

Small-scale galaxy clustering in the EAGLE simulation

M. Celeste Artale^{1*}, Susana E. Pedrosa¹, James W. Trayford², Tom Theuns²,
Daniel J. Farrow⁴, Peder Norberg^{2,3}, Idit Zehavi⁵, Richard G. Bower^{2,3},
and Matthieu Schaller²

¹ *Instituto de Astronomía y Física del Espacio (IAFE, CONICET-UBA), C.C. 67 Suc. 28, C1428ZAA Ciudad de Buenos Aires, Argentina*

² *Institute for Computational Cosmology, Department of Physics, Durham University, South Road, DH1 3LE Durham, UK*

³ *Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, DH1 3LE Durham, UK*

⁴ *Max-Planck-Institut für extraterrestrische Physik, Postfach 1312 Giessenbachstrasse, D-85741 Garching, Germany*

⁵ *Department of Astronomy, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA*

17 November 2016

ABSTRACT

We study present-day galaxy clustering in the EAGLE cosmological hydrodynamical simulation. EAGLE’s galaxy formation parameters were calibrated to reproduce the redshift $z = 0.1$ galaxy stellar mass function, and the simulation also reproduces galaxy colours well. The simulation volume is too small to correctly sample large-scale fluctuations and we therefore concentrate on scales smaller than a few megaparsecs. We find very good agreement with observed clustering measurements from the Galaxy And Mass Assembly (GAMA) survey, when galaxies are binned by stellar mass, colour, or luminosity. However, low-mass red-galaxies are clustered too strongly, which is at least partly due to limited numerical resolution. Apart from this limitation, we conclude that EAGLE galaxies inhabit similar dark matter haloes as observed GAMA galaxies, and that the radial distribution of satellite galaxies as function of stellar mass and colour is similar to that observed as well.

Key words: galaxies: formation – galaxies: evolution – galaxies: haloes – galaxies: statistics – large-scale structure of Universe – cosmology: theory

1 INTRODUCTION

The spatial distribution of galaxies provides a powerful way to probe both cosmology and galaxy formation. Galaxy clustering measurements on scales where density fluctuations are only mildly non-linear, combined with other cosmological data sets such as CMB measurements, put impressively tight constraints on cosmological parameters (e.g. [Hinshaw et al. 2013](#); [Planck Collaboration et al. 2016](#)). In addition, the detection of baryon acoustic oscillations in the clustering of galaxies (e.g. [Cole et al. 2005](#); [Eisenstein et al. 2005](#)) opened up the way to quantify the nature of dark energy (e.g. [Laureijs et al. 2011](#)), and combined with redshift-space distortion measurements, test theories of gravity (e.g. [Linder 2008](#)).

On the galaxy formation perspective, the clustering of galaxies inform us about the relation between galaxies and the underlying dark matter and can provide hints about the physical processes involved in galaxy assembly history. As galaxies reside within the dark matter haloes, their posi-

tions trace the underlying cosmic structure. While the formation and evolution of the dark matter haloes is governed exclusively by gravitational interaction, the assembly of the galaxies is governed by the more complex baryon physics that also affects the distribution of galaxies. This galaxy bias can impact as well cosmological inferences made from galaxy clustering measurements.

The main statistical tool used for characterising galaxy clustering is the two-point correlation function $\xi(r)$, which measures the excess probability over random of finding pairs of galaxies at different separations r ([Peebles 1980](#)). Commonly, when analysing redshift surveys, the projected correlation function integrated along the line-of-sight is used, in order to eliminate in principle redshift-space distortions ([Davis & Peebles 1983](#)). Observations show that brighter, redder and more massive galaxies are more strongly clustered and related trends are also measured as a function of morphology and spectral type (e.g., [Norberg et al. 2002a](#); [Zehavi et al. 2002, 2005](#); [Goto et al. 2003](#); [Li et al. 2006](#); [Coil et al. 2006](#); [Croton et al. 2007](#); [Zheng et al. 2007](#); [Coil et al. 2008](#); [Zehavi et al. 2011](#); [Coupon et al. 2012](#); [Guo et al. 2013](#); [Farrow et al. 2015](#)).

* E-mail: mcartale@iafe.uba.ar

The theoretical modelling of galaxies plays an important role in interpreting clustering data, since, although galaxy clustering on large enough scales is very similar to that of the underlying matter distribution, it is not expected to be identical. Such models are also routinely used to estimate sample variance and verify methods for correcting observational biases. Several theoretical schemes are able to model galaxy clustering in volumes comparable to those probed observationally. These start from a dark matter only (DMO) N -body simulation, and populate haloes or sub-haloes with galaxies. Halo occupation distribution models (HOD, e.g. Cooray 2002; Berlind & Weinberg 2002; Tinker et al. 2012) or sub-halo abundance matching (SHAM, e.g. Conroy et al. 2006; Vale & Ostriker 2004, 2006) are statistical techniques that match galaxies with haloes by abundance based on, for example, their circular velocity. Semi-analytical galaxy formation techniques (e.g. Kauffmann et al. 1993; Cole et al. 1994) use physically motivated schemes to associate galaxies to haloes (see e.g. Baugh 2006, for a review).

Notwithstanding the successes of these methods, they suffer from intrinsic limitations. Stellar and AGN feedback from forming galaxies affect the mass of their halo (Sawala et al. 2013; Velliscig et al. 2014; Schaller et al. 2015), limiting the extent to which any DMO simulation predicts the clustering of haloes as function of mass accurately. Feedback effects are plausibly strong enough to affect the mass distribution itself (van Daalen et al. 2011; Semboloni et al. 2011) and with it galaxy clustering (Hellwing et al. 2016). These effects may be relatively small, but the main limitation of the models is on smaller scales, where several effects that occur when a galaxy becomes a satellite (tidal interactions, ram-pressure stripping, strangulation) come into play. van Daalen et al. (2014) investigate how physics behind galaxy formation affect the clustering of galaxies at small scales through comparing two models of OWLS project (Schaye et al. 2010) with and without AGN feedback. They found that the physics of galaxy formation affects clustering on small scales, and in addition affects larger scales through its impact on the masses of sub halos. Farrow et al. (2015) show how clustering in galaxy light cone mocks generated with two versions of the semi-analytical GALFORM code (Lacey et al. 2016; Gonzalez-Perez et al. 2014) differ significantly with the observations on small scales. McCullagh et al. (*in prep.*) shows that once a more detailed merger scheme is considered, such as that described by Simha & Cole (2013) reasonably good agreement with clustering data on small scale is achieved. This is in agreement with the findings of Contreras et al. (2013) in which different families of galaxy formation models were compared to clustering data. McCarthy et al. (2016) present results from the BAHAMAS cosmological hydrodynamical simulation. These simulations were designed using a similar calibration strategy as EAGLE but using the OWLS implementation of galaxy formation described by Schaye et al. (2010). The simulated volume (400 Mpc/ h on a side) and mass resolution (initial baryonic particle mass $8 \times 10^8 h^{-1} M_\odot$) allow them to probe the galaxy correlation function on large scales, but not to go to lower-mass galaxies and smaller scale clustering that we concentrate on here, nor to investigate clustering as function of galaxy property such as for example colour. Sales et al. (2015) compare the

distribution of satellite galaxies from the ILLUSTRIS simulation (Vogelsberger et al. 2014) as function of colour to Sloan Digital Sky Survey (SDSS, York et al. 2000) observations by Wang et al. (2014). They attribute the better agreement of the simulations compared to semi-analytical models to the more realistic gas contents of satellites at infall.

Galaxy clustering on small scales, and as a function of intrinsic galaxy properties such as luminosity, colour and star formation rate, thus might prove to be a stringent test of galaxy formation models. Performing such a test is the aim of this paper: we explore the clustering of galaxies in the cosmological hydrodynamical EAGLE simulation (Schaye et al. 2015) and its dependence on galaxy properties. The galaxy formation model of EAGLE uses sub-grid modules that are calibrated to reproduce the present-day stellar mass function (as described by Crain et al. 2015). In addition, the EAGLE simulation reproduces relatively well the colours and luminosities of galaxies both in the infra-red (Camps et al. 2016) and at optical wavelengths (Trayford et al. 2016). Thus, although the simulated volume is too small to sample large scale modes accurately, the realism of the simulation might yield reasonable clustering statistics on small scales. We compare the EAGLE predictions to clustering measurements by Farrow et al. (2015) of galaxies in the Galaxy and Mass Assembly redshift survey (GAMA, Driver et al. 2011; Liske et al. 2015), which are in accord with the Zehavi et al. (2011) SDSS measurements. For completeness we note that Crain et al. (2016) show that EAGLE reproduces the observed clustering of $z = 0$ HI sources.

This paper is organised as follows. In § 2 we describe the main characteristics of the simulations used, and briefly discuss the GAMA survey to which we compare. In § 3 we define the notation and present the tools used to measure galaxy clustering. Simulations and observations are compared in § 4. In the discussion section, § 5, we compare EAGLE with the clustering in GALFORM (following the analysis of Farrow et al. 2015) and our own clustering measurements using the database of the ILLUSTRIS (Vogelsberger et al. 2014) simulation. The conclusions are summarised in § 6.

Throughout this paper and unless specified otherwise, we use the Planck Collaboration et al. (2014) values of the cosmological parameters ($\Omega_b = 0.0482$, $\Omega_{\text{dark}} = 0.2588$, $\Omega_\Lambda = 0.693$ and $h = 0.6777$, where $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$). Observational measures of clustering are (most commonly) specified in ‘ h ’-dependent units, and to ease comparison to other clustering studies, we will express distances in $h^{-1} \text{ Mpc}$ and masses in $h^{-2} M_\odot$.

2 SIMULATIONS AND DATA

This section briefly describes the simulations used, and the GAMA survey to which clustering results of the simulations are compared.

2.1 The EAGLE hydrodynamical simulation suite

We use the ‘reference’ EAGLE simulation from Table 2 in Schaye et al. (2015) (i.e. L0100N1504, but hereafter referred to as the EAGLE simulation), a hydrodynamical cosmological simulation that start at $z = 127$ from

Table 1. Numerical parameters of cosmological simulations considered. From left to right: simulation identifier, simulation co-moving side length L , initial mass m_g of baryonic particles, dark matter particle mass, Plummer-equivalent co-moving length (ϵ_{com}) and maximum proper gravitational softening length ϵ_{prop} . The EAGLE and EAGLE-DMO simulations are referred to as L0100N1504 and L0100N1504-DMO by Schaye et al. (2015).

Name	L Mpc	m_g $10^6 M_\odot$	m_{dm} $10^6 M_\odot$	ϵ_{com} kpc	ϵ_{prop} kpc
EAGLE	100	1.81	9.70	2.66	0.7
EAGLE-DMO	100	-	11.51	2.66	0.7
P-MILLENNIUM	800	-	157	3.40	3.40

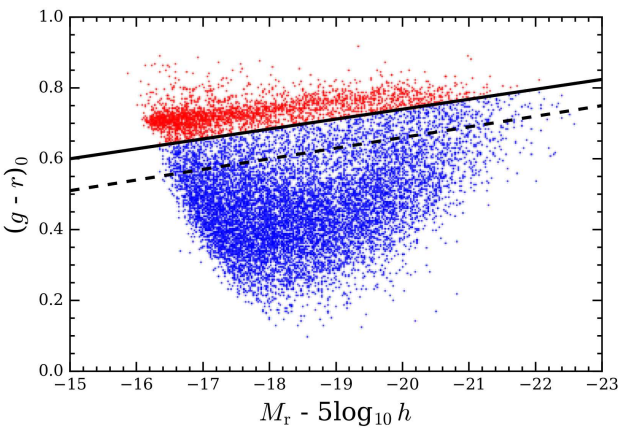


Figure 1. Distribution of rest-frame $(g - r)_0$ colour versus r -band absolute magnitude, for $z = 0.1$ EAGLE galaxies (coloured points). The rapid colour-dependent decline of galaxies fainter than $M_r - 5 \log_{10}(h) = -17$ results from imposing a stellar mass cut of $M_* > 10^{8.66} h^{-2} M_\odot$. The solid black line from Eq. (2) distinguishes red from blue galaxies, EAGLE galaxies above (below) this line are represented by a red (blue) dot. The black dashed-line is the corresponding colour-cut from Farrow et al. (2015) for GAMA galaxies (their Eq. 4, and Eq. 1 in the text)

initial conditions generated using the PANPHASIA multi-resolution phases of Jenkins & Booth (2013), taking the Planck Collaboration et al. (2014) cosmological parameter values. The simulation is part of the EAGLE simulation suite (Schaye et al. 2015; Crain et al. 2015) and Table 1 lists some of the key simulation parameters. The EAGLE simulations were performed with the GADGET-3 code, which is based on GADGET-2 (Springel 2005), and uses sub grid models for physical processes below the resolution limit. Sub grid parameters are calibrated to observations at $z \approx 0$ of the galaxy stellar mass function (e.g. Baldry et al. 2012), galaxy sizes (Shen et al. 2003), and the relation between black hole mass and stellar mass, as described in detail by Crain et al. (2015).

The hydrodynamics used in EAGLE uses a number of improvements to the SPH implementation collectively referred to as ANARCHY and described by Dalla Vecchia (*in prep.*); see Schaller et al. (2015) for a discussion of the relatively small impact of these changes on the properties of simulated galaxies. We briefly summarise below the sub grid modules for unresolved physics relevant for this paper:

- *Photo-heating and radiative cooling* by the optically thin evolving UV/X-ray background of Haardt & Madau (2001) is implemented element-by-element as described by Wiersma et al. (2009a).

