Multi-Epoch Machine Learning 2: Uncovering the variety of galaxy formation pathways in different cosmological simulations

Robert J. McGibbon,[1]⋆ Sadegh Khochfar[1]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ*

ABSTRACT

Using a novel machine learning method, we investigate the buildup of galaxy properties in different simulations, and in various environments within a single simulation. We compare how the stellar mass is dependent on the value of other galaxy and halo properties at points in time. To do this we utilise the feature importance values of a machine learning model which takes galaxy properties over a wide range of redshifts as it's input. By training the model on IllustrisTNG we show that stars are produced at earlier times in higher density regions of the universe than they are in low density regions. We also apply the technique to galaxies from Illustris, EAGLE, and the CAMELS simulations. We find that stellar mass is built up in a similar way in EAGLE and IllustrisTNG, but significantly differently in the original Illustris, suggesting that subgrid model physics is more important than the choice of hydrodynamics method used for a simulation. These differences are driven by the efficiency of supernova feedback. Applying principal component analysis to the CAMELS simulations allows us to identify a component associated with the importance of a halo's gravitational potential and another component representing the time at which galaxies form. We discover that the speed of galactic winds is a more important subgrid parameter than the parameter which gives the total energy per unit star formation. Finally we find that the SIMBA black hole feedback model has a larger effect on galaxy formation than the IllustrisTNG model.

Key words: keyword1 – keyword2 – keyword3

1 INTRODUCTION

Hydrodynamical cosmological simulations have become key tools for helping us to understand both cosmology and galaxy formation. These simulations include baryonic particles or cells alongside dark matter, and include relevant physical processes such as gas cooling, star formation, and feedback. Prominent examples of cosmological simulations include Illustris (Vogelsberger et al. 2014a,b; Genel et al. 2014; Sijacki et al. 2015), IllustrisTNG (Springel et al. 2018; Pillepich et al. 2018b; Naiman et al. 2018; Nelson et al. 2018; Marinacci et al. 2018), Simba (Davé et al. 2019), EAGLE (Schaye et al. 2015; McAlpine et al. 2016), HorizonAGN (Dubois et al. 2014), and FiBY (Johnson et al. 2013). They have been successful at reproducing a large number of observables, such as the column density distribution of the Lyman-α forest (Hernquist et al. 1996), stellar mass functions (Schaye et al. 2015; Pillepich et al. 2018b), galaxy clustering (Springel et al. 2018), the galaxy color bimodality (Kaviraj et al. 2017; Trayford et al. 2015; Nelson et al. 2018), and properties of the circumgalactic medium (Appleby et al. 2021). For recent reviews of cosmological simulations see Somerville & Davé (2015) and Vogelsberger et al. (2020).

There are various methods for simulating gas such as smoothed-particle hydrodynamics (e.g. Springel 2005), adaptive mesh refinement (e.g. Bryan et al. 2014), and moving-mesh (e.g. Springel 2010). Each comes with it's own strengths and weakness, and numerical ef-

fects from the implementations of the various methods can affect the hydrodynamics of the gas in different ways. More variation in simulations comes from the fact that many of the relevant physical processes that need to be modelled occur below the typical resolution limits of cosmological simulations. Therefore various 'sub-grid physics' prescriptions are employed to model them. There many different implementations of different subgrid models, and each model tends to have a number of tunable parameters.

Despite the differences in the modelling, subgrid models from different simulations can be tuned to reproduce observables. An example of this is shown in Figure 1. Although the simulations appear to have a similar stellar population at $z = 0$, the way in which galaxies built up their stellar mass can differ greatly between simulations, which is what the technique presented in this paper is able to distinguish.

One way to compare and contrast simulations to to fit them to a simple semi-analytic model (SAM), as introduced in Neistein et al. (2012), and then examine the resulting coefficients. Mitchell & Schaye (2022) fitted a set of linear coupled differential equations to the EAGLE simulation, and by examining the parameters of their fitted model they were able to gain understanding into what shapes the stellar-halo mass relation. Recently Alarcon et al. (2022) fit the star formation history (SFH) of galaxies from cosmological simulations using a differentiable parametric model. They showed that star formation in the UniverseMachine (Behroozi et al. 2019) simulation is less efficient and burstier than in IllustrisTNG.

As cosmological simulations output large volumes of data, ma-
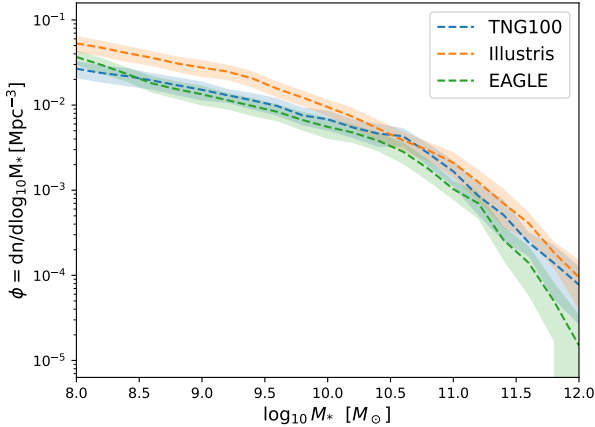
Figure 1. Stellar mass functions from the Illustris, TNG100, and EAGLE simulations. The shaded area corresponds to the spread computed from jackknife re-sampling over eight simulation sub-octants.

chine learning can be a useful technique for gaining understanding of the processes occurring within the simulation. Machine learning methods can be used for obtaining insights into both dark-matter-only and hydrodynamical simulations. We highlight a number of recent examples below.

Using N-body simulations Lucie-Smith et al. (2019) trained a decision-tree based model to predict the final mass on halos by using the initial conditions as input, and examined their model to determine the importance of the tidal shear field in establishing halo mass. In Lucie-Smith et al. (2022) the authors predicted the density profiles of halos, and included the mass accretion history of the halo as model input. They showed how the feature importance from the model could be interpreted in terms of physical timescales.

By training neural networks on hydrodynamical cosmological simulations, and then applying symbolic regression, Shao et al. (2022) were able to recover a version of the virial theorem. Shi et al. (2022) used random forests to learn which galaxy properties are most useful for determining the fraction of accreted stellar mass, and compared the feature importance values for high mass and low mass galaxy samples. Eisert et al. (2022) showed that by using invertible neural networks it is possible to gain information about the properties that are most predictive of the time of a galaxies' last major merger. Wadekar et al. (2020) used saliency maps from a convolutional neural network to show how the HI content of halos exhibits a strong dependence on their local environment. Villanueva-Domingo et al. (2021) trained a graph neural network to predict the mass of a halo based on its host galaxies, and examined which galaxy properties were the most predictive.

These types of machine learning methods are not only restricted for use with simulations, but can also be applied directly to observations to advance our understanding of galaxy formation processes. Bluck et al. (2022) and Piotrowska et al. (2022) both used the feature importance from trained random forest models to examine the most important parameters for predicting whether a galaxy is quenched. Curti et al. (2022) investigated which physical properties are most predictive for determining the position of galaxies on BPT diagrams by using neural networks and random forests, and Holwerda et al. (2022) used self-organizing maps to help analyse galaxy bimodalities.

One common application of machine learning to cosmological simulations has been to train a model to predict the baryonic properties of dark matter only halos (e.g. Moster et al. 2021; Machado Poletti Valle et al. 2021; Icaza-Lizaola et al. 2021; de Santi et al. 2022; Stiskalek et al. 2022; Chittenden & Tojeiro 2022; Jagvaral et al. 2022; Hausen et al. 2022; de Andres et al. 2022; Icaza-Lizaola et al. 2022). In McGibbon & Khochfar (2022) the authors introduced a new method from making these predictions, which used the properties of the subhalo over a wide range of times as model input. They showed how by examining the importance of the various input features that it was possible to extract insights into the physical processes occurring within the simulations used to train the models, specifically information on when was the most important time for determining the values of the different physical properties. In this work we expand on the method introduced in McGibbon & Khochfar (2022). We now use baryonic features as input to the model, rather than just halo properties, and examine the resulting feature importance plots. The purpose of this is to gain insight into the different formation processes that occur within different simulations.

The remainder of this paper is organized as follows. In Section 2 we give an overview of the simulations used in this work, focusing on the differences in their feedback subgrid model implementations. We also describe our machine learning method. In Section 3 we test the robustness of our method and show how it can highlight difference in galaxy populations by applying it to the IllustrisTNG simulation suite. In Section 4 we look at the insights we can gain when comparing different simulations. In Section 5 we apply our method to the CAMELS simulation suite, and show how the feature importance changes when subgrid model parameters are varied. We conclude in Section 6.

