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
- 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
- Peters A., Brown M. L., Kay S. T., Barnes D. J., 2018, *MNRAS*, 474, 3173
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Schneider A., Teyssier R., 2015, *J. Cosmol. Astropart. Phys.*, 2015, 049
- Seljak U., 2000, *MNRAS*, 318, 203
- Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, *MNRAS*, 417, 2020
- Semboloni E., Hoekstra H., Schaye J., 2013, *MNRAS*, 434, 148

- Shao H. et al., 2022a, preprint (arXiv:2209.06843)
 Shao H. et al., 2022b, *ApJ*, 927, 85
 Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
 Sorini D., Davé R., Cui W., Appleby S., 2022, *MNRAS*, 516, 883
 Spergel D. et al., 2015, preprint (arXiv:1503.03757)
 Springel V., 2010, *MNRAS*, 401, 791
 Springel V. et al., 2018, *MNRAS*, 475, 676
 Tenneti A., Mandelbaum R., Di Matteo T., Kiessling A., Khandai N., 2015, *MNRAS*, 453, 469
 Thiele L. et al., 2022, *Phys. Rev. D*, 105, 083505
 Tillman M. T. et al., 2023, preprint (arXiv:2307.06360)
 van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, *MNRAS*, 415, 3649
 van Daalen M. P., McCarthy I. G., Schaye J., 2020, *MNRAS*, 491, 2424
 Villaescusa-Navarro F. et al., 2021a, preprint (arXiv:2109.09747)
 Villaescusa-Navarro F. et al., 2021b, preprint (arXiv:2109.10360)
 Villaescusa-Navarro F. et al., 2021c, *ApJ*, 915, 71
 Villaescusa-Navarro F. et al., 2023a, *ApJS*, 265, 14
 Villaescusa-Navarro F. et al., 2023b, *ApJS*, 265, 54
 Villanueva-Domingo P. et al., 2022, *ApJ*, 935, 30
 Vogelsberger M. et al., 2014a, *MNRAS*, 444, 1518
 Vogelsberger M. et al., 2014b, *Nature*, 509, 177
 Weinberger R. et al., 2017, *MNRAS*, 465, 3291
 Weinberger R., Springel V., Pakmor R., 2020, *ApJS*, 248, 32
 Wright R. J., Lagos C. d. P., Power C., Mitchell P. D., 2020, *MNRAS*, 498, 1668
 Zentner A. R., Semboloni E., Dodelson S., Eifler T., Krause E., Hearin A. P., 2013, *Phys. Rev. D*, 87, 043509

APPENDIX A: RESULTS FROM SUPPLEMENTARY STUDIES

In order to further investigate the performance of training the RF on halo abundances, we performed several supplementary studies comparing our fiducial models to new models which include cosmological parameters as training features. The results for these can be seen in Fig. A1. We continue to show the main results from this work in orange and blue, as done in Fig. 6, in order to provide a more clear comparison to the supplementary studies. We find the following:

(i) We introduced the training feature Ω_m as a substitute to halo abundance (N_{halo} or N_{halo}^j). Scores for these results are shown in purple in Fig. A1. In the left panels, scores for training on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and Ω_m (purple) and those for our fiducial model ($\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j , blue) are within ≈ 1 per cent difference. This indicates that the RF was able to extract important cosmological information given the full range of halo masses and halo abundances. However, in the right panel, we see that in the RF model trained on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ and Ω_m resulted in ≈ 20 per cent on average improvement in scores compared to that trained on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ and N_{halo} . This

indicates that in the case of high-mass haloes in CAMELS, halo abundance is not a sufficient proxy for cosmological information.

(ii) We introduced σ_8 as a third training feature (i.e. models trained on mean baryon content, Ω_m and σ_8). There were no improvements to the results from using σ_8 as an additional training feature and results are not shown in Fig. A1 in order to maintain visual clarity.

(iii) We trained the RF on mean baryon content and σ_8 . The performance was poor compared to the models trained on mean baryon content and Ω_m and results are not shown in order to maintain visual clarity.

We further explored the relation between $\Delta P/P_{\text{DM}}$ and mean baryon content in a regime of more intermediate mass haloes. Because we are limited in the number of realizations with haloes of mass $M_{\text{halo}} \geq 10^{13.5} M_{\odot}$, we repeated the additional RF studies described above, this time with \bar{f}_{bar} of $M_{\text{halo}} \in [10^{12}-10^{14}] M_{\odot}$ as a training feature. We also created eight bins of $M_{\text{halo}}^j \in [10^{12}-10^{14}]$. The results from training the RF on baryon content from intermediate mass haloes can be compared to those using the full range of halo masses. The scores from these results are shown in Fig. A1.

(i) We found that training the RF on the mean baryon content of intermediate mass haloes alone (red) resulted in significantly poorer scores than training on either $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ (right panels shown in orange) or on the $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ of the full range of halo masses (left panels shown in orange).

(ii) The \bar{f}_{bar} of intermediate mass haloes and Ω_m model (right panels shown in green) resulted in significantly improved performance of the RF ability to predict $\Delta P/P_{\text{DM}}$ compared to the $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$ and N_{halo} model (right panels shown in blue). Improved scores average ≈ 33 per cent difference. The results for this improved model are comparable to those shown in green in the left panels, which are for a model using Ω_m and $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ for the range of intermediate masses. This suggests that there is no meaningful information to be extracted by binning intermediate halo masses to obtain the mean baryon content per mass bin versus the \bar{f}_{bar} averaged over all intermediate mass haloes. Furthermore, our fiducial model (RF trained on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ and N_{halo}^j from the full range of halo masses, shown in blue in the left panels) is only, on average, 6 per cent better than that trained on \bar{f}_{bar} of intermediate mass haloes and Ω_m , suggesting the latter is promising and worth further investigation as it is a simpler model.