- *Star formation* is implemented using the pressure law of Schaye & Dalla Vecchia (2008) and the metallicity-dependent star formation threshold of Schaye (2004). Gas particles eligible for star formation are converted to ‘star particles’ stochastically with a probability that depends on their star formation rate and their time step.

- *Stellar evolution and enrichment* is implemented as described by Wiersma et al. (2009b): assuming stars form with the Chabrier (2003) stellar initial mass function (IMF), spanning the range $[0.1, 100] M_\odot$, we use stellar evolution and yield tables to calculate the rate of type Ia and type II (core-collapse) supernovae, and follow the rate at which stars enrich the interstellar medium through AGB, type Ia and type II evolutionary channels.

- *Seeding, accretion and merging* of black holes is implemented following Springel et al. (2005); Booth & Schaye (2009), modified to account for the angular momentum of accreted gas as described by Rosas-Guevara et al. (2015).

- Thermal feedback from stars is implemented as described by Dalla Vecchia & Schaye (2012); feedback from accreting black holes is also implemented thermally.

- *Dark matter haloes* are identified using the friends-of-friends algorithm, with baryonic particles (gas, stars, and black holes) assigned to the same halo as the nearest dark matter particle, if any. The mass of the halo is characterised by $M_h \equiv M_{200,c}$, the mass enclosed within a sphere within which the mean density is 200 times the critical density.

- *Galaxies* are identified using the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009). To avoid including ‘intra-cluster’ mass/light to massive galaxies, we calculate (and quote) galaxy stellar masses/luminosities within 3D spherical apertures of 30 kpc. The aperture size was chosen to broadly approximate a Petrosian aperture, see Schaye et al. (2015) for details. We classify the galaxy that contains the particle with the lowest potential as the ‘central galaxy’, any other galaxies in the same halo is a ‘satellite’.

- *Broad-band absolute magnitudes* of galaxies are computed in the rest-frame, as described by Trayford et al. (2015): stellar emission is represented by the Bruzual & Charlot (2003) population synthesis models, with dust accounted for using the two-component screen model of Charlot & Fall (2000). Dust-screen optical depths depend on the mass of enriched, star-forming gas in galaxies and include an additional scatter to represent orientation effects. This constitutes the fiducial model of Trayford et al. (2015) (referred to as GD+O in that paper).

The analysis of the EAGLE simulations is greatly simplified by using the SQL database described by McAlpine et al. (2016) that contains all the properties of EAGLE galaxies used here. In particular, we extract position, velocity, stellar mass, star formation rate, and broad-band luminosities in a 30 kpc aperture for all galaxies from the database, and then convert them to h dependent units (such as h^{-1} Mpc for lengths and $h^{-2} M_\odot$ for stellar masses).

Galaxies in EAGLE show similar colour bi-modality as those in GAMA (Trayford et al. 2015, and Trayford *in prep.*): a blue cloud of star forming galaxies and a red sequence of

mostly passive galaxies. In the simulation, the appearance of passive red galaxies is due to the suppression of star formation in satellites and by feedback from supermassive black hole, as demonstrated by [Trayford et al. \(2016\)](#). The colour cut used by [Farrow et al. \(2015\)](#) to separate red from blue galaxies in GAMA is (their Eq. 4)

$$(g-r)_0 = 0.618 - 0.03 (M_{r,h} + 18.6), \quad (\text{GAMA red-blue cut}), \quad (1)$$

shown as a dashed black line on the rest-frame colour-magnitude plot of Fig. 1. Here we separate EAGLE galaxies in a red and blue population using

$$(g-r)_0 = 0.7 - 0.028 (M_{r,h} + 18.6) \quad (\text{EAGLE red-blue cut}), \quad (2)$$

shown as the black line in Fig. 1. This separates the passive from the active population well. The slope of observed and simulated colour cuts are virtually identical, and they are offset in colour by less than 0.1 magnitude at $M_{r,h} = -18.6$. This offset is comparable to the offset of ~ 0.15 (but in the opposite direction to the EAGLE offset) for the semi-analytical model considered by [Farrow et al. \(2015\)](#) (see their Fig. 2).

2.2 Dark Matter only simulations

We use two dark matter only simulations to study the impact of the limited simulation volume of the EAGLE simulation on clustering, one with the same volume and initial conditions as EAGLE (referred to as EAGLE-DMO below, this is the simulation L00100N1504-DMO described by [Schaye et al. 2015](#)) and one with a much larger simulation volume (referred to as P-MILLENNIUM below). Combined they allow us to assess to what extent missing large-scale power and sample variance affect inferences on clustering from the relatively small EAGLE volume.

Simulation EAGLE-DMO has the same volume, gravitational softening, cosmology and initial conditions as EAGLE. The masses of the dark matter particles are increased by a factor of $(\Omega_b + \Omega_{\text{dark}})/\Omega_{\text{dark}}$ compared to EAGLE, to account for not including the baryonic mass. We use this simulation to study clustering of haloes compared to other models - without results being affected by galaxy formation. The impact of baryonic effects on the density profiles of haloes in EAGLE was investigated by [Schaller et al. \(2015\)](#).

The P-MILLENNIUM simulation (see Table 1; [Baugh et al. in prep.](#) and [McCullagh et al. in prep.](#)) uses identical cosmological parameters as EAGLE but has a much larger volume (800^3Mpc^3 compared to the 100^3Mpc^3 for the EAGLE simulation used here). With P-MILLENNIUM we can quantify the effects of missing large-scale power and poor sampling of long wavelengths on clustering statistics by comparing the dark matter halo clustering in EAGLE-DMO with P-MILLENNIUM. For details about P-MILLENNIUM see [Baugh et al. in prep](#) and [McCullagh et al. in prep](#).

In both EAGLE-DMO and P-MILLENNIUM, dark matter haloes were identified using the friends-of-friends (FoF) algorithm with the standard value of $b = 0.2$ for the linking length in units of the mean particle separation. The mass of the halo is represented by $M_{200,c}$, the mass enclosed within a sphere with a density 200 times the critical density.

2.3 The GAMA survey

To put the EAGLE simulation results into context, we will compare them primarily to results from the GAMA survey data, and in particular to clustering measurements made by [Farrow et al. \(2015\)](#). The Galaxy And Mass Assembly (GAMA) survey ([Driver et al. 2011](#); [Liske et al. 2015](#)) is a spectroscopic and multi-wavelength survey of galaxies carried out on the Anglo-Australian telescope. In this work, we make use of the main r-band limited data from the GAMA equatorial regions ($\sim 180 \text{sq.deg.}$), which consists of a highly complete (> 98 per cent) spectroscopic catalogue of galaxies selected from the SDSS DR7 ([Abazajian et al. 2009](#)) to $r_{\text{petro}} < 19.8$. Further details of the GAMA survey input catalogue, tiling algorithm, redshifting and survey progress are described by [Baldry et al. \(2010\)](#), [Robotham et al. \(2010\)](#), [Baldry et al. \(2014\)](#) and [Liske et al. \(2015\)](#), respectively.

For the clustering comparisons presented in § 4, we are primarily interested in the following GAMA galaxy properties: r-band absolute magnitude, stellar mass and rest-frame $(g-r)_0$ colour. We describe in turn how each of those properties have been estimated in the clustering measurements of [Farrow et al. \(2015\)](#).

- *r-band absolute magnitude* ($M_{r,h}$): we apply evolution corrections and k-corrections to $z_{\text{ref}} = 0$ Petrosian r-band absolute magnitudes, where $M_{r,h} \equiv M_r - 5 \log_{10}(h)$. For further details, see § 2.1.4 of [Farrow et al. \(2015\)](#).

- *$(g-r)_0$ colour*: the rest-frame colours are derived from SDSS model magnitudes, with colour and redshift dependent k-corrections as per [Loveday et al. \(2012\)](#) and [McNaught-Roberts et al. \(2014\)](#). For further details, see § 2.1.4 of [Farrow et al. \(2015\)](#).

- *stellar mass* (M_* in units of $h^{-2} \text{M}_{\odot}$): the clustering measurements in [Farrow et al. \(2015\)](#) use the relation between rest-frame $(g-i)_0$ colour and stellar mass as derived by [Taylor et al. \(2011\)](#), using the [Bruzual & Charlot \(2003\)](#) synthetic stellar population models with a [Calzetti et al. \(2000\)](#) dust attenuation law to correct for dust in the Milky Way. For further details, see § 2.1.5 of [Farrow et al. \(2015\)](#) and [Taylor et al. \(2011\)](#).

For completeness, we refer the reader to § 3 of [Farrow et al. \(2015\)](#) for the modelling of the GAMA selection function. Accurate modelling of the selection function of a redshift survey is key for precise clustering measurements. The modelling approach described in [Farrow et al. \(2015\)](#), which is based on the method of [Cole \(2011\)](#), enables a uniform modelling for all galaxy samples split by stellar mass, luminosity and colour.

3 GALAXY CLUSTERING

In this section we present the analysis methods used, starting with the estimators we use for calculating the two-point correlation function and its associate errors. We then briefly discuss how we compute the effective bias.

3.1 The two-point correlation function

The spherically averaged two-point correlation function, $\xi(r)$, defined as (e.g. Peebles 1980)

$$\xi(r) = \frac{1}{\langle n \rangle} \frac{dP}{dV} - 1, \quad (3)$$

provides a statistical description of a sample's spatial distribution. Here, dP/dV is the probability of finding a galaxy in the volume dV at a given (co-moving) distance r from another galaxy, and $\langle n \rangle$ is the mean (co-moving) number density of galaxies. In practice for a volume with periodic boundary conditions, the correlation function can be estimated by counting the number of pairs of galaxies, $N_s(r)$, in a shell of volume $V_s(r)$ at distance r from each other using (e.g. Rivolo 1986)

$$\xi(r) = \frac{1}{\langle n \rangle^2 V} \frac{N_s(r)}{V_s(r)} - 1, \quad (4)$$

where V is the total volume of the periodic simulation.

However, when the volume has boundaries (as is the case for any observed survey, or when we want to restrict the analysis to a small region within a larger periodic simulation volume), these equations cannot be used. In such cases, ξ can be computed by comparing the distribution of galaxy pairs to the clustering of a set of points uniformly distributed within the survey volume, using e.g. the Landy & Szalay (1993) estimator:

$$\xi(r) = \frac{DD - 2DR + RR}{RR}, \quad (5)$$

where DD is the suitably normalised number of galaxy pairs at a distance r from each other, RR the corresponding normalised number of pairs from the random distribution, and DR the suitably normalised numbers of galaxies and random pairs separated by distance r .

The co-moving distance between galaxies cannot be measured directly from a redshift survey due to galaxy peculiar velocities and large-scale redshift space distortions. However, by splitting the information into projected separation, r_p , and distance parallel to the line of sight, π , one can estimate the 2D correlation function, $\xi(r_p, \pi)$, which in turn is used to estimate the projected correlation function,

$$w_p(r_p) = 2 \int_0^{\pi_{\max}} \xi(r_p, \pi) d\pi, \quad (6)$$

with π_{\max} set to a value adequate for the sample considered (here π_{\max} is fixed to ~ 34 Mpc/ h , which represents $\sim L/2$ of EAGLE simulation. See Table 1). We select π_{\max} to be sufficiently large to account for most redshift space distortions, while keeping it small enough to not add noise to our estimate of the projected correlation function. In addition the π_{\max} value chosen is in line with what is commonly used in observational clustering measurements. To a very good approximation, $w_p(r_p)$ is independent of redshift space effects, making this statistic ideal for model comparisons. We compute $w_p(r_p)$ along three orthogonal directions in the simulations, to improve the signal-to-noise of the clustering measurements.

To reduce the dynamical range when plotting w_p , we will often divide by the projected correlation function of the reference power law, $\xi(r) = (r_0/r)^\gamma$ with $r_0 = 5.33h^{-1}$ Mpc and $\gamma = 1.8$, where the constants are from the fit function

that corresponds to this power-law is

$$w_p^{\text{ref}}(r_p) = r_p \left(\frac{r_0}{r_p} \right)^\gamma \frac{\Gamma(1/2)\Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)}, \quad (7)$$

where Γ denotes the Gamma function.

3.2 Error estimates on clustering statistics

We compute and quote jackknife errors on the simulated two-point correlation function in EAGLE to mimic observational errors. However, sample variance is likely to dominate the error budget. We estimate sample variance by subsampling EAGLE sized-volumes in the P-MILLENNIUM simulation, which also allows us to examine any effects due to missing large-scale power. Unfortunately we can only estimate these errors for the clustering of haloes - not galaxies - since the P-MILLENNIUM simulation is dark matter only.

We apply the jackknife technique by partitioning the EAGLE (or EAGLE-DMO) simulation volume in N_{sub} tiles of equal volume, with $N_{\text{sub}} = 8$. We then compute the two-point correlation function ξ_k^{JK} by omitting the k -th tile, and compute the variance

$$\sigma^2(r) = \frac{(N_{\text{sub}} - 1)}{N_{\text{sub}}} \sum_{k=1}^{N_{\text{sub}}} (\xi_k^{JK}(r) - \xi^{\text{tot}}(r))^2, \quad (8)$$

where $\xi^{\text{tot}}(r)$ is the correlation function of the total volume. Such jackknife error estimates have been used extensively to estimate errors in galaxy clustering (e.g., Zehavi et al. 2002, 2011; Favole et al. 2016), and Zehavi et al. (2002) shows that such errors accurately reflect uncertainties in the clustering on the scales investigated here. However, it has of course its limitations, as pointed out by e.g. Norberg et al. (2009). For example the technique may underestimate errors when a few systems dominate the signal. We will see below this is in fact the case in EAGLE, where the clustering of low-mass red galaxies on small scales is dominated by satellites in a few massive clusters, as we illustrate in Fig. 7 below.