2 METHODS

2.1 Simulations

In this subsection we summarise the hydrodynamical cosmological simulations used in this work. Each simulation includes all significant physical processes required to track the evolution of dark matter, cosmic gas, luminous stars, and supermassive blackholes (SMBHs) from high redshifts ($z \sim 100$) to the present day $z = 0$. We will focus on the different subgrid implementations of supernova and black hole feedback. We quote the dark matter particle mass, m_{DM}, as a measure of the resolution of the simulation rather than attempting to directly compare the mass resolution of different baryonic elements. For all simulations the halos are first located using the friend-of-friends algorithm (FoF; Davis et al. (1985)), then substructure is identified using the SUBFIND subhalo finder (Springel et al. 2001).

2.1.1 Illustris

Illustris[1] (Vogelsberger et al. 2014a,b; Genel et al. 2014; Sijacki et al. 2015) is run with the moving mesh code AREPO (Springel 2010). The simulation adopts a WMAP-9 (Bennett et al. 2013) consistent cosmology, and merger trees are constructed using the SUBLINK algorithm (Rodriguez-Gomez et al. 2015). The box size is $(75 \ h^{-1}\text{Mpc})^3$, with $m_{DM} = 6.3 \times 10^6 M_\odot$.

Feedback associated with star formation is assumed to drive galactic scale outflows. The generated winds have a velocity scaled to the local dark matter velocity dispersion. The mass loading factor of the

---

[1] https://www.illustris-project.org/

Table 1. The parameters that are varied for each run of the CAMELS simulations. **Min** and **Max** give the minimum and maximum values that the parameters take on. **log scale** indicates whether the values are sampled linearly, or if they are varied with a logarithmic scale. **IllustrisTNG effect** (**SIMBA effect**) gives information about how the value of the parameter changes the feedback models within IllustrisTNG (SIMBA). For more details about the parameters see Villaescusa-Navarro et al. (2022).

Parameter	Min	Max	log scale	IllustrisTNG effect	SIMBA effect
Ω_m	0.1	0.5	No	Initial conditions	Initial conditions
σ_8	0.6	1	No	Initial conditions	Initial conditions
A_{SN1}	0.25	4	Yes	Energy output per unit star formation	Wind mass outflow rate per unit star formation
A_{AGN1}	0.25	4	Yes	Prefactor for power injected in kinetic mode	Prefactor for momentum flux of outflows
A_{SN2}	0.5	2	Yes	Speed of galactic winds	Speed of galactic winds
A_{AGN2}	0.5	2	Yes	Burstiness and temperature	Speed of jets

wind is calculated using the desired wind speed and available supernova energy. The direction of the wind is determined by the parent gas cell in such a way that wind particles are ejected preferentially along the rotation axis of spinning objects. SMBH feedback occurs in two different modes. If the Eddington ratio is below 0.05, a radio-mode model injects highly bursty thermal energy into large, ~50 kpc 'bubbles' which are displaced away from the central galaxy. Above this accretion rate, a quasar-mode model injects thermal energy into the immediately surrounding gas.

2.1.2 IllustrisTNG

The IllustrisTNG suite[2] (Springel et al. 2018; Pillepich et al. 2018b; Naiman et al. 2018; Nelson et al. 2018; Marinacci et al. 2018) is an update to Illustris simulation. It is also run using the Arepo code (Springel 2010), but a notable addition is the inclusion of magnetic fields. Cosmological parameters are set to the Planck 2015 values (Planck Collaboration et al. 2016). There are 3 different box sizes, each with its own resolutions. Comparable with the original Illustris simulation, TNG100 has a box size of $(75\ h^{-1}\mathrm{Mpc})^3$ and dark matter particles have $m_{\mathrm{DM}} = 7.5 \times 10^6 M_\odot$. TNG100 uses the same initial conditions as Illustris, although they have been adjusted for the updated cosmology. TNG300 has a box size of $(205\ h^{-1}\mathrm{Mpc})^3$ and $m_{\mathrm{DM}} = 5.9 \times 10^7 M_\odot$, while TNG50 has a box size of $(35\ h^{-1}\mathrm{Mpc})^3$ and $m_{\mathrm{DM}} = 4.5 \times 10^5 M_\odot$. Two sets of merger trees are available: one generated using Sublink (Rodriguez-Gomez et al. 2015), the second created by LHaloTree (Springel et al. 2005).

The TNG model for stellar feedback is based on the Illustris model, with some modifications. Winds are now ejected isotropically, although will still naturally propagate along the direction of least resistance. The wind velocity is now redshift-dependent, and a wind velocity floor is also introduced. The result of these changes is that stellar feedback in the TNG is more effective at suppressing star formation. The TNG also features two modes of SMBH feedback, with the mode being dependent on whether the Eddington ratio is above a critical value. For the TNG this critical value increases with the mass of the black hole. The high accretion thermal mode is the same as Illustris, but for low accretion rates a kinetic mode is used which adds momentum to neighbouring gas cells. For more details regarding the specific implementation of the IllustisTNG simulations, including all relevant subgrid models, we refer the reader to Weinberger et al. (2017) and Pillepich et al. (2018a).

2.1.3 EAGLE

EAGLE[3] (Schaye et al. 2015; McAlpine et al. 2016) is a suite of cosmological simulations run with the smoothed particle hydrodynamics code GADGET-3 (Springel 2005). It adopts a Planck 2013 cosmology (Planck Collaboration et al. 2014). Merger trees are built using the D-Trees algorithm (Jiang et al. 2014). The fiducial EAGLE simulation, named Ref-L100N1504, has a box size of $(100\ \mathrm{Mpc})^3$, and dark matter particles have $m_{\mathrm{DM}} = 9.7 \times 10^6 M_\odot$.

Supernova feedback in EAGLE is implemented by injecting thermal energy in a stochastic manner to nearby particles. The energy injected per unit stellar mass varies based on the metallicity and density of the interstellar medium (ISM). SMBH feedback in EAGLE is achieved using a single mode of feedback that operates at any Eddington ratio, in contrast to the dual modes of Illustris, TNG, and SIMBA. Feedback energy is stored until it is sufficient to heat the surrounding particles by $\Delta T = 10^{7.5}\mathrm{K}$ and then is stochastically injected as thermal energy. This 'pulsed' nature of the thermal feedback prevents the energy being immediately radiated away and offsets cooling. This makes it more efficient at quenching the galaxy than the corresponding thermal mode in Illustris and TNG.

Alongside the fiducial EAGLE simulation, we consider some variants run with the same resolution, but differing subgrid models. We utilise the FBconst and FBZ simulations, which are described in Crain et al. (2015). Both of these runs have been calibrated to match the observed stellar mass function. The FBconst model injects into the ISM a fixed amount of energy per unit stellar mass formed. For the FBZ simulation the energy associated with supernova feedback depends on the ISM metallicity, but unlike the fiducial model the energy is independent of gas density. We also examine the NoAGN run, which has the same subgrid models as the reference EAGLE simulation, but does not include black holes.

2.1.4 CAMELS simulations

The CAMELS project[4] (Villaescusa-Navarro et al. 2021, 2022) contains two different suites of state-of-the-art hydrodynamic simulations.

The simulations in the first suite have been run with the Arepo code (Springel 2010) and employ the same subgrid physics model as

---

[2] https://www.tng-project.org/

[3] http://icc.dur.ac.uk/Eagle; https://eagle.strw.leidenuniv.nl

[4] https://camels.readthedocs.io/

the IllustrisTNG simulations. See section 2.1.2 for more details on the subgrid models for this suite.

The simulations in the second suite have been run with the GIZMO code (Hopkins 2015) and employ the same subgrid physics model as the SIMBA simulation (Davé et al. 2019), which built on its precursor MUFASA (Davé et al. 2016) with the addition of supermassive black hole growth and feedback (Anglés-Alcázar et al. 2017). Star formation in the SIMBA model drives winds similar to those found in the TNG. The mass loading factor is scaled by redshift, and is constant for low mass galaxies. The SIMBA model has two SMBH kinetic feedback modes. At high Eddington ratios SMBHs drive multi-phase winds at velocities of $\sim 10^3$ km s^{-1}. At low Eddington ratios gas is heated to the virial temperature of the halo and ejected at velocities of $\sim 10^4$ km s^{-1}. For both of these modes the ejection is bipolar and parallel to the angular momentum vector of the SMBH accretion disc. X-ray feedback from SMBHs is also implemented, but it has a minimal effect on the galaxy mass function.