(iii) We further limited the halo mass range of intermediate mass haloes to $M_{\text{halo}} \in [10^{12}-10^{13}] M_{\odot}$ and repeated our experiments. Scores were on average ≈ 10 per cent worse than those discussed in this section and results are not shown in order to maintain visual clarity. However, this further demonstrates that there is important information to be extracted from the baryon content of more massive haloes.

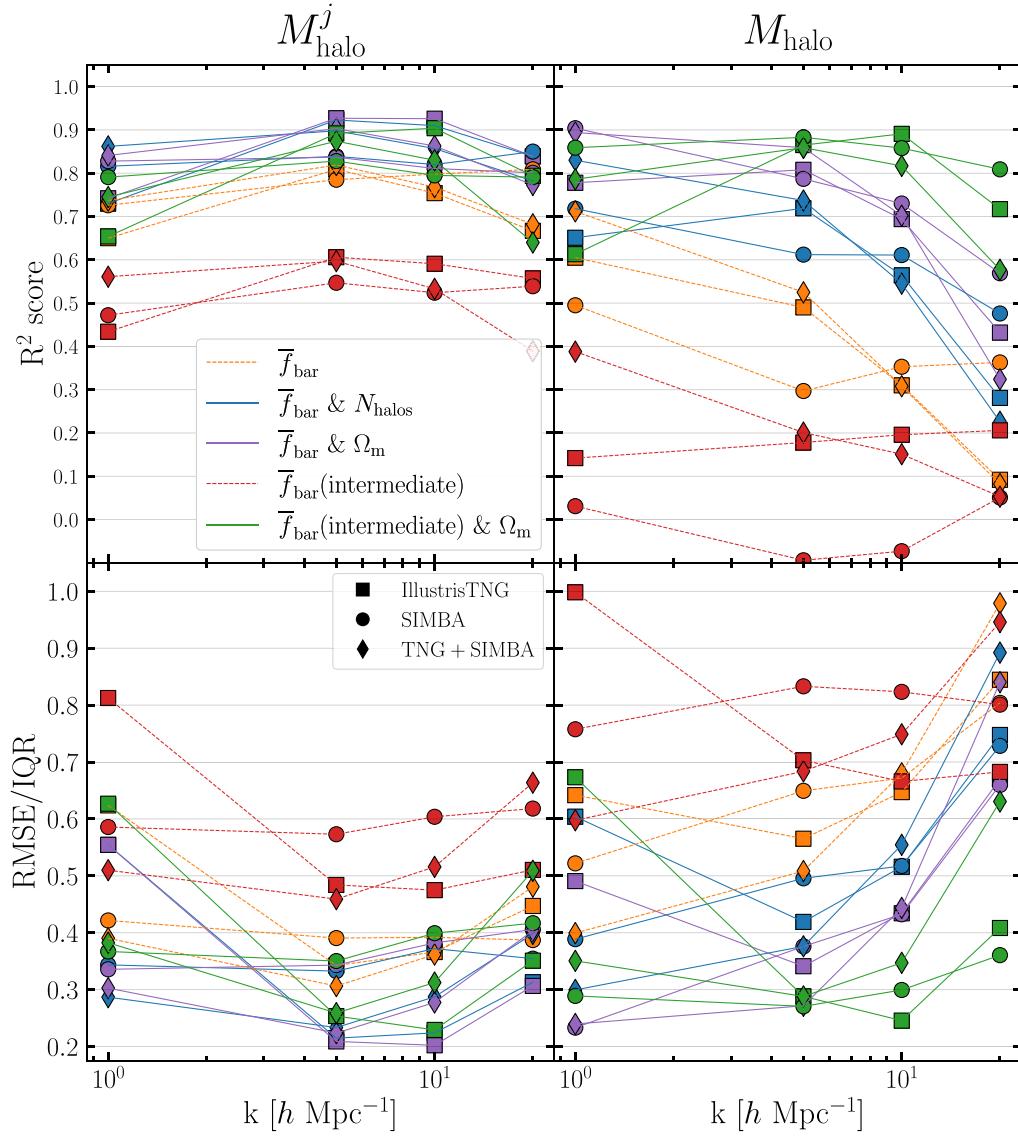


Figure A1. Performance scores for additional RF experiments. Results from the main studies in this paper are still shown in blue and orange as in Fig. 6. Models are trained on mean baryon fraction with some models having additional training features of halo abundance or Ω_m as indicated in the legend. Left panel: Results shown in orange, blue and purple are for models trained on $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ of the full range of halo masses, while those shown in red and green are for $\bar{f}_{\text{bar}}(M_{\text{halo}}^j)$ of intermediate mass haloes, $M_{\text{halo}}^j \in [10^{12}-10^{14}]$. Right panel: Results shown in orange, blue and purple are for models trained on $\bar{f}_{\text{bar}}(M_{\text{halo}} > 13.5)$. Results shown in red and green are for those trained on \bar{f}_{bar} of $M_{\text{halo}} \in [10^{12}-10^{14}]$.

This paper has been typeset from a \LaTeX file prepared by the author.