We estimate errors on the clustering of haloes due to sample variance and missing large-scale power on EAGLE volumes using the P-MILLENNIUM simulation as follows. We partition the P-MILLENNIUM simulation in $N_{\text{sub}} = 512$ tiles of volume equal to that of EAGLE. We then calculate the correlation function of haloes in the i -th tile, $\xi_i(r)$, using Eq. (5), as well as the correlation function of the total volume, $\bar{\xi}(r)$. The variance is then calculated as

$$\sigma^2(r) = \frac{1}{N_{\text{sub}} - 1} \sum_{i=0}^{N_{\text{sub}}} (\xi_i(r) - \bar{\xi}(r))^2. \quad (9)$$

We use the mean number of haloes in the total volume to compute $\xi_i(r)$ for each tile, rather than the mean number in that tile, to better mimic the observed analysis and that applied to EAGLE. This has a non-negligible effect on the correlation functions, as evidenced in § 5.

3.3 Effective bias

The bias, b , of a tracer population is the ratio of the correlation function of that tracer over that of the mass (e.g. Davis et al. 1985), with the effective bias, $b_{\text{eff}}(X)$, given by (e.g. Porciani et al. 2004)

$$b_{\text{eff}}(X) = \frac{\int b(M_h) N_{\text{gal}}(M_h, X) n(M_h) dM_h}{\int N_{\text{gal}}(M_h, X) n(M_h) dM_h}, \quad (10)$$

where $n(M_h)$ is the halo mass function (the number density of haloes of mass M_h), $b(M_h)$ the linear bias factor of haloes of mass M_h , and $N_{\text{gal}}(M_h, X)$ the mean number of galaxies of property X in haloes of mass M_h (the mean halo occupation). The property X could select galaxies in a given stellar mass, colour or luminosity range, for example. We approximate the integral as a sum over all haloes in the simulation, obtaining

$$b_{\text{eff}}(X) = \frac{\sum_{i=0}^{N_{\text{haloes}}} b(M_h^i) N_{\text{gal}}(M_h^i, X)}{\sum_{i=0}^{N_{\text{haloes}}} N_{\text{gal}}(M_h^i, X)}, \quad (11)$$

where $N_{\text{gal}}(M_h^i, X)$ is the number of galaxies with property X in a halo of mass M_h^i . In practice we estimate the effective bias of samples split by stellar mass and hence evaluate this sum for all galaxies in narrow stellar mass bins.

To estimate the linear halo bias $b(M_h)$, we follow Mo & White (1996),

$$b(M_h) = 1 + (\nu(M_h)^2 - 1)/\delta_c, \quad (12)$$

where $\delta_c = 1.686$ is the spherical collapse density threshold, and $\nu(M_h) = \delta_c/\sigma(M_h)$ the dimensionless amplitude of fluctuations that produce haloes of mass M_h (at a given redshift z). The (linear) matter variance, $\sigma(M_h)$ at a given z , can be computed numerically for a given linear power spectrum (see Eq. (9) in Murray et al. 2013) using the web-portal HMFcalc¹ (and adopting the spectral index $n_s = 0.9611$ used in EAGLE). Finally, we adopt the fit provided by Jenkins et al. (2001) for the halo mass function, $n(M_h)$, which provides a good description of the halo mass function of all simulations used here.

4 RESULTS

We begin by considering the clustering of dark matter haloes in EAGLE, followed by a quick look at the real and redshift space clustering of EAGLE galaxies with well resolved stellar mass. In §4.3 we present the main results of this study, namely the clustering of galaxies in EAGLE compared to that of GAMA, when split by stellar mass, luminosity, colour or star-formation rate.

4.1 Halo clustering in EAGLE

The real-space clustering of haloes with $M_h > 10^{12} h^{-1} M_\odot$ in the P-MILLENNIUM dark matter simulation is plotted in Fig. 2. We sub-divide the volume of this simulation into 512 non-overlapping tiles, each with the same volume as EAGLE-DMO, and compute the correlation function $\xi_i(r)$ for each of the tiles, but using the overall mean number density of haloes in the full volume in Eq.(5) (rather than the mean in each tile i). We plot the mean correlation function averaged over all tiles, $\bar{\xi}(r)$, as the black line, with the scatter around the mean shown as 1σ error bars. The mean correlation function $\bar{\xi}(r)$ follows the correlation function of the full

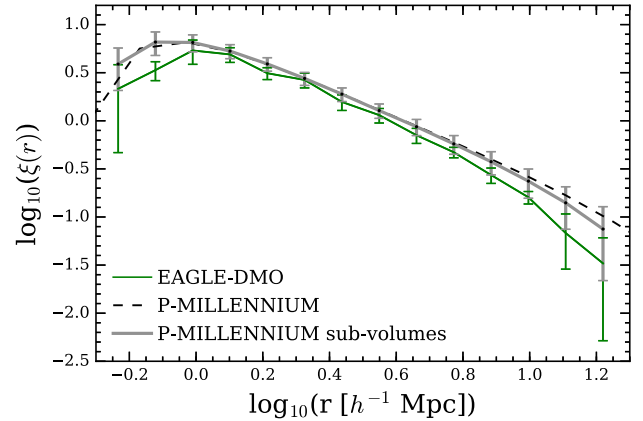


Figure 2. The $z = 0$ real space two-point correlation function of haloes with mass $M_h > 10^{12} M_\odot h^{-1}$ from P-MILLENNIUM (black dashed line), the mean correlation function, $\bar{\xi}(r)$, of the 512 non-overlapping tiles of that simulation (grey line), and the 1σ scatter around this mean (grey error bars). The green curve is the correlation function for simulation EAGLE-DMO with jackknife error bars. Finite-volume effects cause the green curve to fall increasingly below the grey line, with jackknife errors yielding nevertheless a realistic error estimate.

volume (grey dashed line) very closely, falling well within the scatter between volumes, as expected.

The correlation function of EAGLE-DMO (green line) falls below $\bar{\xi}(r)$ on scales smaller than $1 h^{-1}$ Mpc, remains well within sample variance up to scales $5 - 6 h^{-1}$ Mpc ($\log r/(h^{-1} \text{Mpc}) = 0.7 - 0.78$), then falls increasingly below $\bar{\xi}(r)$ above this scale. As both simulations have identical power-spectra and cosmological parameters, the deviations are due to sample variance and to the integral constraint on ξ . We note that numerical resolution is not likely to play a role in the apparent differences in clustering, as these haloes are resolved by $\approx 10^4$ particles or more.

We compute jackknife errors for EAGLE-DMO, as described above, and plot them in green. Of course these do not quantify finite-volume effects nor sample variance, but nevertheless the green and black curves are within the EAGLE-DMO jackknife errors. This motivates us to use such errors when calculating errors on the correlation function of galaxies, rather than haloes, below, since we do not have access to larger hydrodynamical simulations to estimate finite-volume effects. Given the level of convergence between EAGLE-DMO and P-MILLENNIUM, we will plot correlation functions up to scales of $10 h^{-1}$ Mpc (14 per cent of the full extent), with Fig. 2 quantifying the limitations on halo clustering.

4.2 Galaxy clustering in EAGLE

The real-space correlation function, $\xi(r)$, of galaxies with stellar mass $M_* > 10^{8.83} h^{-2} M_\odot$ from EAGLE is plotted in Fig. 3 (top panel). Distinguishing between central galaxies (typically but not necessarily the most massive galaxy in a given halo) and satellites galaxies, we compute the one- and two-halo contributions separately (grey dashed and grey dotted lines, respectively). The contribution to ξ from these is equal at a separation of approximately $r = 1.3 h^{-1}$ Mpc. It is important to note that, while the two-point correlation

¹ <http://hmf.icrar.org>

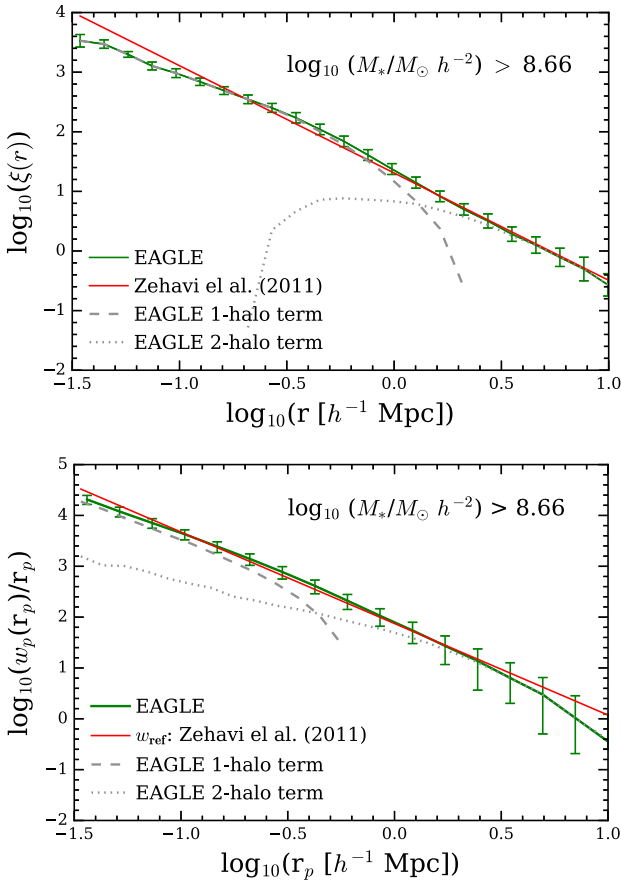


Figure 3. *Top panel:* the $z = 0.1$ two-point correlation function for EAGLE galaxies with $M_* > 10^{8.66} h^{-2} M_\odot$ (green curve with jackknife errors bars), decomposed in the one- and two-halo terms (dashed and dotted grey lines, respectively). The red line is the reference power law model for galaxies with $-21.0 < M_{r,h} < -20.0$ from the fit by Zehavi et al. 2011. *Bottom panel:* corresponding projected correlation function, $w_p(r_p)$ from Eq. (6). The red line is the projected correlation function from the fit by Zehavi et al. 2011, w_p^{ref} from Eq. (7). The galaxy selection is different in detail for EAGLE and the observations: the red lines are shown to guide the eye.

function from the dark matter haloes of EAGLE-DMO is underestimated at distances below $r \sim 1 h^{-1} \text{Mpc}$, this does not imply that the same is true for the galaxy correlation function, since there are of course generally many galaxies per halo.

The real-space correlation function quantifies the physical clustering of galaxies, independently of any peculiar velocities. However, observations can only measure clustering in redshift space, and peculiar velocities then distort the signal. To ameliorate the effects of such redshift space distortion, it is convenient to integrate $\xi(r_p, \pi)$ over a narrow range in π , and compute the projected correlation function $w_p(r_p)$, see Eq. (6). This is plotted for the same EAGLE galaxies (those with $M_* > 10^{8.83} h^{-2} M_\odot$) in the bottom panel of Fig. 3, again plotting one- and two-halo terms as well.

The reference model $w_p^{\text{ref}}(r_p)$ of Eq. (7) provides a relatively good fit to EAGLE’s projected correlation function. We note, however, that the galaxy selections differ between EAGLE and the SDSS galaxies fit by Zehavi et al. (2011) that

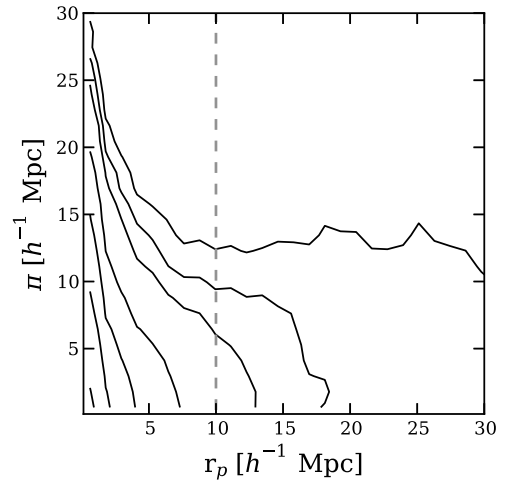


Figure 4. The $z = 0.1$ two-dimensional redshift space correlation function $\xi(r_p, \pi)$, as function of projected separation r_p and line-of-sight separation π , for EAGLE galaxies with stellar masses greater than $10^{8.66} h^{-2} M_\odot$. Contours levels correspond to $\xi(r_p, \pi) = 5, 2, 1, 0.5, 0.2, 0.1, 0.01$. The vertical dashed line corresponds to projected separations of $r_p = 10 h^{-1} \text{Mpc}$ beyond which the clustering of haloes is increasingly suppressed due to the limited extent of the EAGLE simulation (see Fig. 2).

gives rise to w_p^{ref} – the two therefore did not have to agree: we show the comparison to guide the eye, and because we use w_p^{ref} as a normalisation below.

The contribution of one- and two-halo terms to w_p are equal at a projected separation of $r_p \sim 0.4 h^{-1} \text{Mpc}$. Comparing top and bottom panel from Fig. 3, it is clear that the two-halo term contributes more to the projected correlation function on scales comparable to the virial radius of haloes than it does to the two-point correlation function: at a scale of $\approx 0.1 h^{-1} \text{Mpc}$, the two-halo term contributes nearly 10 per cent to w_p .

The two-dimensional redshift space correlation function of this stellar mass limited sample of EAGLE galaxies is plotted in Fig. 4 in terms of the projected separation, r_p , and line of sight separation, π . It exhibits the familiar elongation in the π direction at small r_p resulting from virial motion of galaxies in haloes (the ‘fingers-of-god’ effect), and the flattening in the π direction at large r_p , due to coherent streaming motions of galaxies into haloes and out of voids (the ‘Kaiser’ effect, Kaiser (1987)). We do not compare this correlation function directly to GAMA, mainly because of the complexity of making sure the selection of galaxies in the π direction is the same in simulation and data.