All simulations from both suites follow the evolution of 2×256^3 dark matter plus fluid elements in a periodic comoving volume of $(25 \ h^{-1}\mathrm{Mpc})^3$. All simulations share the value of the following cosmological parameters: $\Omega_b = 0.049$, $h = 0.6711$, $n_s = 0.9624$, $\sum m_\nu = 0.0$ eV, $w = -1$. However, each simulation has a different value of Ω_m and σ_8. Thesimulations also vary the values of four astrophysical parameters that control the efficiency of supernova and SMBH feedback: A_{SN1}, A_{SN2}, A_{AGN1}, and A_{AGN2}. Details about the effect of these parameters and the range of values they can take is given in Table 1.

The simulations with the different parameters are arranged into 4 sets. In the LH set, which contains 1000 simulations for each code, values are arranged on a latin-hypercube, and each simulation has a different random seed for initial conditions. We note that the latin-hypercubes of the IllustrisTNG and SIMBA simulations are different, i.e. there is no correspondence between simulations among the two sets. The 1P set contains simulations in which a single parameter varies, and all other parameters are kept fixed at their fiducial value. There are 11 simulations for each parameter, meaning there are 61 simulations for each code. The same initial conditions random seed is used for all the 1P simulations. The CAMELS project also contains a cosmic variance and extreme set, but we do not use either of these in our work.

The merger trees available in the CAMELS project are created using the CONSISTENT TREES code (Behroozi et al. 2013b) which is built on top of halos located using the ROCKSTAR algorithm (Behroozi et al. 2013a). However the rockstar halo catalogs do not contain all the galaxy properties we require for this work, such as star formation rate (SFR). Therefore we match the rockstar and subfind halos so that we have access to the data that we require. We use the method described in Gómez et al. (2022), but allow halos to be within 3x half mass radius to increase our sample size. Details about how well the catalogs are matched can be found in Appendix B.

2.2 Input and output features

As in McGibbon & Khochfar (2022), the inputs to our model are properties of a subhalo taken from a wide range of redshifts. A summary of our method is shown in Figure 2. Decision trees are invariant to the scaling of the input features, therefore we do not scale the input features in any way, despite the fact that their values span multiple orders of magnitude. As discussed in McGibbon & Khochfar (2022) we log the output features to prevent high mass galaxies being given a significantly higher weight than low mass ones.
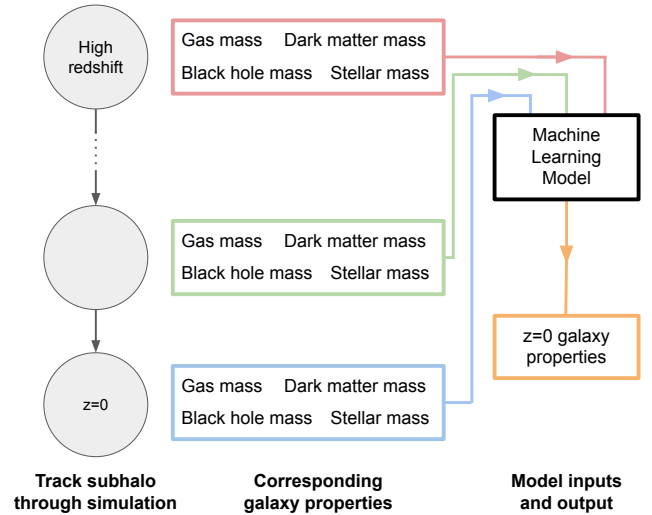

Figure 2. Summary of the method used in this work. The machine learning model takes in the four input features of the base model, but from a range of snapshots, not just redshift zero. The output for all models is the subhalo's baryonic properties at redshift zero.
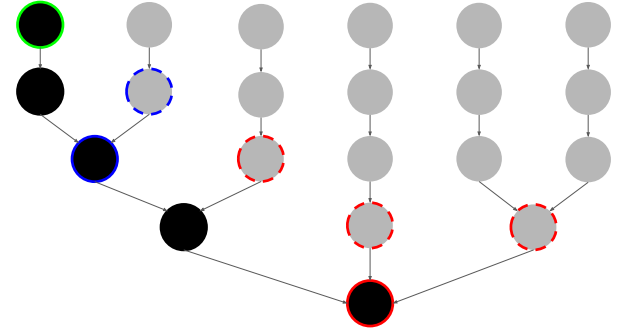

Figure 3. How subhalos that have merged are passed as input to the model. The black circles indicate the main progenitor branch of the merger tree, the grey circles indicates subhalos that merge with the main subhalo. Three snapshots are considered as input here, the green circle, blue circle, and red circle. The merger input feature is defined as the sum of the properties of the subhalos that have merged since the last snapshot that was used as input. This means that for the blue snapshot the single subhalo with the blue dashed line is used as the merger feature. For the red snapshot the merger feature is equal to the sum of the masses of the three subhalos with a red dashed outline.

In McGibbon & Khochfar (2022) we only considered the main progenitor branch as input to our model. In this work we consider the impact of other branches that merge. A schematic of how we include this information is shown in Figure 3.

2.3 Machine learning algorithms

2.3.1 ERT

Random forest regressors (Tin Kam Ho 1995; Breiman 2001) use an ensemble of decision trees to make a prediction. Each decision tree is constructed top-down from a root node. At each node the data is split into two bins based on the values of its input parameters. The
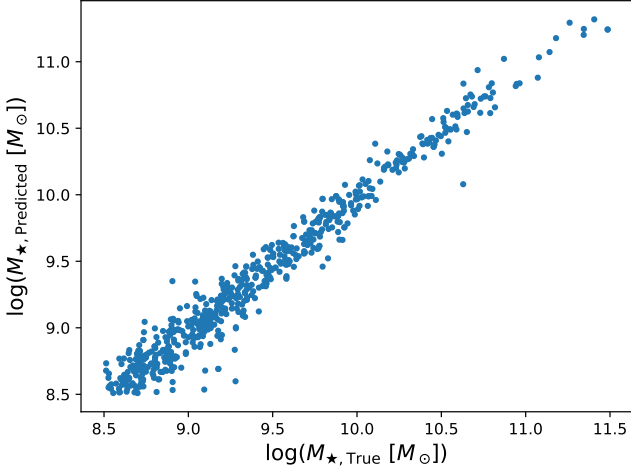
Figure 4. Model performance on 1000 randomly sampled galaxies from the TNG100 test set. The x axis shows the stellar mass from the simulation, and the y axis shows the value predicted by the machine learning model. This shows that the model has learned the relationship between the input features and stellar mass.

Figure 5. Feature importance of an ERT model trained to predict stellar mass of galaxies in the TNG100-1 simulation. Merger trees were generated using the LHaloTree algorithm.

plots the differences in the relative importance of each input feature should be considered, rather than their absolute values.

2.3.2 Principal Component Analysis (PCA)

In order to help visualize the results of applying our method to the CAMELS simulations, we make use of PCA (Shlens 2014; Jolliffe & Cadima 2016). We choose to use PCA over other available non-linear dimensionality reduction methods as it allows us to determine the feature importance of the reduced components that the algorithm finds.

3 SUBSAMPLES IN A SIMULATION

3.1 Is the model learning relationships?

In this work we show how the feature importance changes for different simulations and different subsamples of galaxies. In order for the feature importance to be meaningful, we need to ensure that the model has been able to successfully learn a relationship between the input and output features. In Figure 4 we show the true vs predicted stellar mass value for 1000 randomly sampled galaxies from TNG100. A model that made perfect predictions would correspond to all points lying on the diagonal. The small scatter in the figure tells us that the ERT model has successfully learnt a function mapping the input features to stellar mass, so the feature important values can be trusted.

3.2 Reading feature importance plots

Figure 5 show the feature importance obtained from a model trained to predict $z = 0$ stellar mass of galaxies in the TNG100 simulation. Merger trees were extracted using the LHaloTree algorithm. Each point on the plot corresponds to the importance of an input property at a certain time in the simulation. We include all input properties other than the one we are predicting, e.g. we do not use stellar mass as an input feature when predicting stellar mass. The sum of the feature importance is normalized to one for all models, so we only consider the relative importance of the input properties at different times rather than focusing on the absolute values. We highlight the fact that a large feature importance value does not necessarily that the

splits are chosen such that the weighted average of the mean squared error (MSE) of the two bins is minimized (Breiman et al. 1984). This partitioning of the data results in each leaf node at the bottom of the tree containing a small subset of the data, where almost all members of the subset have a similar output value. Predictions from decision trees are based on the assumption that test data points will have a similar output value to the other members of the leaf node it is placed into. A random forest is made up of a number of decision trees. There is a bootstrapping procedure such that each decision tree within the forest is trained on a randomly generated subset of the training data. Further randomness is added in that for each split only a subset of input features can be used. The prediction from a random forest is the average prediction of its component decision trees. A major advantage of random forests is that they are significantly less prone to overfitting data compared with a single decision tree. This results from the randomness added when training the individual decision trees.