4.3 Galaxy clustering in EAGLE compared to GAMA

We use the volume-limited samples of GAMA galaxies presented by Farrow et al. (2015), which we can split by stellar mass, luminosity or colour. For samples split in bins of stellar mass or luminosity only, we refer the reader to Table 2 of Farrow et al. (2015), while we present in Table 2 the properties of the additional GAMA samples used in here, for which

we have computed clustering statistics following the methods outlined by Farrow et al. (2015)².

Throughout this section, EAGLE galaxies are selected from the $z = 0.1$ snapshot (a redshift close to the median redshift of the GAMA samples). Some statistics of the EAGLE samples used are provided in Table 3. By construction, the stellar mass functions from EAGLE and GAMA show good agreement in the stellar mass range we are analysing. Hence, the mean number density of galaxies in each bin of stellar mass, agree well between data and simulation. Therefore we do not need to compare clustering at given number density, but directly compare samples selected by stellar mass - or indeed luminosity.

4.3.1 Stellar mass dependent clustering

A comparison of the clustering of EAGLE galaxies to that in GAMA as a function of stellar mass is shown in Fig. 5, using the same mass as used by Farrow et al. (2015). The bottom panels of Fig. 5 highlight the differences between the two measurements, by presenting the ratio of the projected correlation functions with respect to the reference power law model adopted (following Farrow et al. 2015). The projected correlation function of EAGLE galaxies (green lines with jackknife errors) is in remarkably good agreement with the GAMA data (solid black lines, with $1\text{-}\sigma$ uncertainty range shown as a grey shaded area): the deviations are typically within the measured uncertainty range.

It is well known (and clear from the figure) that the clustering strength of galaxies increases with stellar mass. There is little or no evidence for such a trend in the simulation, however the errors are relatively large and the simulation is not inconsistent with such a trend either. We re-iterate that the jackknife errors do not capture sample variance on the scale of the sample, and any missing large scale power likely affects the overall normalisation of $w_p(r_p)$ rather than the shape. The good agreement in shape of the correlation functions of GAMA and EAGLE is therefore encouraging, but the simulated volume too small to detect a clear dependence of w_p on M_* .

The calibration of sub-grid parameters in EAGLE was based on one-point statistics, as described in Section 2. The clustering of galaxies is therefore a genuine model prediction. The good agreement then implies that EAGLE galaxies tend to inhabit haloes in a way that mimics accurately the way GAMA galaxies do. Finally we note that the decrease of the clustering signal on scales greater than $5 h^{-1}\text{Mpc}$ scales in EAGLE is related to the limited simulation box size - and mimics the corresponding fall in clustering of EAGLE haloes.

4.3.2 Luminosity dependent clustering

A comparison of the clustering of EAGLE galaxies to that in GAMA as a function of r -band luminosity, is shown in Fig. 6. The agreement is very good, and well within the relatively large jackknife error estimates. Similar to the case of clustering as function of mass, the amplitude of observed

clustering increases with luminosity (see e.g. Norberg et al. 2001; Zehavi et al. 2005), but again there is little or no evidence for such a trend in EAGLE. As in the previous section, we suggest this is mostly due to the finite volume of the simulation: more luminous galaxies are biased to more massive haloes, of which there are relatively few in the EAGLE volume, and their clustering is underestimated because of lack of large-scale power. The reference power law w_p^{ref} from Eq. (7), describes the clustering of EAGLE galaxies in the middle panel of Fig. 6 very well. In this panel, EAGLE galaxies are selected in the same way, $-21.0 < M_{r,h} < -20.0$, as in the sample of Zehavi et al. (2011) to which w_p^{ref} was fit, so they can be compared directly. The good agreement in clustering, combined with the fact that EAGLE also fits the galaxy luminosity function well (Trayford et al. 2015), implies that EAGLE galaxies form in similar haloes, and have similar stellar populations and star formation histories as those in GAMA. This encourages us to look in clustering as function of galaxy colour in more detail next.

4.3.3 Colour dependent clustering

Farrow et al. (2015) present r -band magnitude limited samples of GAMA galaxies split by rest-frame $(g-r)_0$ colour and in bins of stellar mass. Galaxies are classified as ‘red’ or ‘blue’ using the r -band magnitude colour cut of Eq. (1). We use this data to compute w_p for these galaxies using the method described by Farrow et al. (2015). Some statistics of these samples are summarised in Table 2. We note that the lowest mass bin in red galaxies, $9.5 < \log_{10} M_*/h^{-2} M_\odot < 10.0$, is only partially volume limited, as ~ 4 per cent of the galaxies in that sample are only volume limited over a smaller redshift range than the one nominally considered. Given the small galaxy fraction affected by this, we can consider this sample still to be volume limited when computing clustering statistics. The advantages of keeping the exact same volumes for all stellar mass samples split by colour (hence they all sample the same underlying large scale structure) overcomes this minor subtlety, which is primarily driven by uncertainties in the measured colours and adopted k -corrections for a small subset of galaxies.

We then select EAGLE galaxies in the same way, using identical bins in colour and stellar mass, but applying the (almost identical) colour cut to split red and blue galaxies of Eq. (2). Some statistics of the EAGLE galaxies are summarised in Table 3 which includes the fraction of EAGLE galaxies that are satellites.

The projected correlation functions for red and blue galaxies, split in bins of stellar mass, is plotted in Fig. 7. To ease the interpretation of the EAGLE clustering results, we show the correlation function for all central and satellite galaxies (i.e. irrespective of colour) within each stellar mass bin as green dotted and dashed lines respectively. Therefore, the central/satellite clustering results are the same in the left and right panels, but vary with stellar mass (from top to bottom).

In EAGLE, red galaxies (left column) cluster more strongly than blue galaxies (right panel) of given mass, and similarly satellite galaxies (dashed lines) cluster more strongly than centrals (dotted lines), in all stellar mass ranges studied. It is also apparent that the red population follows closely the clustering of satellites, in particular

² Farrow et al. (2015) uses a flat $\Omega_m = 0.25$ cosmology to infer distances for their clustering measurements. On the scales considered, this difference in cosmology is totally negligible.

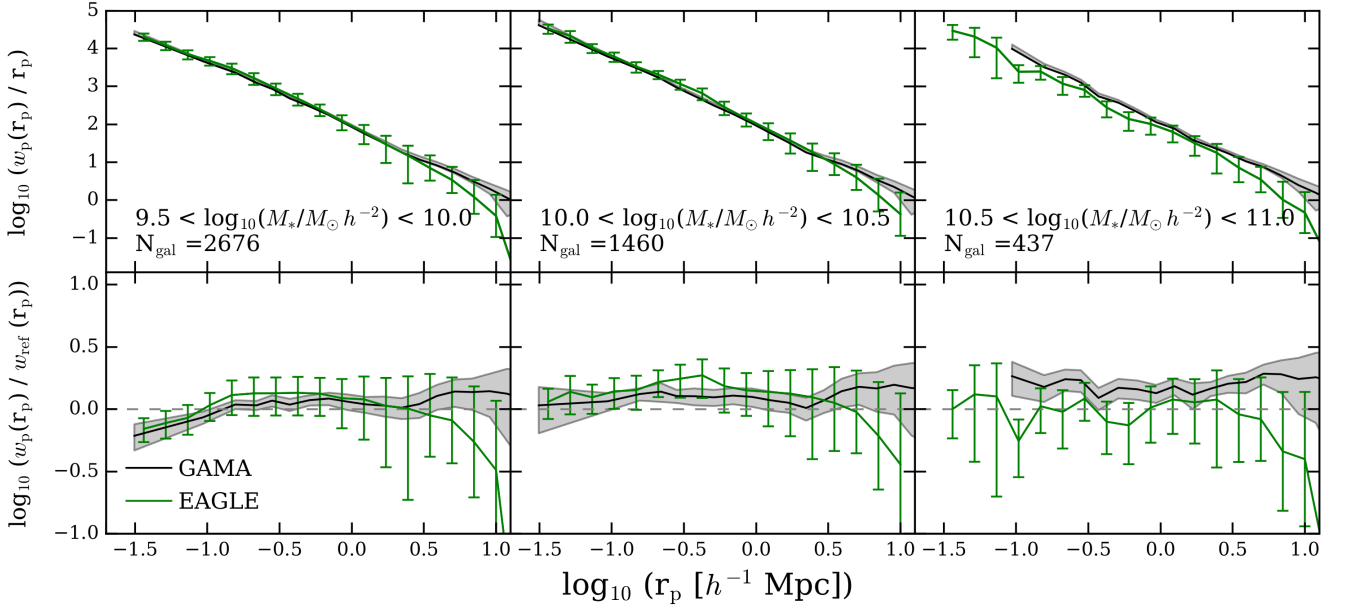


Figure 5. Clustering as a function of stellar mass; the mass bin is indicated in each column. *Top panels:* Projected correlation function $w_p(r_p)$ from Eq. (6). *Green curve* is the EAGLE result at $z = 0.1$, with jackknife error bars; N_{gal} is the number of EAGLE galaxies in each mass bin. *Black curve* is GAMA result, with the grey shading including the $1\text{-}\sigma$ error range. *Bottom panels:* Same as top panels, but the correlation functions are divided by the reference function $w_p^{\text{ref}}(r_p)$ from Eq. (7). The dashed line indicates where the ratio is unity.

Table 2. Statistics of GAMA galaxy samples (mostly volume limited), that are not already described in Table 2 of Farrow et al. (2015), following a similar table structure. The stellar mass restricted samples are split into red and blue galaxies, presented in turn. Columns from left to right are: stellar mass range, minimum and maximum sample redshift, total number of galaxies, the galaxy number density, median sample redshift, median r -band absolute magnitude, median stellar mass, median $(g-r)_0$ rest-frame colour, and the fraction of truly volume limited galaxies (see text for further details).

stellar mass range	z_{min}	z_{max}	N_{gals}	\bar{n} (Mpc/h) $^{-3}$	z_{med}	$M_{r,h}^{\text{med}}$	$\log_{10} M_{*}^{\text{med}}$ ($h^{-2} M_{\odot}$)	$(g-r)_0^{\text{med}}$	f_{vlim}
Red									
$9.5 < \log_{10} M_{*}/h^{-2} M_{\odot} < 10.0$	0.02	0.14	5407	4.45×10^{-3}	0.11	-19.16 (0.38)	9.77 (0.14)	0.72 (0.07)	0.96
$10.0 < \log_{10} M_{*}/h^{-2} M_{\odot} < 10.5$	0.02	0.14	5527	4.55×10^{-3}	0.11	-20.16 (0.40)	10.23 (0.14)	0.75 (0.10)	1.00
$10.5 < \log_{10} M_{*}/h^{-2} M_{\odot} < 11.0$	0.02	0.14	1945	1.60×10^{-3}	0.12	-21.11 (0.37)	10.65 (0.12)	0.77 (0.09)	1.00
Blue									
$9.5 < \log_{10} M_{*}/h^{-2} M_{\odot} < 10.0$	0.02	0.14	4663	3.84×10^{-3}	0.11	-19.71 (0.39)	9.71 (0.14)	0.54 (0.08)	1.00
$10.0 < \log_{10} M_{*}/h^{-2} M_{\odot} < 10.5$	0.02	0.14	1870	1.54×10^{-3}	0.12	-20.61 (0.34)	10.18 (0.13)	0.61 (0.06)	1.00
$10.5 < \log_{10} M_{*}/h^{-2} M_{\odot} < 11.0$	0.02	0.14	248	0.20×10^{-3}	0.12	-21.44 (0.34)	10.61 (0.11)	0.67 (0.06)	1.00

Table 3. Statistics of EAGLE stellar mass selected samples from the $z = 0.1$ snapshot. Columns from left to right are: stellar mass range, number of galaxies, galaxy number density, fraction of satellites, fraction of blue galaxies (following Eq. 2), the SFR limits used to define ‘low’ and ‘high’ SFR galaxy samples respectively, number density of red and blue galaxies, respectively.

Sample stellar mass range	N_{gals}	\bar{n} (Mpc/h) $^{-3}$	f_{sat}	f_{blue}	SFR_{low} $M_{\odot} \text{ yr}^{-1}$	SFR_{high} $M_{\odot} \text{ yr}^{-1}$	\bar{n}_{red} (Mpc/h) $^{-3}$	\bar{n}_{blue} (Mpc/h) $^{-3}$
$9.5 < \log_{10} M_{*}/h^{-2} M_{\odot} < 10.0$	2676	8.6×10^{-3}	0.43	0.82	< 0.28	$1.02 <$	1.5×10^{-3}	7.1×10^{-3}
$10.0 < \log_{10} M_{*}/h^{-2} M_{\odot} < 10.5$	1460	4.7×10^{-3}	0.39	0.77	< 0.26	$1.99 <$	1.2×10^{-3}	3.6×10^{-3}
$10.5 < \log_{10} M_{*}/h^{-2} M_{\odot} < 11.0$	437	1.4×10^{-3}	0.22	0.81	< 0.65	$3.43 <$	2.7×10^{-4}	1.1×10^{-3}

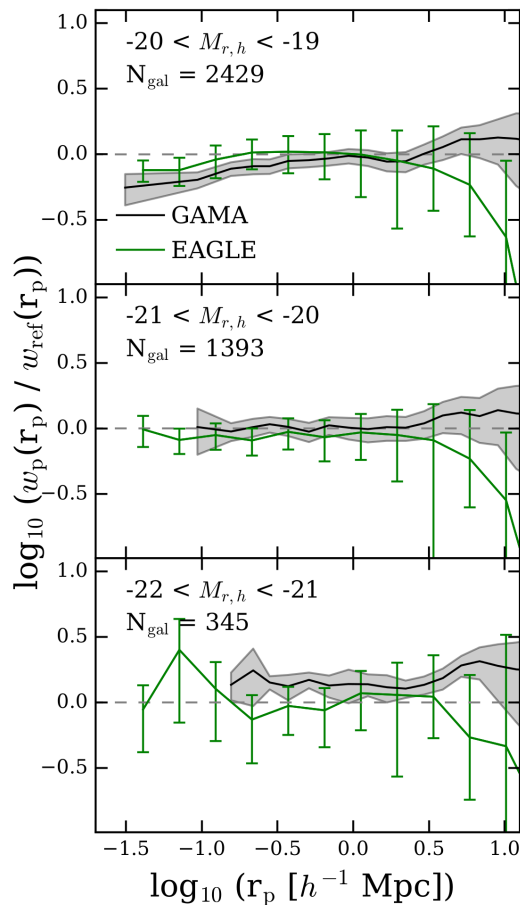


Figure 6. Same as the bottom panels of Fig. 5, but for clustering as a function of r -band luminosity. Ranges of r -band absolute magnitudes are shown from faint to bright (top to bottom panels), with the number of EAGLE galaxies in this luminosity range labelled in each panel.

for galaxies with stellar masses greater than $10^{10} h^{-2} M_{\odot}$. In contrast, blue galaxies follow more closely the clustering of centrals, again particularly so for the two more massive galaxy stellar bins.