For this work we use extremely randomised tree ensembles (ERT; Geurts (2006)). This is the algorithm used in previous work (e.g. Kamdar et al. 2016; Agarwal et al. 2018; Jo & Kim 2019; Moews et al. 2021; Lovell et al. 2022; McGibbon & Khochfar 2022) for predicting galaxy properties, and we found it to slightly outperform the standard random forest. It adds in additional randomization by computing a random split for each feature at each node, rather than the optimal split.

One major benefit to using ensembles based on decision trees is the ability to extract information on which input features are providing the information that is used to make the final predictions. For each decision tree the importance of an input feature can be determined by the number of times it is used for a split, and how close to the top of the tree those splits are. By averaging the importance values over all the decision trees within the ensemble model, a set of the feature importances can be determined for the model as a whole. However, one must be aware that correlations between input features will affect their importance values, and can make the results more difficult to interpret. The sum of the feature importance of all input features is normalised to one. Therefore when examining feature importance
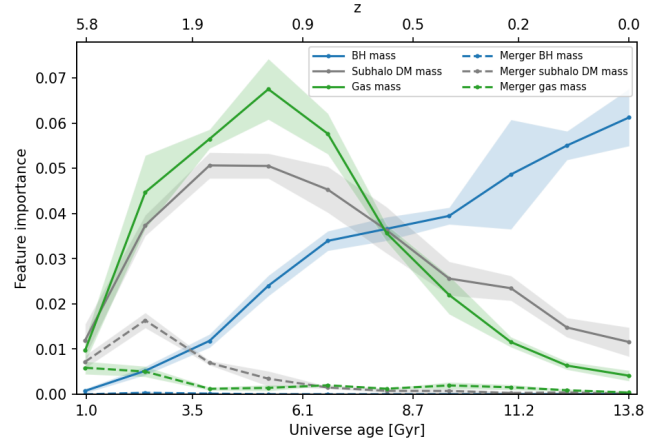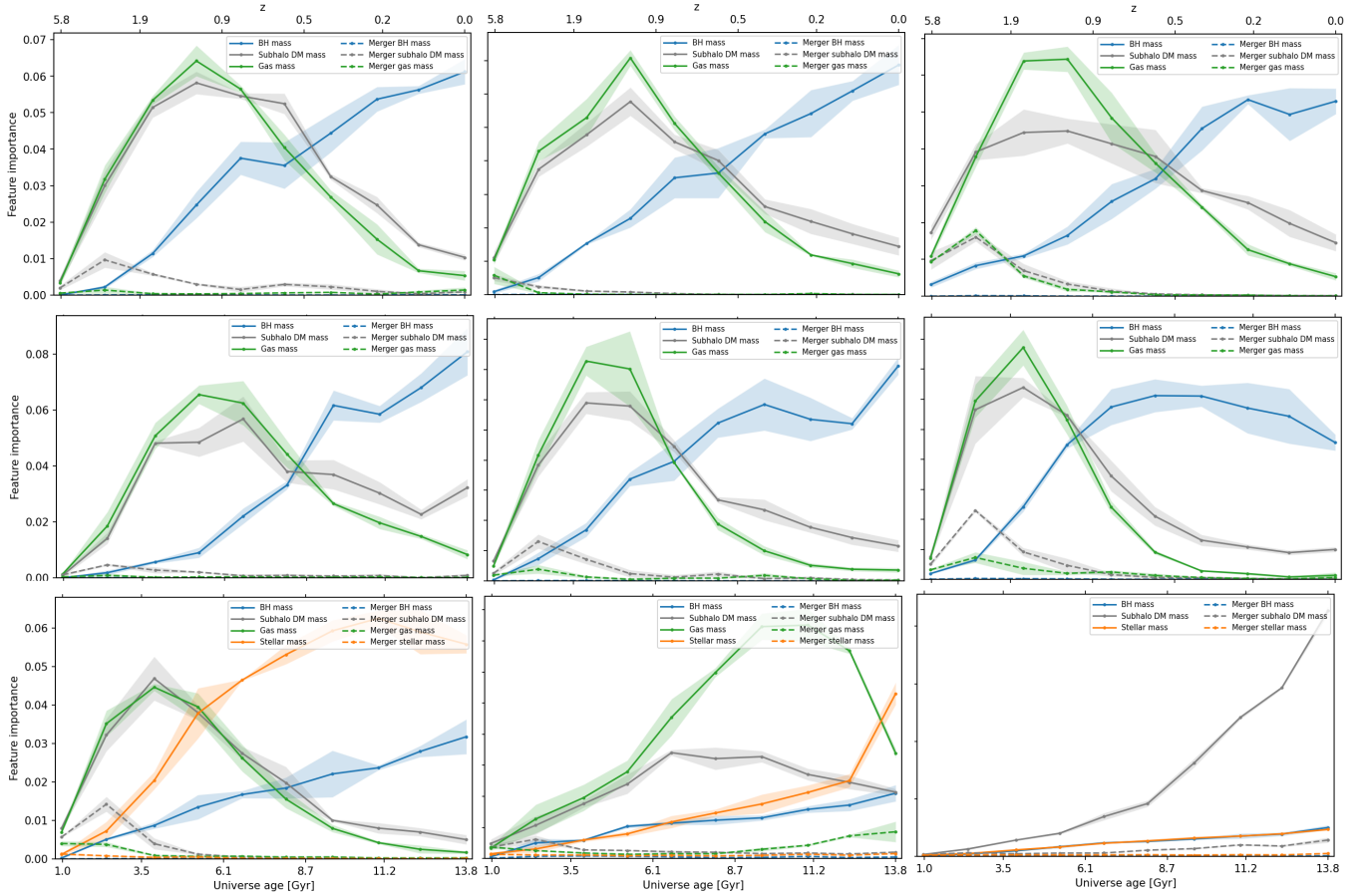
Figure 6. Feature importance plots from IllustrisTNG. All have the same subgrid models. **Top left** TNG50 - higher resolution simulation, **Top centre** TNG100 - sublink merger trees, **Top right** TNG300 - lower resolution simulation, **Middle left** $10^{12} < M_{FOF} < 10^{13}$ - Low density environment, **Middle centre** $10^{13} < M_{FOF} < 10^{14}$ - Medium density environment, **Middle right** $10^{14} < M_{FOF} < 10^{15}$ - High density environment, **Bottom left** Predicting $z = 0$ stellar metallicity, **Bottom centre** Predicting $z = 0$ SFR, **Bottom right** Predicting $z = 0$ gas mass

input feature has a large value at that point. For example, in Figure 5 the feature importance for the dark matter mass peaks at early times then drops off. This does not mean that the halos within IllustrisTNG are decreasing in dark matter mass, instead it indicates that the dark matter mass at early times is more informative for predicting the stellar mass at $z = 0$ than the dark matter mass at late times. We also note that a high feature importance value does not mean that the input feature has a positive correlation with the output feature being predicted. For example, we find that black hole mass is an important factor when predicting SFR, but for large galaxies black hole mass is negatively correlated with SFR. To give an error on the feature importance we train a model for 10 different train/test splits of the data. The shaded region in Figure 5 corresponds to the standard error taken on the feature importance values from the 10 different models.

We now interpret the feature importance plot shown in Figure 5. We see that the dark matter and gas mass feature importance peak at early times, then decrease. Physically this corresponds to the initial period of formation when the universe SFRD is highest and most stars are being formed. The black hole feature importance is similar in magnitude to that from gas and dark matter mass, but peaks at late times. Relative to the importance of the main progenitor branch, the feature importance of mergers is very small, and peaks earlier than the main progenitor feature importance. Black holes from mergers are

deemed to be completely unimportant. This is unsurprising as most halos that merge will be too small to host a black hole. However, the fact that the feature importance for black hole mergers is zero provides evidence that our model is not overfitting, since an overfitted model would end up using uninformative features to make splits.

3.3 Comparing feature importance plots

Figure 6 contains a number of feature importance plots from models trained on the IllustrisTNG simulation suite. In the top row we show the effect of varying resolution. The top centre plot gives the feature importance of a model trained to predict the stellar mass of TNG100 galaxies. This is the same as Figure 5, except the merger trees were generated using the SUBLINK algorithm. We see the same trends in the two plots, with all progenitor input properties peaking at the same point, and having the same relative importance. There is a minor difference in the importance of the merger features, with them being deemed less important for the SUBLINK merger trees. However, the overall agreement provides evidence of the robustness of this method to the choice of merger tree algorithm. For the remaining IllustrisTNG plots we use the LHaloTree merger trees.

The top left and top right panels show a model trained using galaxies from TNG50 and TNG300 respectively. The general trends are
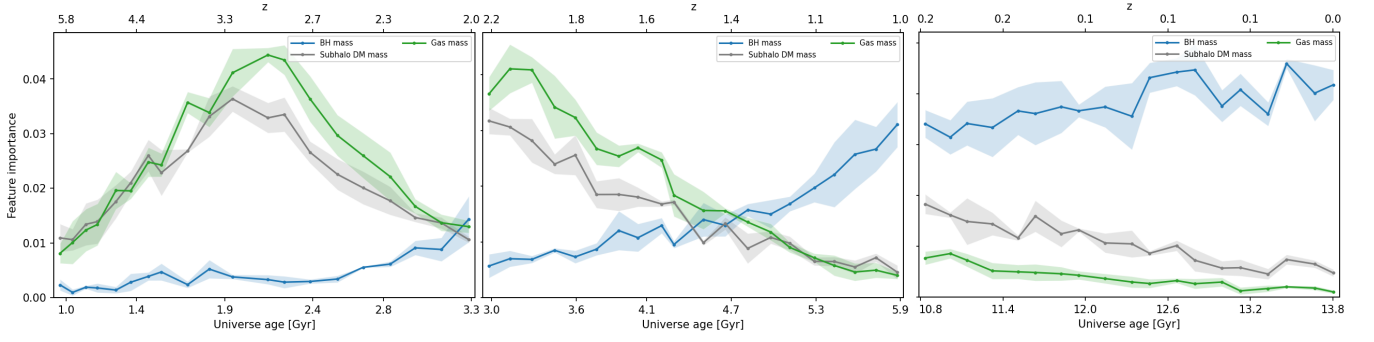
Figure 7. Feature importance plots for predicting the stellar mass of TNG100 galaxies at different redshifts. **Left** $z = 2$, **Centre** $z = 1$, **Right** $z = 0$

very similar to the model trained on TNG100, but some differences appear. For TNG50 the peak of the gas mass feature importance has a similar magnitude to the peak of the dark matter mass, whereas for TNG300 the peaks are further apart, and for TNG100 the distance between peaks is intermediate. Thus there is a clear trend in the relative importance of the gas mass and dark matter mass with increasing resolution. For low resolution you need a large amount of gas in order to have enough resolution elements to actually be able to collapse??? This isn't a problem for the higher resolution simulation??? There are no significant differences in the black hole feature importance.

In the middle row of Figure 6 we show the feature importance from models trained on galaxies taken from different density environments. All galaxies are taken from TNG100 and are split into three samples based on the mass of their FOF halo. The bin edges are given by $\log M_{FOF} = 12, 13, 14, 15$. The time at which the peak in gas and dark matter importance occurs shifts as the environment is varied. For galaxies in low density areas the peak occurs at later times, indicating delayed galaxy formation. For the galaxies in high density regions the majority of star formation occurs at earlier times. This also causes the black hole feature importance to peak prior to $z = 0$. As we increase density the importance of merger features relative to the progenitor features increases. This is to be expected as galaxies within groups and clusters will experience a large number of mergers.

In the final row of Figure 6 we show the feature importance from models trained to predict other properties of TNG100 galaxies at $z = 0$. The left, centre, and right panels correspond to output features of stellar metallicity, SFR, and gas mass respectively. The feature importance of dark matter and gas, and black hole mass in the stellar metallicity plot is similar to that from stellar mass plots.

The feature importance plot for SFR is significantly different to those for stellar mass, with features peaking close to or at $z = 0$. This is because SFR is an instantaneous property unlike stellar mass which builds up over time. As the dark matter mass gives the gravitational potential, which indicates how much gas will fall onto the halo. However, gas must cool before it can form stars, which explains why the dark matter feature importance peaks at earlier times than any other input property, as any gas which was recently accreted is unlikely to contribute to star formation. Of the merger features only gas mass is important, but it does indicate that mergers have a minor effect on the overall galaxy population SFR at $z = 0$.

For the prediction of gas mass the dark matter mass dominates. This is because we are predicting the gas mass of the halo, of which the majority is hot gas, rather than the mass of the ISM. Stellar mass and black hole mass do have some importance, but this plot shows

that in general the feedback in the IllustrisTNG model is insufficient to eject gas from halos.

3.4 Predictions at different redshifts

Figure 7 shows models trained to predict stellar mass for galaxies from TNG100 at different redshifts. The left, centre, and right panels show predictions for $z = 2, 1, 0$ galaxies respectively. For each redshift we use all snapshots from the 3 Gyr prior as input to the ERT model.

The leftmost $z = 2$ shows a clear peak in gas and dark matter mass around $z = 3$. This is despite the fact that there is still a large amount of star formation ongoing at $z = 2$. However, the gas needs time to collapse into the gravitational well of the halo, as well as radiate away energy, before it is able to form stars. This explains why the peak is located about 1Gyr prior to $z = 2$. In the $z = 1$ panel we see gas and dark matter mass continuously dropping, indicating that 3Gyr before $z = 1$ is already past the peak of star formation. However, the black hole feature importance continues to increase, indicating it's coupling to the stellar mass of galaxies. For the $z = 0$ plot the feature importance is nearly flat across time. This is a reflection of the lack of star formation in the majority of galaxies in the present epoch. A trend across the three panels is the increasing importance of black holes. This is because it takes time for black holes to build up and become effective, at $z = 2$ they are not that massive for most galaxies.

4 COMPARING SIMULATIONS

In Figure 8 we compare the feature importance from models trained on simulations with different subgrid model implementations. As these simulations have slightly differing cosmologies, the horizontal axes are not exactly aligned, but they are close enough for comparisions to still be valid. The top left panel shows the results from TNG100, as discussed in section 3. The top centre panel shows a model trained on the original Illustris simulation. There are three differences when compared with the TNG. Firstly the gas mass importance relative to dark matter is significantly increased for Illustris. Secondly the peaks in gas and dark matter mass are less pronounced for Illustris, and they occur at later times compared to the TNG. Both of these differences are reflections of the changes to the supernova feedback implementations. The feedback in TNG is more effective, which means a deep gravitational potential is needed to hold on to gas in star forming halos. This explains why dark matter importance increases for the TNG. The reduced efficiency of stellar feedback in Illustris means that star formation can continue for longer, which
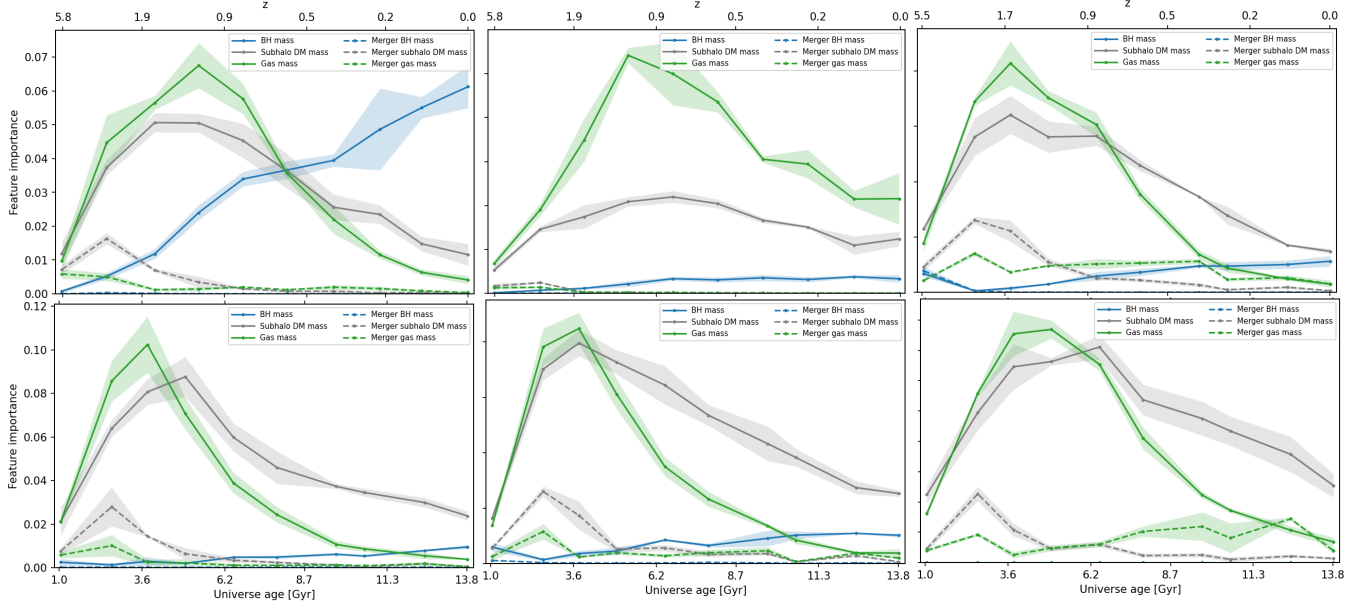
Figure 8. Feature importance plots for predicting stellar mass of simulations with different subgrid models **Top left** TNG100, **Top centre** Illustris, **Top right** EAGLE fiducial, **Bottom left** EAGLE FBconst, **Bottom centre** EAGLE FBZ, **Bottom right** EAGLE NoAGN. As the simulations have slightly different cosmologies, the horizontal axis are not exactly aligned.

is reflected in the shape and location of the peaks. The third difference is the relative importance of black holes, which are much more prominent in the TNG. This is a result of the new black hole feedback mode introduced in the TNG which boosts black hole feedback at low accretion rates. The merger trees for the Illustris simulation are created using the SUBLINK algorithm, which explains the differences in the merger feature importance.