The clustering of blue EAGLE galaxies tracks that of blue GAMA galaxies well, in particular on scales up to $r_p \sim 4 h^{-1} \text{Mpc}$ ($\log_{10}(r_p) = 0.6$). However, red EAGLE galaxies cluster noticeably more strongly than red GAMA galaxies. As a consequence, the trend that red galaxies cluster more strongly than blue galaxies of given mass, which is clearly present in GAMA, is too strong in EAGLE. Trayford et al. (2016) already noticed that red galaxies are overabundant in EAGLE, and that this is at least partly due to lack of numerical resolution in that simulation (see the Appendix of Trayford et al. 2016). The strong clustering of red galaxies in EAGLE is significantly influenced by the presence of a few massive haloes. To demonstrate this, we re-compute w_p for red EAGLE galaxies after excluding all galaxies in the three most massive haloes. We show the result by the orange line in Fig. 7. The overall amplitude of the clustering signal of red galaxies is dramatically reduced. A larger simulation box is likely required to provide detailed insight on the clustering of red massive galaxies.

The colour dependences of the clustering of EAGLE galaxies of given stellar mass is clearly partly due to the relative fractions of satellite and central galaxies. In Table 3 we show the fraction of satellites and blue galaxies for each stellar mass range. We find that the fraction of satellite galaxies decreases at higher masses, while the fraction of red and blue satellites is not strongly dependent on stellar mass. For example, in the stellar mass range $9.5 < \log M_*/M_{\odot} h^{-2} < 10.0$ we find that 43% of the galaxies are satellites, of which $\sim 19\%$ are red and the remaining $\sim 81\%$ blue. That table also shows that the vast majority of galaxies are blue, and most of these are centrals. Therefore the good agreement with GAMA in the clustering of blue galaxies is consistent with the fact that EAGLE galaxies of given mass cluster similarly to GAMA galaxies, as shown in Fig. 5. A relatively modest improvement of the numerical resolution of the simulation could be enough to reproduce the clustering of red galaxies equally well. However the difference seen could equally be due to the missing large-scale power and the impact the few rare objects have in the 100 Mpc EAGLE volume.

4.3.4 Star formation rate dependent clustering

We divide EAGLE galaxies of a given stellar mass in three bins of star formation rate (SFR), the 30 per cent with the lowest star formation rate (‘low SFR’), the 30 per cent with the highest star formation rate (‘high SFR’), and the remainder (‘int SFR’). Their galaxy clustering is plotted in Fig. 8. The low SFR galaxies are clustered most strongly, which is particularly evident on the smaller scales, and this is the case in all stellar mass bins investigated. The difference in the amplitude of clustering between the high and intermediate SFR galaxies is not very large, with high SFR galaxies generally clustering least. Jackknife errors bars are not plotted to avoid clutter, but should be of order ± 0.35 dex (i.e. a factor $\sqrt{3}$ higher than the typical ± 0.2 dex errors seen in Fig. 5 below $r_p = 1 h^{-1} \text{Mpc}$). Any difference in clustering between the blue (‘high SFR’) and red (‘int SFR’) curves are therefore not very significant. At the largest scales shown, the amplitude of clustering for all populations becomes similar, with any difference now much smaller than jackknife errors.

We do not compare these clustering measurements to those from GAMA presented by Gunawardhana et al. (*in prep*), for the following main reason: the SFR and stellar mass range probed by GAMA and EAGLE are in detail poorly matched, in the sense that the limitations in each of the GAMA and EAGLE star-forming samples are hard to account for all at the same time. For GAMA, the SFR is only measurable for a galaxies with sufficiently high SFR, resulting in volume limited samples with SFR and stellar mass ranges restricted to $\text{SFR} \gtrsim 0.3 M_{\odot} \text{yr}^{-1}$ and $M_* \gtrsim 10^{9.5} h^{-2} M_{\odot}$. (see Gunawardhana et al. (*in prep.*) for details). Hence for a detailed comparison to take place, it would be necessary to work with samples defined by absolute cuts in SFR. This in turn requires the GAMA SFR measurements to be directly compatible with those in EAGLE, as defining samples by SFR ranking is not possible. As the EAGLE number densities are between 60 and 70 per cent larger than the GAMA number densities for the same stellar mass range and SFR constraints a detailed clustering comparison becomes futile:

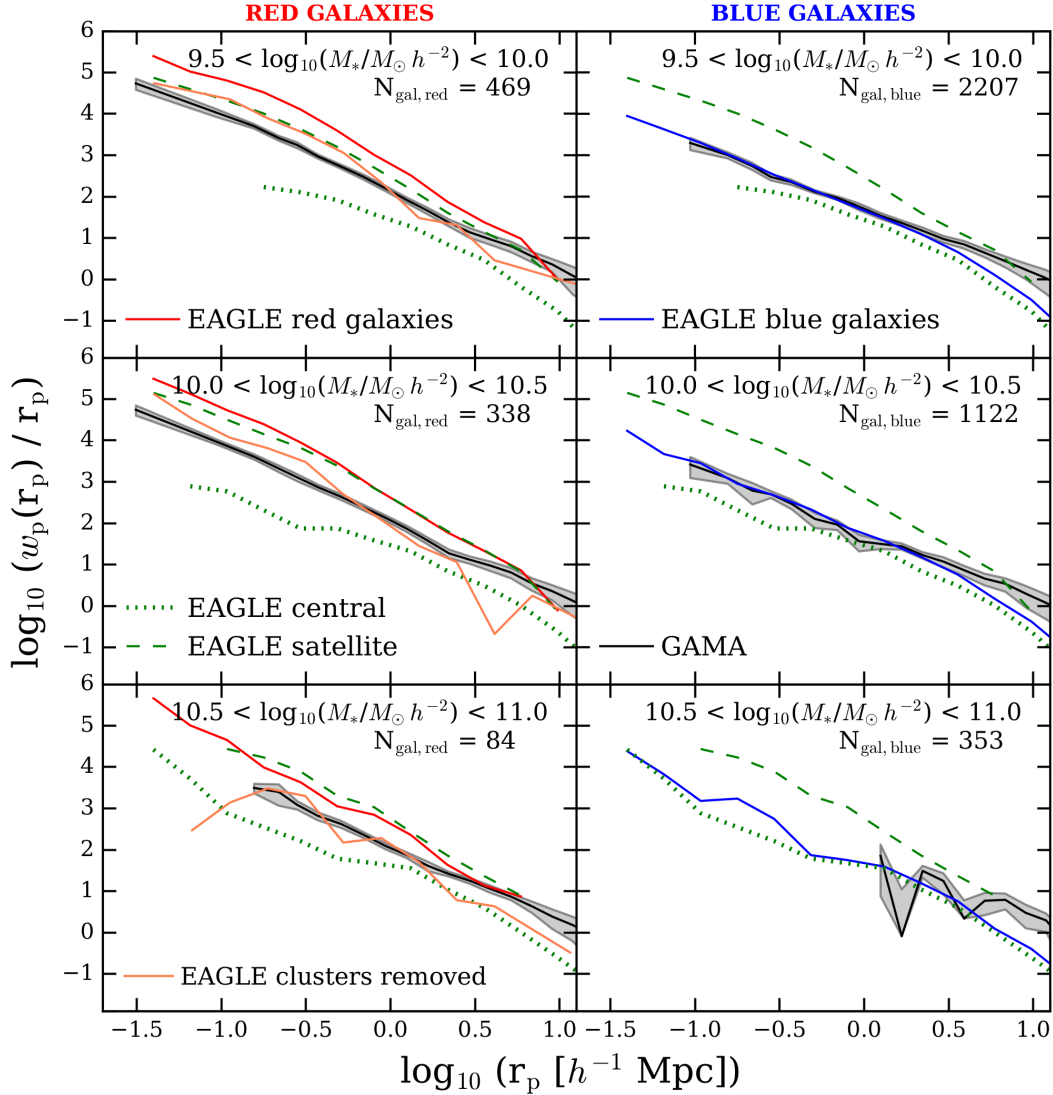


Figure 7. Clustering as function of $(g-r)_0$ colour for red and blue galaxies (left and right column, respectively), in bins of stellar mass (top to bottom rows). Limits of each mass bin are labelled in each panel. The projected correlation function for EAGLE galaxies in each mass bin is shown by the *red* and *blue* curves for red and blue galaxies, respectively, N_{gal} is the number of EAGLE galaxies that contributes to the calculation. The corresponding clustering of GAMA galaxies is shown by the *black* line with *grey* shaded region the $1\text{-}\sigma$ error range. The clustering of central and satellite EAGLE galaxies split by mass but not by colour, is plotted as *dotted green* and *dashed green* lines, respectively. In the left column, the *orange* lines are w_p of red EAGLE galaxies, after excluding the three most massive haloes.

any difference observed in the clustering can be attributed to the difference in number densities. This is in agreement with the results of Furlong et al. (2015), who pointed out that the specific SFR in EAGLE are typically 0.2 to 0.5 dex lower than observed. A proper understanding of those SFR difference between data and simulations is required before a detailed and informative clustering comparison of SFR selected samples can be made.

The differences in clustering of galaxies of the same mass that are satellites versus centrals (Fig. 7), or red versus blue galaxies (Fig. 8), should be dependent on the mass of the halo they inhabit, with more strongly clustered galaxies residing in more massive haloes. To verify that this is the case in EAGLE, we plot the halo occupation distribution - the fraction of galaxies of a given M_* that inhabit

haloes of mass M_h - for galaxies in the stellar mass range $10^{9.5} < M_*/h^{-2}M_\odot < 10^{10}$ split in centrals and satellites (Fig. 9), or in bins of SFR for three ranges in M_* (Fig. 10).

Central galaxies in the mass range $10^{9.5} < M_*/h^{-2}M_\odot < 10^{10}$, inhabit haloes with a narrow range of masses, from $10^{11.2}h^{-1}M_\odot$ to $\sim 10^{13}h^{-1}M_\odot$. In contrast, satellites of the same stellar mass, inhabit haloes with a wide range of masses, from $10^{11}h^{-1}M_\odot$ to $\sim 10^{14.5}h^{-1}M_\odot$ (Fig. 9). The (much) stronger clustering of satellites is therefore clearly due to the significant fraction that resides in these much more massive (and hence more clustered) haloes (see e.g., Guo et al. 2014).

At a given stellar mass, EAGLE galaxies with a higher star formation rate inhabit lower mass haloes than those with a lower value of SFR (at $z = 0.1$), as shown in Fig. 10

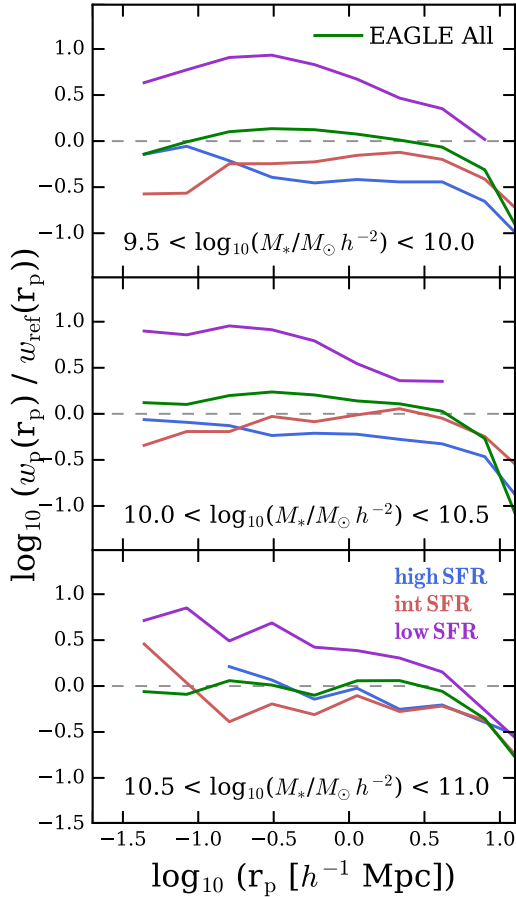


Figure 8. Clustering as function of star formation rate. The projected galaxy correlation function of EAGLE galaxies, divided by the reference power-law of Eq. (7), is plotted for galaxies in bins of stellar mass, as labelled in each panel. In each galaxy stellar mass bin, we show the result for all galaxies (green line), the 30 per cent least star forming galaxies (‘low SFR’, purple), the 30 per cent highest star forming galaxies (‘high SFR’, blue), and the remainder (‘int SFR’, red). Table 3 lists the corresponding cuts in star formation rate.

for three ranges in M_* . The figure also demonstrates that the halo occupation is similar for galaxies with a high or intermediate SFR. This explains the results from Fig. 8 that, at given M_* , galaxies with a low SFR cluster more strongly than those with higher SFR.