The top right panel shows the results of the model trained on the fiducial EAGLE simulation. The gas mass and dark matter mass are close to those from the TNG, but the black hole importance is similar to Illustris. This is interesting as EAGLE is run with an SPH code, but both TNG and Illustris use a moving mesh to model the gas hydrodynamics. The fact that the EAGLE gas mass and dark matter mass importances are much closer to the TNG than the TNG is to Illustris shows that it is the subgrid models that are key to determining the correct build up of galaxy properties, rather than the hydrodynamic code which is used. The black hole importance shows that AGN feedback in EAGLE has similar efficiency to that in Illustris, despite the significantly different implementations.

The bottom right panel shows the EAGLE run without any black holes. Compared with the fiducial EAGLE run the gas and dark matter mass is more important at late times in the NoAGN run. This confirms that AGN do have any effect in shutting off some star formation in EAGLE.

The bottom left panel shows the EAGLE run where supernova feedback is independent of environment. The ISM dependence in the fiducial run makes feedback more efficient for low mass galaxies. Thus when the feedback is constant more star formation can occur in low mass galaxies, which means more stellar mass will build up at early times. This is reflected in the fact that the gas and dark matter peaks move to the left for the FBconst plot. There is also a decrease in the black hole importance.

The bottom centre panel displays the EAGLE run where supernova feedback depends only on metallicity, unlike the fiducial run where

it also depends on density. The peak occurs at a similar time to the FBconst run, showing that it is the density rather than the metallicity which is the factor determining the location of the peak of star formation density. However, the dark matter feature importance is more similar to the fiducial run, indicating the metallicity dependence is responsible for this feature.

5 CAMELS

In this section we apply to the method to the CAMELS suite. We first focus on the effect of varying the A_{SN2} parameter in the IllustrisTNG simulations from CAMELS, then compare the other parameters and the SIMBA simulations.

5.1 Correlations between supernova feedback and PCA components

For each of the $N = 1061$ IllustrisTNG simulations in the CAMELS suite we train an ERT model to predict the stellar mass at $z = 0$. From each of these models we extract the feature importances, and concatenate them into a vector. The feature importance vector from each model has a length of $M = 30$, which corresponds to 3 input properties (black hole mass, dark matter mass, gas mass) at 10 different snapshots. Combining the feature importances from all the simulations gives us a matrix of size $N \times M$. We apply PCA to this matrix, and show the results in Figure 9. In this plot each point corresponds to a single simulation. The horizontal axis corresponds to the first PCA component, and the vertical axis corresponds to the second PCA component. We color each of the points by the value of the A_{SN2} parameter of the simulation. Since A_{SN2} is sampled uniformly in log space for the LH and 1P sets, the colorbar is also logged.

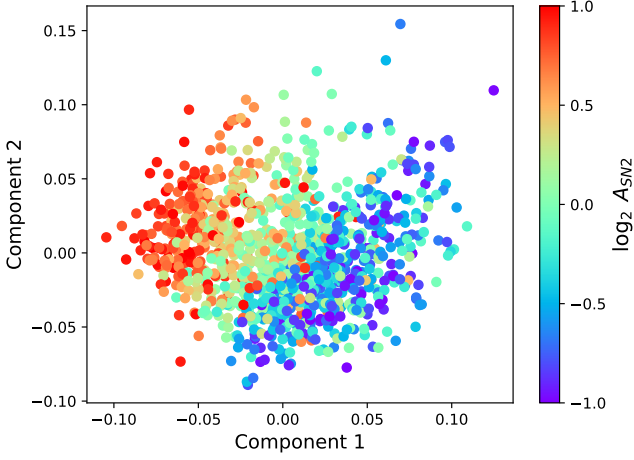We can see that there is a clear trend with supernova feedback in

Figure 9. PCA plot of stellar mass feature importance applied to the IllustrisTNG CAMELS simulations. Each point represents a single simulation. Points are colored by the speed of the supernova winds within the simulation.

the value of the PCA components, with a large A_{SN2} value corresponding to a low component 1 coefficient. There is a slight trend in component 2, but this is minor compared to the variation in component 1. Rather than showing scatter plots for all the PCA components, we summarise the information in Figure 9 in the bottom centre panel of Figure 11. Here we plot the mean value of the coefficient of the ith component for different A_{SN2} bins. This shows the negative correlation between the value of A_{SN2} and the first PCA component, along with the minor trend in the second component. This plot also allows us to examine other PCA components. We can see that there is some correlation with A_{SN2} and the third component. We have not plotted the fourth component to avoid overcrowding, but it does not show any correlation.

5.2 Physical interpretation of PCA components

Now that we have established the relationship between each of the PCA components and A_{SN2} we wish to determine what each component corresponds to physically. To do this we plot the feature importance of the PCA mean and each of the first three components in Figure 10. We would expect the feature importance of the mean to be similar to the one obtained from TNG100 (shown in Figure 5). We find the relative importance of each of the input properties to be similar, but the peak occurs later. This is a result of three factors which reduce our ability to track the halos through the simulation. Firstly the mass resolution is lower for CAMELS than for TNG100, which means resolution effects are more likely for small halos, and halos which exist at early times in TNG100 might not have enough particles to be identified in CAMELS. Another factor is that the spacing between snapshots is larger for CAMELS, meaning that it is more difficult to track halos between snapshots as they may have moved further, and gained/lost more mass. Finally, while we are able to match a high fraction of ROCKSTAR halos with their SUBFIND counterparts, some halos are matched incorrectly or no match can be found.

Unlike the previous feature importance plots, in Figure 10 each of the components must sum to zero by definition. This means it is possible to have negative importance values. This can be seen in the plot of component 1, where dark matter mass always has a negative value.
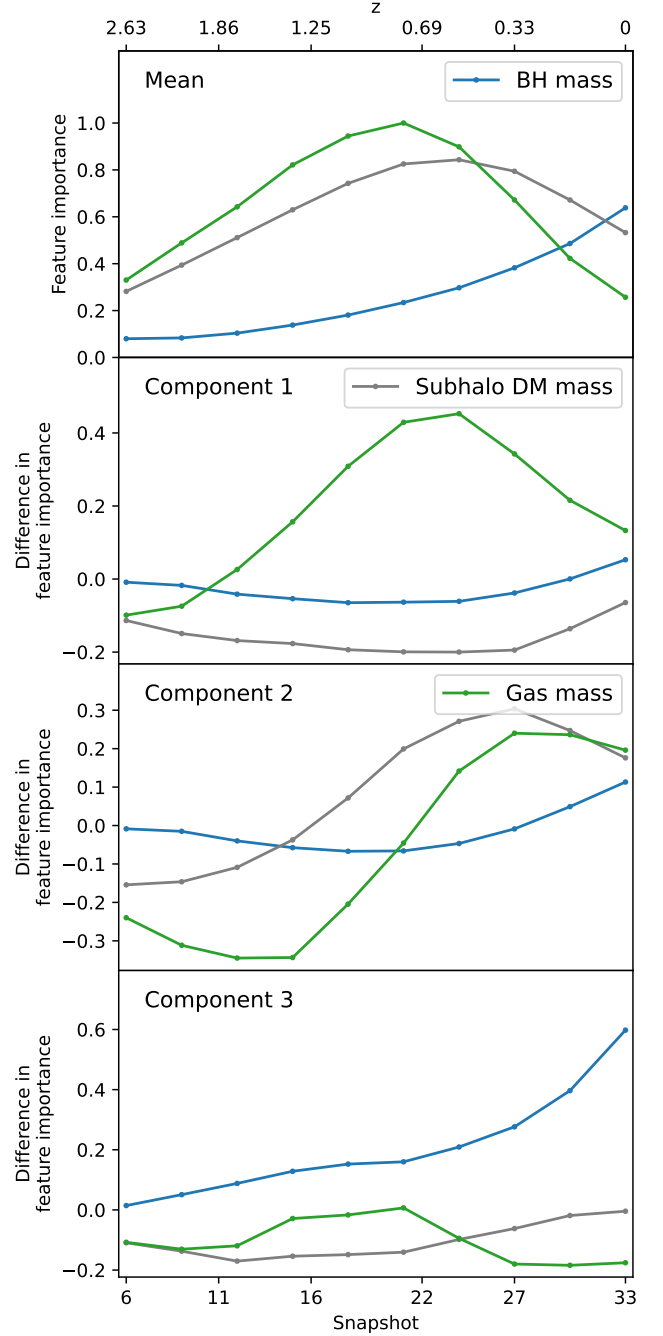

Figure 10. Plots of the value of each component resulting from PCA applied to the feature importance vectors of models trained to predict the stellar mass of galaxies from the CAMELS simulations.

Consider two simulations, A and B, where simulation A has a larger component 1 value than simulation B, and the coefficients for all the other components are identical. In this case the relative importance of gas mass to dark matter mass would be greater for simulation A than for simulation B. This means that increasing component 1 corresponds to a decreasing importance of dark matter mass, which gives the gravitational potential of the halo. When we increase the speed of supernova winds, they are more likely to blow gas out of the halo. This is why the dark matter mass is more important for
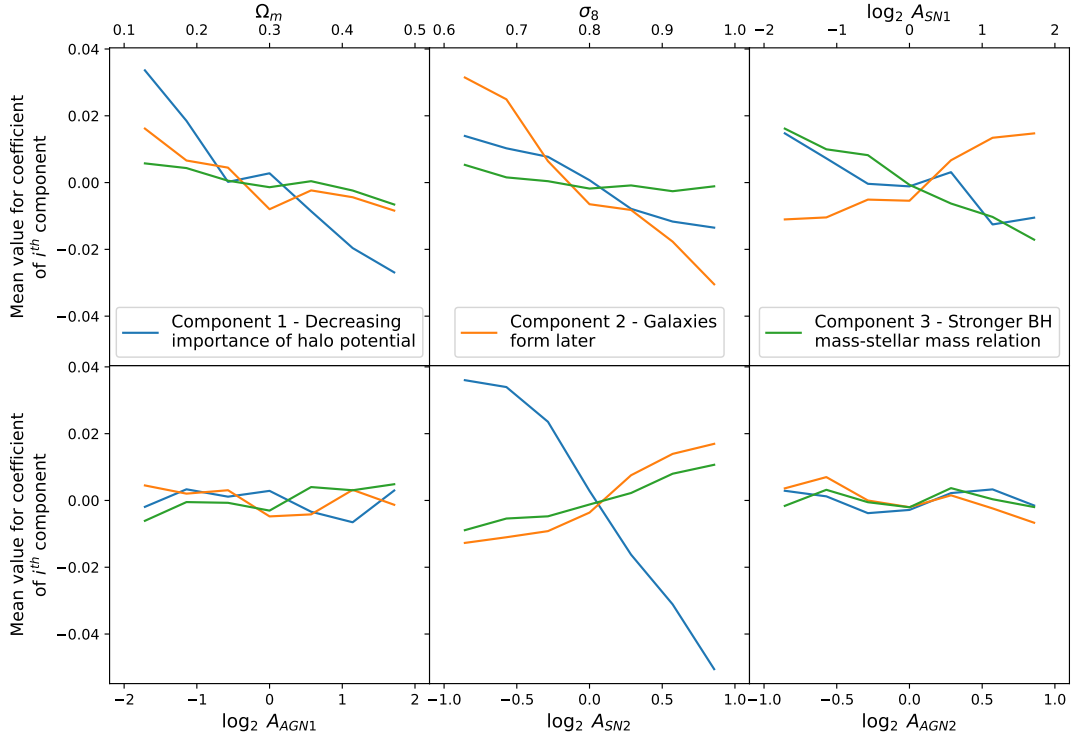
Figure 11. The effect of varying the simulation parameters on the PCA components. Models are trained on the IllustrisTNG simulations from CAMELS. **Top left** Omega matter, **Top centre** Sigma 8, **Top right** A_{SN1}, **Bottom left** A_{AGN1}, **Bottom centre** A_{SN2}, **Bottom right** A_{AGN2}. For a description of each parameter see Table 1.
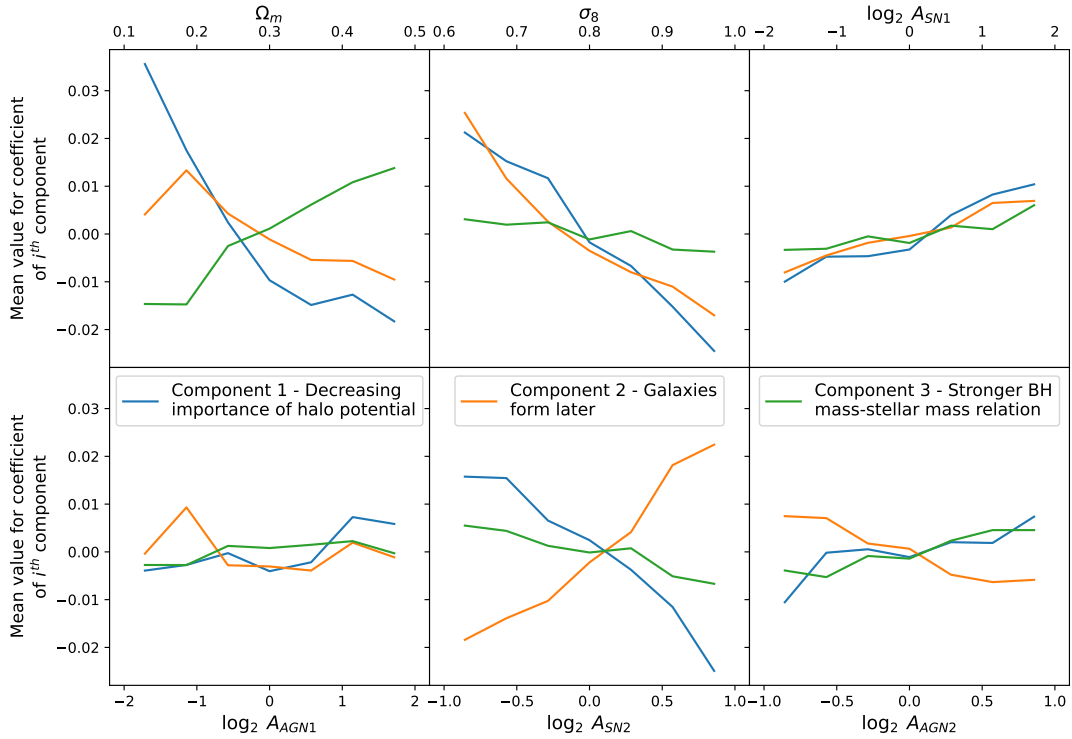

Figure 12. Same as Figure 11, but using the SIMBA simulation from the CAMELS suite. Components are determined from PCA applied to the IllustrisTNG simulations only

simulations with a large A_{SN2} value, and explains the bifurcation shown in Figure 9.

When plotting component 2, we see that there are negative values of gas and dark matter mass at early times, and positive values at late times. Therefore increasing the component 2 value means that the peak in gas and dark matter mass will occur at a later point in time, corresponding to galaxies which form later. Increasing A_{SN2} causes galaxies to form later as the gas is ejected further from the galaxy, and so takes longer to cool and return before it can form stars.

For component 3 the difference in gas and dark matter mass is always negative, but is relatively flat. Thus component 3 corresponds to an increasing importance of the black hole.

5.3 Comparing parameters

We now consider the effect of the varying the other simulation parameters. We have created an online application [5] which estimates the feature importance for a given set of parameters. The plots shown by the app validate the results found by PCA in the previous section. For example, if we are predicting stellar mass and we increase the value of the supernova 2 parameter, we see that the dark matter feature importance increases. To generate the plots for this app we train a Random Forest model that takes simulation parameters as input and predicts feature importance.

From Figure 11 we can see the effect of modifying the simulation parameters on the feature importance of the CAMELS TNG simulations. The vertical axis is set to the same scale for all plots to allow for comparison.

The top left panel shows the effect of changing Ω_m. For low Ω_m values galaxies form earlier. It is interesting to compare this with the feature importance plots from the different density environments as shown in Figure 6. In that case we found that galaxies in low density regions tended to form later. Low density regions can be regarded as a separate universe with a low Ω_m value, so this agrees with our findings from the PCA components. However, for the different density regions there was no difference in the gas and dark matter mass relative importance, but in CAMELS we do find a correlation with Ω_m.

Changing the value of σ_8 also effects the time at which galaxies form. Since for low σ_8 values the density peaks are smaller, it takes longer for the halos to collapse and allow galaxies to form.

Varying A_{SN1}, which increases the energy per unit star formation, causes similar but less pronounced effects to A_{SN2}. However, it has the opposite effect on the black hole mass relation.