The SFR (and hence also colour) dependence of clustering in EAGLE is related to the mechanism that causes some galaxies to have a low SFR for their mass: the reduction in SFR once a galaxy becomes a satellite³, which is discussed in detail for the EAGLE simulations by Trayford et al. (2016). The reduction of the SFR of satellites results from two related physical processes that operate in hydrodynamical simulations: ram-pressure stripping, mainly of the outer parts of satellites as shown by Bahé & McCarthy (2015) using the GIMIC simulations (Crain et al. 2009), and the strong suppression - by many orders of magnitude - of the accretion rate of gas onto satellites shown by Van de Voort et al.

³ AGN feedback plays a role at higher M_* as well.

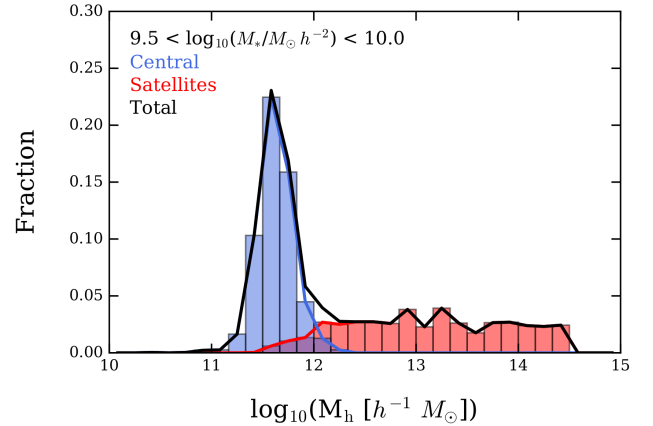


Figure 9. Halo occupation distribution of EAGLE galaxies with stellar mass in the range $9.5 < \log_{10}(M_*/M_\odot h^{-2}) < 10.0$. The distribution for all galaxies in that stellar mass range is plotted as the black histogram, which is normalised to unit integral. The red and blue histograms show the fraction of those galaxies that are satellites and centrals, respectively. The coloured histograms integrate separately to the fraction of galaxies that are satellite (red) or central (blue) in this range of M_* .

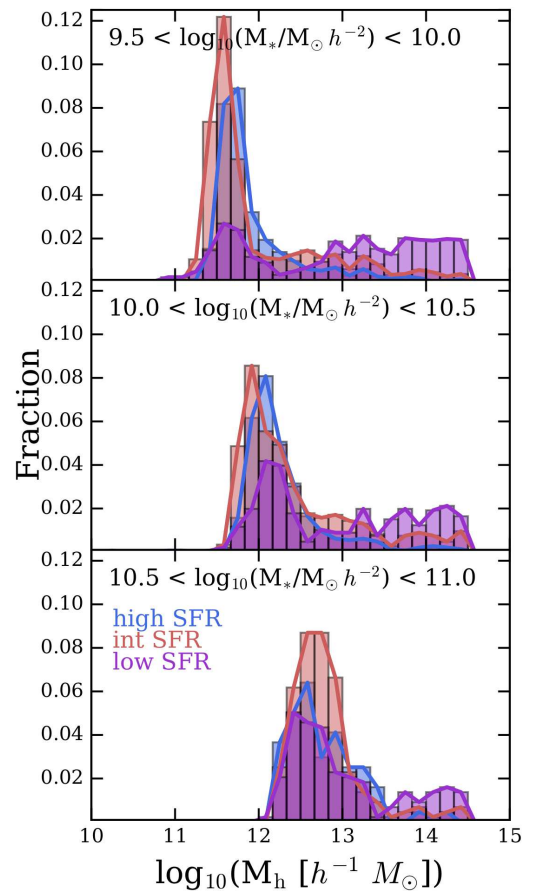


Figure 10. As in Fig. 9, but for EAGLE galaxies split by star formation rate and for a three ranges in stellar mass as indicated in each panel (with increasing stellar mass from top to bottom). The sample cuts are the same as in Fig. 8 (and listed in Table 3).

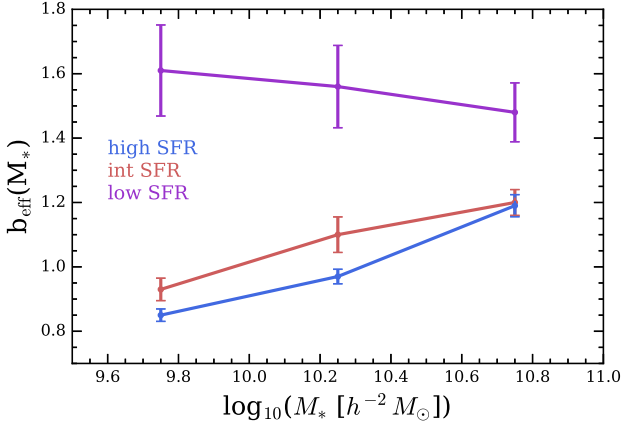


Figure 11. Effective bias, estimated using Eq. 11, with jackknife errors for EAGLE galaxies as a function of stellar mass, for three cuts in specific star formation rate, low (purple), intermediate (red) and high (blue). The sample cuts are the same as in Figs. 8 and 10 and are listed in Table 3.

(*in prep.*) in EAGLE. The much reduced gas fraction of such satellites then also implies that their interstellar medium rapidly increases in metallicity (Bahé et al., *in prep.*) - a testable prediction of the scenario.

Another way to demonstrate the bias of quenched galaxies to inhabit more massive haloes is shown in Fig. 11, which plots the effective bias of galaxies as function of stellar mass, split by SFR. The effective bias of the low SFR population is nearly independent of M_* , and considerably higher than that of the intermediate or high SFR population. For the active galaxies with intermediate or high SFR, the bias increases with stellar mass - simply reflecting that for those galaxies SFR increases with M_* , which in turn increases with halo mass.

5 COMPARISON WITH OTHER MODELS

In this section we compare the EAGLE clustering results to two sets of models: *i*) two incarnations of the GALFORM semi-analytical model, namely the version of Gonzalez-Perez et al. (2014) (hereafter GP14), and an early version of Lacey et al. (2016) (hereafter L14), both of which were used in the GAMA clustering study of Farrow et al. (2015); *ii*) the ILLUSTRIS hydrodynamical simulation described by Vogelsberger et al. (2014).

The GALFORM model assumes that galaxies form in dark matter haloes, and it uses analytical prescriptions to describe galaxy formation processes. These phenomenological prescriptions have free parameters controlling different physical processes necessary for the model to be realistic and which are tuned to fit a set of observational constraints at low redshift. GP14 & L14 use halo merger trees from the MILLENNIUM-MR7 simulation (Guo et al. 2013), which uses cosmological parameters set by WMAP7 (Komatsu et al. 2011). The Gonzalez-Perez et al. (2014) model has the same physical prescriptions as the Lagos et al. (2012) model, but set in a different cosmology to that used by Lagos et al. (2012), resulting in the need to retune the free parameters as described by Gonzalez-Perez et al. (2014). The main dif-

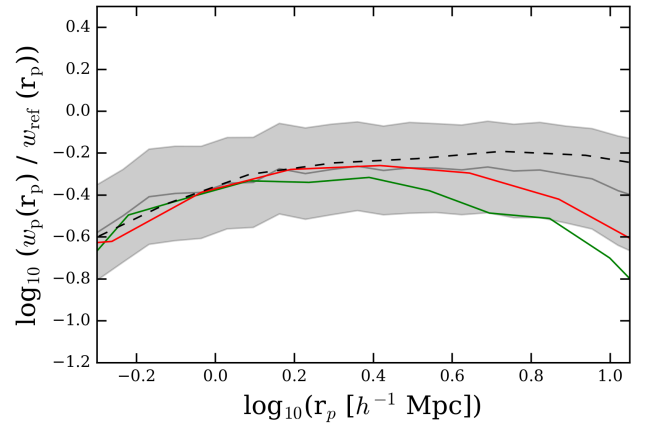
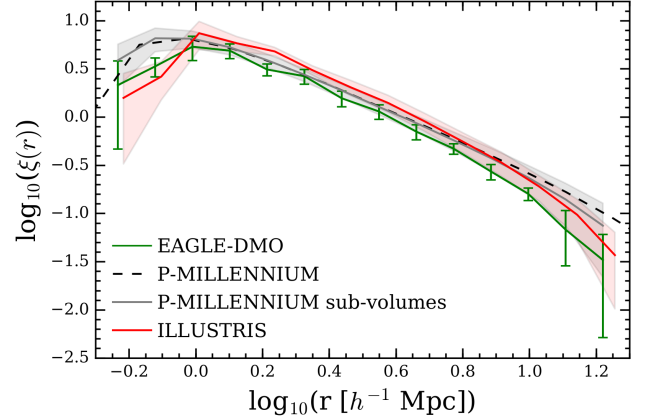


Figure 12. Correlation function of $z = 0$ of dark matter haloes with mass $M_h > 10^{12} h^{-1} M_\odot$ in real space (*top panel*), and their corresponding projected correlation function divided by a reference power law (*bottom panel*). Simulations shown are EAGLE-DMO (green line with jackknife error bars), ILLUSTRIS dark matter only (red line with jackknife errors represented by red shaded area), and P-MILLENNIUM with line styles as in Fig. 2.

ferences between GP14 and L14 are: *i*) the assumed stellar initial mass function, with GP14 using a Kennicutt (1983) IMF, while L14 switches from this IMF to a top-heavy IMF in star bursting galaxies (see Lacey et al. 2016, for further details); *ii*) the treatment of merging satellite galaxies, with GP14 using the Chandrasekhar dynamical friction timescale in an isothermal sphere as given in Lacey & Cole (1993), while L14 uses the Jiang et al. (2008) and Jiang et al. (2014) formula for the time-scale, which is empirically calibrated on N-body simulations to account for the tidal stripping of the accreting haloes; *iii*) the assumed stellar population synthesis (SPS) model, with GP14 using an updated version of the Bruzual & Charlot (1993) SPS model, while L14 adopts the Maraston (2005) SPS model. Both GP14 and L14 were tuned to reproduce the b_J -band and K-band luminosity functions of Norberg et al. (2002b) and Cole et al. (2001), respectively. As GAMA is r-band selected and as Farrow et al. (2015) analysis covered a larger redshift range not probed by those galaxy luminosity functions used to calibrate the GP14 and the L14 GALFORM models, Farrow et al. (2015) adjusted the GAMA GALFORM lightcone mocks constructed following Merson et al. (2013) to closely

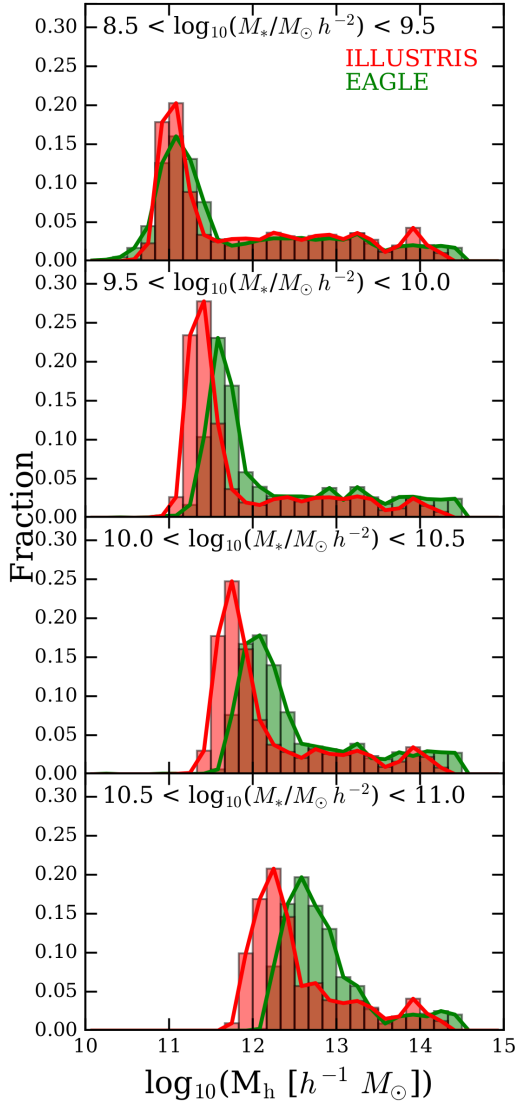


Figure 13. Halo occupation distribution of galaxies at $z = 0.1$ in four different stellar mass bins, as indicated in each panel, comparing EAGLE (green) and ILLUSTRIS (red). With the exception of the lowest stellar mass bin, galaxies of a given mass tend to reside in lower mass haloes in ILLUSTRIS compared to EAGLE.

reproduce the GAMA r-band selection function (which in turn is well described by the GAMA r-band luminosity functions of Loveday et al. (2012, 2015)). We note that the L14 model used here and by Farrow et al. (2015) has marginally different parameters from the model discussed by Lacey et al. (2016); we have not investigated whether this impacts any of the results presented below.