Changing the two parameters associated with the AGN feedback strength has no effect on the mean PCA coefficients. The reason for this could be because the number of galaxies without supermassive black holes is significantly larger than the number of galaxies with black holes. Ideally we would consider only galaxies above a certain mass cut to exclude those without any black hole activity, but the $25(Mpc/h)^3$ box size of the CAMELS simulation is not large enough for us to do this.

5.4 Comparing IllustrisTNG and SIMBA

We now apply the same analysis to the SIMBA simulations from the CAMELS suite. We decompose the feature importance vectors from SIMBA using the PCA components we obtained from the TNG galaxies. This allows for a more direct comparison between the two

codes. We remind the reader that due to the different subgrid model implementations the subgrid model parameters (e.g. A_{SN1}) have different meaning for TNG and SIMBA. See Table 1 for a description of each parameter.

In Figure 12 we show the results of this analysis. In general we find the same trends as in the TNG simulations, but there is often a difference with component 3, which is mainly linked to the black hole importance. This is the case for Ω_m, σ_8, and A_{SN2}.

When increasing the value A_{SN1} for SIMBA we find that galaxies form later, which also occurs for the TNG, but that the halo potential importance decreases, opposite to the TNG.

Unlike for the TNG, changing the black hole feedback parameters does have an impact on the feature importance. For the bottom right panel we see that decreasing A_{AGN2} means that galaxies form later, due to the fact that their star formation is not being shut down. However, A_{AGN1} still shows no clear trend with any of the PCA components.

6 CONCLUSIONS

The last numbered section should briefly summarise what has been done, and describe the final conclusions which the authors draw from their work.

DATA AVAILABILITY

The IllustrisTNG and Illustris data used in this work is available from the IllustrisTNG website (http://www.tng-project.org). The CAMELS data The EAGLE data The code used to produce this paper is available at https://github.com/robjmcgibbon/baryon_tree_ml

The analysis in this paper depended on the following packages of the python programming language: NumPy (Harris et al. 2020), Matplotlib (Hunter 2007), and Scikit-learn (Pedregosa et al. 2011). We are thankful to the developers of these tools. This research has made intensive use of NASA's Astrophysics Data System (http://ui.adsabs.harvard.edu) and the arXiv eprint service (http://arxiv.org).

REFERENCES

Agarwal S., Davé R., Bassett B. A., 2018, MNRAS, 478, 3410
Alarcon A., Hearin A. P., Becker M. R., Chaves-Montero J., 2022, arXiv e-prints, p. arXiv:2205.04273
Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017, MNRAS, 464, 2840
Appleby S., Davé R., Sorini D., Storey-Fisher K., Smith B., 2021, MNRAS, 507, 2383
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bennett C. L., et al., 2013, ApJS, 208, 20
Bluck A. F. L., Maiolino R., Brownson S., Conselice C. J., Ellison S. L., Piotrowska J. M., Thorp M. D., 2022, A&A, 659, A160
Breiman L., 2001, Mach. Learn., 45, 5–32

---

[5] Available at https://camels-feature-importance.streamlitapp.com/

Breiman L., Friedman J. H., Olshen R. A., Stone C. J., 1984, Classification and Regression Trees. Wadsworth and Brooks, Monterey, CA

Bryan G. L., et al., 2014, ApJS, 211, 19

Chittenden H. G., Tojeiro R., 2022, arXiv e-prints, p. arXiv:2203.12702

Crain R. A., et al., 2015, MNRAS, 450, 1937

Curti M., et al., 2022, MNRAS, 512, 4136

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

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

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

Eisert L., Pillepich A., Nelson D., Klessen R. S., Huertas-Company M., Rodriguez-Gomez V., 2022, arXiv e-prints, p. arXiv:2202.06967

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

Geurts P., 2006, Mach. Learn., 63, 3–42

Gómez J. S., Padilla N. D., Helly J. C., Lacey C. G., Baugh C. M., Lagos C. D. P., 2022, MNRAS, 510, 5500

Harris C. R., et al., 2020, Nature, 585, 357

Hausen R., Robertson B. E., Zhu H., Gnedin N. Y., Madau P., Schneider E. E., Villasenor B., Drakos N. E., 2022, arXiv e-prints, p. arXiv:2204.10332

Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51

Holwerda B. W., et al., 2022, MNRAS, 513, 1972

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

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Icaza-Lizaola M., Bower R. G., Norberg P., Cole S., Schaller M., Egan S., 2021, MNRAS, 507, 4584

Icaza-Lizaola M., Bower R. G., Norberg P., Cole S., Schaller M., 2022, arXiv e-prints, p. arXiv:2205.08799

Jagvaral Y., Lanusse F., Singh S., Mandelbaum R., Ravanbakhsh S., Campbell D., 2022, arXiv e-prints, p. arXiv:2204.07077

Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115

Jo Y., Kim J.-h., 2019, MNRAS, 489, 3565

Johnson J. L., Dalla Vecchia C., Khochfar S., 2013, MNRAS, 428, 1857

Jolliffe I. T., Cadima J., 2016, Philosophical Transactions of the Royal Society of London Series A, 374, 20150202

Kamdar H. M., Turk M. J., Brunner R. J., 2016, MNRAS, 457, 1162

Kaviraj S., et al., 2017, MNRAS, 467, 4739

Lovell C. C., Wilkins S. M., Thomas P. A., Schaller M., Baugh C. M., Fabbian G., Bahé Y., 2022, MNRAS, 509, 5046

Lucie-Smith L., Peiris H. V., Pontzen A., 2019, MNRAS, 490, 331

Lucie-Smith L., Adhikari S., Wechsler R. H., 2022, arXiv e-prints, p. arXiv:2205.04474

Machado Poletti Valle L. F., Avestruz C., Barnes D. J., Farahi A., Lau E. T., Nagai D., 2021, MNRAS, 507, 1468

Marinacci F., et al., 2018, MNRAS, 480, 5113

McAlpine S., et al., 2016, Astronomy and Computing, 15, 72

McGibbon R. J., Khochfar S., 2022, MNRAS, 513, 5423

Mitchell P. D., Schaye J., 2022, MNRAS, 511, 2948

Moews B., Davé R., Mitra S., Hassan S., Cui W., 2021, MNRAS, 504, 4024

Moster B. P., Naab T., Lindström M., O'Leary J. A., 2021, MNRAS, 507, 2115

Naiman J. P., et al., 2018, MNRAS, 477, 1206

Neistein E., Khochfar S., Dalla Vecchia C., Schaye J., 2012, MNRAS, 421, 3579

Nelson D., et al., 2018, MNRAS, 475, 624

Pedregosa F., et al., 2011, Journal of machine learning research, 12, 2825

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

Pillepich A., et al., 2018b, MNRAS, 475, 648

Piotrowska J. M., Bluck A. F. L., Maiolino R., Peng Y., 2022, MNRAS, 512, 1052

Planck Collaboration et al., 2014, A&A, 571, A1

Planck Collaboration et al., 2016, A&A, 594, A13

Rodriguez-Gomez V., et al., 2015, MNRAS, 449, 49

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

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

Shi R., Wang W., Li Z., Han J., Shi J., Rodriguez-Gomez V., Peng Y., Li Q., 2022, MNRAS,

Shlens J., 2014, arXiv e-prints, p. arXiv:1404.1100

Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575

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

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

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

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

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

Stiskalek R., Bartlett D. J., Desmond H., Anbajagane D., 2022, MNRAS,

Tin Kam Ho 1995, in Proceedings of 3rd International Conference on Document Analysis and Recognition. pp 278–282 vol.1, doi:10.1109/ICDAR.1995.598994

Trayford J. W., et al., 2015, MNRAS, 452, 2879

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

Villaescusa-Navarro F., et al., 2022, arXiv e-prints, p. arXiv:2201.01300

Villanueva-Domingo P., et al., 2021, arXiv e-prints, p. arXiv:2111.08683

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

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

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Wadekar D., Villaescusa-Navarro F., Ho S., Perreault-Levasseur L., 2020, arXiv e-prints, p. arXiv:2012.00111

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

de Andres D., Yepes G., Sembolini F., Martínez-Muñoz G., Cui W., Robledo F., Chuang C.-H., Rasia E., 2022, arXiv e-prints, p. arXiv:2204.10751

de Santi N. S. M., Rodrigues N. V. N., Montero-Dorta A. D., Abramo L. R., Tucci B., Artale M. C., 2022, MNRAS,

APPENDIX A: BOX SIZE NEEDED?

APPENDIX B: MATCHING ROCKSTAR AND SUBFIND CATALOGS

This paper has been typeset from a T_EX/L^AT_EX file prepared by the author.