The ILLUSTRIS simulation suite was performed with the AREPO moving-mesh code of Springel (2010). The simulation volume, $(75h^{-1}\text{Mpc})^3$, is comparable to that of EAGLE; sub-grid modules for star formation and feedback are as described by Vogelsberger et al. (2014) and the assumed cosmological parameters⁴ are close to those of

⁴ $\Omega_m = 0.2726$, $\Omega_\Lambda = 0.7274$, $\Omega_b = 0.0456$, $\sigma_8 = 0.809$, $n_s = 0.963$, and $H_0 = 100h\text{kms}^{-1}\text{Mpc}^{-1}$ with $h = 0.704$

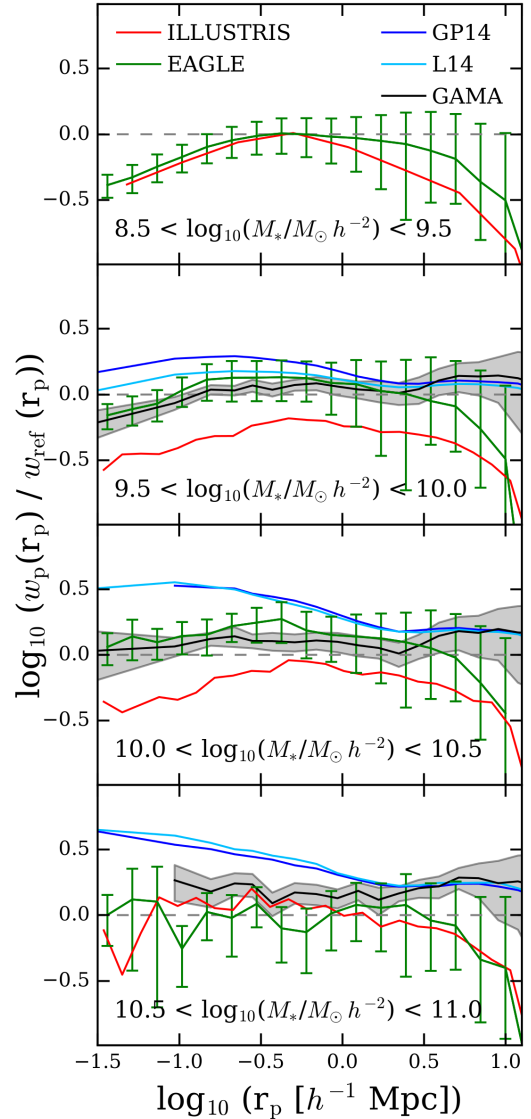


Figure 14. Comparison of galaxy clustering between different models and GAMA, as a function of stellar mass (M_* increases from top to bottom). Different panels show the projected correlation function, $w_p(r_p)$, divided by the reference fit of Eq. (7). The models shown are EAGLE (green curve with jackknife errors) ILLUSTRIS (red curve - jackknife errors are not shown but are similar to those of EAGLE), GP14 (dark blue curve) and L14 (cyan curve). The observed correlation function from GAMA is shown in black with grey shading encompassing jackknife errors. The latter three sets of results are from Farrow et al. (2015). Models GP14 and L14 and GAMA curves are taken from Farrow et al. 2015

Planck Collaboration et al. (2014) assumed in EAGLE. Properties of ILLUSTRIS galaxies were released by the collaboration through a database⁵ with content described by Nelson et al. (2015). The simulation reproduces several observed properties of galaxies such as for example the colours of satellites (Sales et al. 2015) and the distribution of galaxy morphologies (Snyder et al. 2015). However, the galaxy stellar mass function simulation has an excess of galaxies at

⁵ <http://www.illustris-project.org>

both high ($M_\star \geq 10^{11.5} M_\odot$) and low ($M_\star \leq 10^{10} M_\odot$) stellar masses at redshift $z \leq 1$, see [Vogelsberger et al. \(2014\)](#). Here we use the ‘Illustris-1’ run (hereafter ILLUSTRIS) and extract galaxy properties directly from the the ILLUSTRIS database.

The ILLUSTRIS simulation suite also includes dark matter-only runs. This enables us to compare the clustering of haloes between EAGLE-DMO and the dark matter-only ILLUSTRIS (Fig. 12). The correlation function of ILLUSTRIS haloes is higher than that of EAGLE-DMO, both in real and redshift space (except for the smallest scales plotted), with the difference consistent with sample variance as judged from the scatter obtained from EAGLE like simulation sub-volumes extracted from the significantly larger P-MILLENNIUM. As discussed before, lack of large-scale power in the smaller boxes and the absence of integral constraint corrections on the clustering estimate cause the correlation functions to drop below that of P-MILLENNIUM on larger scales.

Given that the EAGLE and ILLUSTRIS dark matter halo functions are very similar, whereas their galaxy stellar mass function are not (see Fig. 5 of [Schaye et al. 2015](#)), we expect some differences between the simulations in how galaxies populate the underlying dark matter haloes. This is indeed born-out by Fig. 13: ILLUSTRIS galaxies of given stellar mass prefer lower mass haloes, by about 0.2 dex.

In Figure 14 we compare the two-point correlation function, w_p , from EAGLE galaxies with the results from ILLUSTRIS, the GP14 and L14 GALFORM models, and the GAMA survey, in four stellar mass bins. We divided w_p by the reference model w_p^{ref} of Eq. (7) to decrease the dynamic range in the plot. The shape of the correlation functions of EAGLE and ILLUSTRIS are very similar (we do not plot jackknife errors on the ILLUSTRIS curves to avoid clutter, but they are nearly identical to those of EAGLE), but with ILLUSTRIS offset to smaller values except for the lowest bin in stellar mass (the top panel). The poor sampling of large-scale modes in both hydrodynamical simulations, combined with the integral constraint, may lead to an net offset of w_p - as we demonstrated explicitly in Fig. 2 for the dark matter haloes. The level of the offset is consistent with sample variance - as shown by the comparing to the P-MILLENNIUM results. However, somewhat surprisingly, whereas haloes in ILLUSTRIS are more strongly clustered than those in EAGLE (red line above green line in Fig. 12), ILLUSTRIS galaxies are *less* strongly clustered than those in EAGLE at given M_\star (Fig. 14). This is related to the differences in stellar mass function: the galaxy number density is higher in ILLUSTRIS compared to EAGLE, therefore ILLUSTRIS galaxies of given M_\star inhabit haloes of lower mass (Fig. 13) which are less clustered. This effect is relatively small, however, and we conclude that the clustering is consistent in both models given the relatively large jackknife errors.

We can partially compensate for differences in the stellar mass function of EAGLE and ILLUSTRIS by comparing galaxy clustering at a given number density, rather than stellar mass. To do so, we select all EAGLE galaxies with $M_\star > 10^{9.5} h^{-2} M_\odot$, yielding a galaxy number density of $n_{\text{gal}} \simeq 1.49 \times 10^{-2} h^3 \text{Mpc}^{-3}$ and then find the corresponding stellar mass range of $M_\star > 10^{9.73} h^{-2} M_\odot$ in ILLUSTRIS above which the number density is equal to n_{gal} . The clustering of these two samples is compared in Figure 15, with ILLUSTRIS in red and EAGLE in green, with sample variance

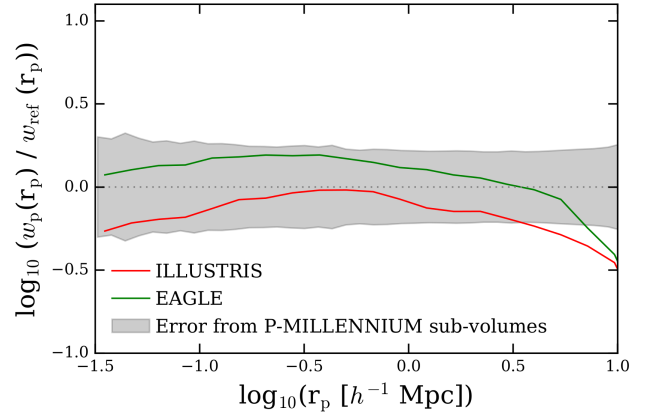


Figure 15. Clustering of the most massive galaxies corresponding to a mean number density of $n_{\text{gal}} \sim 1.4910^{-2} h^3 \text{Mpc}^{-3}$ at $z = 0.1$. The projected correlation function w_p is divided by the reference model w_p^{ref} of Eq. (7). The EAGLE and ILLUSTRIS simulations are plotted in green and red, respectively. Sample variance in the clustering of dark matter haloes, mass ranked and selected to have the same mean number density, estimated using from sub-sampling the P-MILLENNIUM volume, are shown by the grey shaded area.

in the clustering of dark matter haloes with the same number density estimated by sub-sampling P-MILLENNIUM volumes in grey. Selected in this way, the clustering in EAGLE is higher than in ILLUSTRIS, but the differences are consistent with sample variance.

We now turn to the semi-analytical models plotted in Fig. 14. The GALFORM model shown are the average of 26 GAMA light cone mocks as described in [Farrow et al. \(2015\)](#). Like in that paper, we assume that the errors on those model clustering results are negligible compared to the errors measured on the GAMA sample. See [Farrow et al. \(2015\)](#) for a quantitative description of how adequate the GP14 and L14 GALFORM models are in describing the observed GAMA clustering. The correlation functions of GP14 and L14 are very similar, except in the second panel from the top where the GP14 model is above that of L14 on scales below $r_p \sim 1 h^{-1} \text{Mpc}$. Both models show stronger clustering than observed below $r_p \sim 1 h^{-1} \text{Mpc}$. As discussed by [Farrow et al. \(2015\)](#), the high values of w_p in the GALFORM models are caused by an excess of satellites galaxies and/or their radial distribution in clusters. The satellite merging scheme is the principal mechanism which impacts directly on the number of satellites within haloes. The two versions of GALFORM we use, include the default scheme in which satellite galaxies merge onto the central galaxy after an analytically determined dynamical friction merger time-scale. [Campbell et al. \(2015\)](#) compared the standard GALFORM scheme with a different one ([Simha & Cole 2016](#)), in which the merger time-scale make use of the information from dark matter subhalo of each satellite galaxy. They find that the new scheme reduces the amplitude at small scales and show good agreement with observational results. McCullagh et al. (*prep*) show that implementing the [Simha & Cole \(2016\)](#) merger scheme within the GP14 model and applying it to the P-MILLENNIUM simulation results in clustering measurements that are in significantly better agreement with the observed ones on small scales. More detailed studies of the

GALFORM clustering predictions on small scales are needed to address the known limitations of samples split by colour (see e.g. Fig. 14 of Farrow et al. 2015). At large-scales the semi-analytic models show a good agreement with observational data.

To summarise: we find that both EAGLE and ILLUSTRIS reproduce the clustering of galaxies in GAMA on scales below $r_p \sim 4h^{-1}$ Mpc, but both jackknife errors and sample variance are still relatively large. Above this scale, both simulations are affected by the relatively small simulation volume. The good agreement show that these models both reproduce the dominant effects of environment on satellites, a crucial ingredient in getting w_p right. Larger simulation volumes, that would yield smaller errors, are needed to make the clustering constraint more stringent. The hydrodynamical models perform better on these smaller scales than the GALFORM models of GP14 and L14.

6 CONCLUSIONS

We have studied the two-point correlation function of galaxies at $z = 0.1$ in the EAGLE cosmological hydrodynamical simulation (Schaye et al. 2015). The sub-grid parameters of EAGLE are calibrated as described by Crain et al. (2015) to the present-day galaxy stellar mass function (amongst other observables, such as galaxy sizes), but the clustering properties of galaxies is a prediction of the simulation. We have compared the results to the clustering of observed galaxies from the GAMA survey (Driver et al. 2011), as well as two incarnations of the GALFORM semi-analytical model (the GP14 model described by Gonzalez-Perez et al. (2014), and an early version of the Lacey et al. (2016) model, referred to as L14 model described in Farrow et al. (2015)), and the ILLUSTRIS simulation described by Vogelsberger et al. (2014).

The simulation volume of the largest EAGLE simulation we use here, is still relatively small at $(100 \text{ Mpc})^3$. We examine how lack of (and poor sampling of) large-scale modes and sample variance might affect clustering by comparing the real-space clustering of dark matter haloes with $M_h > 10^{12}h^{-1}M_\odot$ in a dark matter-only version of EAGLE (called EAGLE-DMO) to that in the much larger volume, $(800 \text{ Mpc})^3$, of the P-MILLENNIUM dark matter only simulation that uses the same Planck Collaboration et al. (2014) cosmology. We find that the clustering amplitude is similar in the range $r \sim 1 - 6h^{-1}$ Mpc, while the clustering amplitude in EAGLE-DMO is smaller on larger scales. We therefore focus our attention on scales up to $\sim 6h^{-1}$ Mpc. We also show that jackknife-error estimates of the clustering amplitude calculated for EAGLE-DMO are similar to the variance measured between EAGLE-sized volumes drawn from P-MILLENNIUM. This encourages us to quote jackknife errors for clustering of EAGLE galaxies as well.

We use the optical broad-band magnitudes and colours of EAGLE galaxies calculated using the fiducial model of Trayford et al. (2015). Briefly, this calculation combines the Bruzual & Charlot (2003) population synthesis model for stars with a two-component screen model for dust. The dust model is based on that of Charlot & Fall (2000), with dust-screen optical depth depending on the enriched star-forming gas content of galaxies, and including additional scatter to represent orientation effects.

Trayford et al. (2015) show that the r-band luminosity function of EAGLE agrees well with observations. The luminosity-dependent fraction of red and blue galaxies also performs reasonably well, although an excess of blue galaxies is found at high mass. Using these predicted fluxes allows us to compare the projected two-point correlation function, $w_p(r_p)$ (defined in Eq. 6) in EAGLE directly to that measured in GAMA for galaxies selected by r-band absolute magnitude $M_{r,h}$ and/or $(g-r)_0$ rest-frame colour, in addition to stellar mass. We do so using the (nearly) volume-limited sample of GAMA galaxies described by Farrow et al. (2015).

Our findings can be summarised as follows:

- The (projected) clustering of EAGLE galaxies in bins of stellar mass agrees well with that from GAMA, with differences consistent within the errors, and we find similar good agreement for galaxies selected in bins of r-band absolute magnitude, $M_{r,h}$. Given that the number densities of these galaxies in simulation and data agree as well, this gives us confidence that EAGLE galaxies of given mass or luminosity, inhabit haloes of similar mass as those in GAMA. The observed clustering amplitude increases with mass and luminosity. This trend is not seen in EAGLE, however the EAGLE clustering results are still consistent with the data given the relatively large jackknife errors and the effect of missing large-scale power.

- At a given stellar mass, red EAGLE galaxies are more strongly clustered than blue galaxies. In EAGLE, red galaxies are either satellites - with ram-pressure stripping of gas (Bahé & McCarthy 2015) and a reduction in the cosmological accretion rate of fresh gas (Van de Voort et al., *in prep.*) both playing in role in reducing the star formation rate and making the galaxy red, or their star formation is reduced by their central black hole (Trayford et al. 2016). The stronger clustering of red galaxies is then a consequence of the higher halo mass they inhabit, compared to blue galaxies of the same M_* . The difference in clustering amplitude between red and blue galaxies is too strong in EAGLE compared to GAMA. This overabundance of red galaxies in EAGLE is at least partly due to lack of numerical resolution (Schaye et al. 2015; Trayford et al. 2016), although poor sampling of massive groups and clusters plays a role as well.

- On small scales, low SFR galaxies cluster more strongly for all the stellar mass bins studied. This is because galaxies with low SFR at a given mass tend to be satellites, and as before, the enhanced clustering reflects that of their more massive dark matter hosts.

We conclude that the galaxy clustering predicted by EAGLE is in very good agreement with GAMA on projected scales up to $r_p \sim 4h^{-1}$ Mpc, also when galaxies are split by colour. EAGLE and observed galaxies therefore inhabit haloes of similar mass, and the reduction in the SFR of EAGLE galaxies when they become a satellite, mimics that of observed galaxies. However, the limited simulation volume of the simulation yields relatively large jackknife errors as well as large sample variance. Better tests of the realism of EAGLE requires clustering studies in somewhat large volumes.

Comparing to other models, we find that both the GP14 and L14 semi-analytical models overestimate the galaxy clustering amplitude at small scales, $r_p \lesssim 1h^{-1}$ Mpc, while showing good agreement with GAMA on larger scales. We speculate that the excess at small scales is caused by the

satellite merging schemes implemented, which are crucial and impact directly on the number of satellites and their radial distribution (Contreras et al. 2013; Campbell et al. 2015). The ILLUSTRIS simulation yields very similar clustering measures to EAGLE. At a given stellar mass, the clustering amplitude in ILLUSTRIS is lower than in EAGLE, although the difference is consistent given the jackknife error estimates. This good agreement is slightly fortuitous: the fact that ILLUSTRIS galaxies tend to inhabit haloes of lower mass than EAGLE galaxies (consistent with ILLUSTRIS over predicting the galaxy stellar mass function for most values of M_* , Vogelsberger et al. 2014, - which would yield lower clustering - is partially compensated by ILLUSTRIS haloes clustering more strongly than EAGLE haloes (with the difference consistent with sample variance)

Galaxy clustering measurements provide powerful constraints on galaxy formation models. Here we have shown that the EAGLE simulation reproduces the spatial distribution of galaxies measured in the GAMA survey, even when galaxies are split by stellar mass, luminosity and colour. This increases our confidence in the realism of the simulation. However, sample variance is still relatively large, given the small volume simulated, and better constraints require larger simulations, even when studying clustering on the smaller scales where the galaxy formation modelling is tested most stringently.

ACKNOWLEDGEMENT

The authors would like to thank to Joop Schaye and Robert Crain for useful comments. MCA acknowledges support from the European Commission's Framework Programme 7, through the Marie Curie International Research Staff Exchange Scheme LACEGAL (PIRSES-GA-2010-269264). This work has been partially supported by PICT Raices 2011/959 of Ministry of Science (Argentina). IZ is supported by NSF grant AST-1612085 and by a CWRU Faculty Seed Grant. MCA, SP and IZ acknowledge the hospitality of the ICC at Durham, where this project was started. This work was supported by the Science and Technology Facilities Council [grant number ST/F001166/1], by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office ([AP P7/08 CHARM]). We used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-Infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC is part of the National E-Infrastructure. PN acknowledges the support of the Royal Society through the award of a University Research Fellowship and the European Research Council, through receipt of a Starting Grant (DEGAS-259586). TT, PN, RGB and MS acknowledge the support of the Science and Technology Facilities Council (ST/L00075X/1). The data used in the work is publically available in the EAGLE database described by McAlpine et al. (2016). GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Sur-

vey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including GALEX MIS, VST KiDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is <http://www.gama-survey.org/>.

REFERENCES

- Abazajian K. N., et al., 2009, *ApJS*, **182**, 543
 Bahé Y. M., McCarthy I. G., 2015, *MNRAS*, **447**, 969
 Baldry I. K., et al., 2010, *MNRAS*, **404**, 86
 Baldry I. K., et al., 2012, *MNRAS*, **421**, 621
 Baldry I. K., et al., 2014, *MNRAS*, **441**, 2440
 Baugh C. M., 2006, *Reports on Progress in Physics*, **69**, 3101
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, **575**, 587
 Booth C. M., Schaye J., 2009, *MNRAS*, **398**, 53
 Bruzual A. G., Charlot S., 1993, *ApJ*, **405**, 538
 Bruzual G., Charlot S., 2003, *MNRAS*, **344**, 1000
 Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storch-Bergmann T., 2000, *ApJ*, **533**, 682
 Campbell D. J. R., et al., 2015, *MNRAS*, **452**, 852
 Camps P., Trayford J. W., Baes M., Theuns T., Schaller M., Schaye J., 2016, *MNRAS*, **462**, 1057
 Chabrier G., 2003, *PASP*, **115**, 763
 Charlot S., Fall S. M., 2000, *ApJ*, **539**, 718
 Coil A. L., Newman J. A., Cooper M. C., Davis M., Faber S. M., Koo D. C., Willmer C. N. A., 2006, *ApJ*, **644**, 671
 Coil A. L., et al., 2008, *ApJ*, **672**, 153
 Cole S., 2011, *MNRAS*, **416**, 739
 Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, *MNRAS*, **271**, 781
 Cole S., et al., 2001, *MNRAS*, **326**, 255
 Cole S., et al., 2005, *MNRAS*, **362**, 505
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, **647**, 201
 Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, *MNRAS*, **432**, 2717
 Cooray A., 2002, *ApJ*, **576**, L105
 Coupon J., et al., 2012, *A&A*, **542**, A5
 Crain R. A., et al., 2009, *MNRAS*, **399**, 1773
 Crain R. A., et al., 2015, *MNRAS*, **450**, 1937
 Crain R. A., et al., 2016, preprint, ([arXiv:1604.06803](https://arxiv.org/abs/1604.06803))
 Croton D. J., Norberg P., Gaztañaga E., Baugh C. M., 2007, *MNRAS*, **379**, 1562
 Dalla Vecchia C., Schaye J., 2012, *MNRAS*, **426**, 140
 Davis M., Peebles P. J. E., 1983, *ApJ*, **267**, 465
 Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, **292**, 371
 Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, **399**, 497
 Driver S. P., et al., 2011, *MNRAS*, **413**, 971
 Eisenstein D. J., et al., 2005, *ApJ*, **633**, 560
 Farrow D. J., et al., 2015, *MNRAS*, **454**, 2120
 Favole G., McBride C. K., Eisenstein D. J., Prada F., Swanson M. E., Chuang C.-H., Schneider D. P., 2016, *MNRAS*, **462**, 2218
 Furlong M., et al., 2015, *MNRAS*, **450**, 4486
 Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, *MNRAS*, **439**, 264
 Goto T., Yamauchi C., Fujita Y., Okamura S., Sekiguchi M., Smail I., Bernardi M., Gomez P. L., 2003, *MNRAS*, **346**, 601
 Guo H., et al., 2013, *ApJ*, **767**, 122
 Guo H., et al., 2014, *MNRAS*, **441**, 2398

- Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. ([arXiv:astro-ph/0106018](#))
- Hellwing W. A., Schaller M., Frenk C. S., Theuns T., Schaye J., Bower R. G., Crain R. A., 2016, *MNRAS*, **461**, L11
- Hinshaw G., et al., 2013, *ApJS*, **208**, 19
- Jenkins A., Booth S., 2013, preprint, ([arXiv:1306.5771](#))
- Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, *MNRAS*, **321**, 372
- Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, *ApJ*, **675**, 1095
- Jiang C. Y., Jing Y. P., Han J., 2014, *ApJ*, **790**, 7
- Kaiser N., 1987, *MNRAS*, **227**, 1
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, **264**, 201
- Kennicutt Jr. R. C., 1983, *ApJ*, **272**, 54
- Komatsu E., et al., 2011, *ApJS*, **192**, 18
- Lacey C., Cole S., 1993, *MNRAS*, **262**, 627
- Lacey C. G., et al., 2016, *MNRAS*, **462**, 3854
- Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T. A., Fanidakis N., Geach J. E., 2012, *MNRAS*, **426**, 2142
- Landy S. D., Szalay A. S., 1993, *ApJ*, **412**, 64
- Laureijs R., et al., 2011, preprint, ([arXiv:1110.3193](#))
- Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, *MNRAS*, **368**, 21
- Linder E. V., 2008, *Astroparticle Physics*, **29**, 336
- Liske J., et al., 2015, *MNRAS*, **452**, 2087
- Loveday J., et al., 2012, *MNRAS*, **420**, 1239
- Loveday J., et al., 2015, *MNRAS*, **451**, 1540
- Maraston C., 2005, *MNRAS*, **362**, 799
- McAlpine S., et al., 2016, *Astronomy and Computing*, **15**, 72
- McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2016, preprint, ([arXiv:1603.02702](#))
- McNaught-Roberts T., et al., 2014, *MNRAS*, **445**, 2125
- Merson A. I., et al., 2013, *MNRAS*, **429**, 556
- Mo H. J., White S. D. M., 1996, *MNRAS*, **282**, 347
- Murray S. G., Power C., Robotham A. S. G., 2013, *Astronomy and Computing*, **3**, 23
- Nelson D., et al., 2015, *Astronomy and Computing*, **13**, 12
- Norberg P., et al., 2001, *MNRAS*, **328**, 64
- Norberg P., et al., 2002a, *MNRAS*, **332**, 827
- Norberg P., et al., 2002b, *MNRAS*, **336**, 907
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, **396**, 19
- Peebles P. J. E., 1980, The large-scale structure of the universe
- Planck Collaboration et al., 2014, *A&A*, **571**, A16
- Planck Collaboration et al., 2016, *A&A*, **594**, A13
- Porciani C., Magliocchetti M., Norberg P., 2004, *MNRAS*, **355**, 1010
- Rivolo A. R., 1986, *ApJ*, **301**, 70
- Robotham A., et al., 2010, *Publ. Astron. Soc. Australia*, **27**, 76
- Rosas-Guevara Y. M., et al., 2015, *MNRAS*, **454**, 1038
- Sales L. V., et al., 2015, *MNRAS*, **447**, L6
- Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, *MNRAS*, **431**, 1366
- Schaller M., et al., 2015, *MNRAS*, **451**, 1247
- Schaye J., 2004, *ApJ*, **609**, 667
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, **383**, 1210
- Schaye J., et al., 2010, *MNRAS*, **402**, 1536
- Schaye J., et al., 2015, *MNRAS*, **446**, 521
- Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, *MNRAS*, **417**, 2020
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *MNRAS*, **343**, 978
- Simha V., Cole S., 2013, *MNRAS*, **436**, 1142
- Simha V., Cole S., 2016, preprint, ([arXiv:1609.09520](#))
- Snyder G. F., et al., 2015, *MNRAS*, **454**, 1886
- 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., Di Matteo T., Hernquist L., 2005, *MNRAS*, **361**, 776
- Taylor E. N., et al., 2011, *MNRAS*, **418**, 1587
- Tinker J. L., et al., 2012, *ApJ*, **745**, 16
- Trayford J. W., et al., 2015, *MNRAS*, **452**, 2879
- Trayford J. W., Theuns T., Bower R. G., Crain R. A., Lagos C. d. P., Schaller M., Schaye J., 2016, *MNRAS*, **453**, 189
- Vale A., Ostriker J. P., 2004, *MNRAS*, **353**, 189
- Vale A., Ostriker J. P., 2006, *MNRAS*, **371**, 1173
- Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, *MNRAS*, **442**, 2641
- Vogelsberger M., et al., 2014, *MNRAS*, **444**, 1518
- Wang W., Sales L. V., Henriques B. M. B., White S. D. M., 2014, *MNRAS*, **442**, 1363
- Wiersma R. P. C., Schaye J., Smith B. D., 2009a, *MNRAS*, **393**, 99
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, *MNRAS*, **399**, 574
- York D. G., et al., 2000, *AJ*, **120**, 1579
- Zehavi I., et al., 2002, *ApJ*, **571**, 172
- Zehavi I., et al., 2005, *ApJ*, **630**, 1
- Zehavi I., et al., 2011, *ApJ*, **736**, 59
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, **667**, 760
- van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, *MNRAS*, **415**, 3649
- van Daalen M. P., Schaye J., McCarthy I. G., Booth C. M., Dalla Vecchia C., 2014, *MNRAS*, **440**, 2997