

# The SAMI Galaxy Survey: comparing 3D spectroscopic observations with galaxies from cosmological hydrodynamical simulations

Jesse van de Sande<sup>1,2\*</sup>, Claudia D.P. Lagos<sup>2,3</sup>, Charlotte Welker<sup>2,3</sup>,  
 Joss Bland-Hawthorn<sup>1,2</sup>, Felix Schulze<sup>4,5</sup>, Rhea-Silvia Remus<sup>4</sup>, Yannick Bahé<sup>6</sup>,  
 Sarah Brough<sup>2,7</sup>, Julia J. Bryant<sup>1,2,8</sup>, Luca Cortese<sup>2,3</sup>, Scott M. Croom<sup>1,2</sup>,  
 Julien Devriendt<sup>9</sup>, Yohan Dubois<sup>10</sup>, Michael Goodwin<sup>11</sup>, Iraklis S. Konstantopoulos<sup>12</sup>,  
 Jon S. Lawrence<sup>11</sup>, Anne M. Medling<sup>13,14,15</sup>, Christophe Pichon<sup>10,16</sup>,  
 Samuel N. Richards<sup>17</sup>, Sebastian F. Sanchez<sup>18</sup>, Nicholas Scott<sup>1,2</sup> and Sarah M. Sweet<sup>2,19</sup>

*Affiliations are listed at the end of the paper*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

Cosmological hydrodynamical simulations are rich tools to understand the build-up of stellar mass and angular momentum in galaxies, but require some level of calibration to observations. We compare predictions at  $z \sim 0$  from the EAGLE, HYDRANGEA, HORIZON-AGN, and MAGNETICUM simulations with integral field spectroscopic (IFS) data from the SAMI Galaxy Survey, ATLAS<sup>3D</sup>, CALIFA and MASSIVE surveys. The main goal of this work is to simultaneously compare structural, dynamical, and stellar population measurements in order to identify key areas of success and tension. We have taken great care to ensure that our simulated measurement methods match the observational methods as closely as possible, and we construct samples that match the observed stellar mass distribution for the combined IFS sample. We find that the EAGLE and HYDRANGEA simulations reproduce many galaxy relations but with some offsets at high stellar masses. There are moderate mismatches in  $R_e$  (+),  $\epsilon$  (−),  $\sigma_e$  (−), and mean stellar age (+), where a plus sign indicates that quantities are too high on average, and minus sign too low. The HORIZON-AGN simulations qualitatively reproduce several galaxy relations, but there are a number of properties where we find a quantitative offset to observations. Massive galaxies are better matched to observations than galaxies at low and intermediate masses. Overall, we find mismatches in  $R_e$  (+),  $\epsilon$  (−),  $\sigma_e$  (−) and  $(V/\sigma)_e$  (−). MAGNETICUM matches observations well: this is the only simulation where we find ellipticities typical for disk galaxies, but there are moderate differences in  $R_e$  (+),  $\sigma_e$  (−),  $(V/\sigma)_e$  (−) and mean stellar age (+). Our comparison between simulations and observational data has highlighted several areas for improvement, such as the need for improved modelling resulting in a better vertical disk structure, yet our results demonstrate the vast improvement of cosmological simulations in recent years.

**Key words:** cosmology: observations – galaxies: evolution – galaxies: formation – galaxies: kinematics and dynamics – galaxies: stellar content – galaxies: structure

## 1 INTRODUCTION

In the present-day Universe, the majority of galaxies (> 85 percent) are consistent with being axisymmetric rotating oblate spheroids and only a minor fraction of galaxies have

\* jesse.vandesande@sydney.edu.au

complex dynamics (for a review, see Cappellari 2016). The ratio of ordered to random stellar motion in galaxies has a strong dependence on luminosity or stellar mass (Illingworth 1977; Davies et al. 1983; Emsellem et al. 2011; Brough et al. 2017; Veale et al. 2017a; van de Sande et al. 2017a; Green et al. 2018), which suggests a link between the build-up of stellar mass and angular momentum over time. Many theoretical studies are aimed at explaining the build-up and removal of angular momentum in galaxies through mergers (Naab et al. 2014, and citations within).

Binary galaxy merger simulations are a commonly used tool for studying the dynamical evolution of galaxies. These simulations showed that most merger remnants are consistent with being fast rotating galaxies (Bois et al. 2010, 2011), similar to what is found in the observational data. The dominant process for creating realistic slow rotating galaxies, however, is still a matter of debate (e.g., Bendo & Barnes 2000; Jesseit et al. 2009; Bois et al. 2011). The mass ratio of the progenitors in binary-disk mergers appears to be the most critical parameter for creating slow rotators, but there is also a strong dependence on specific spin-orbit alignments (Jesseit et al. 2009; Bois et al. 2010, 2011). Merger remnants formed from dissipational (wet) mergers of equal-mass disk galaxies better match the observed data than dissipationless (dry) merger remnants (Cox et al. 2006). This suggests that the presence of gas during mergers is crucial for creating slow rotators. However, this is in contrast with Taranu et al. (2013) who show that dissipation is not a prerequisite for producing slow-rotating galaxies. Instead, multiple, mostly dry, minor mergers are sufficient.

To disentangle the relative importance of major and minor mergers, and large-scale environment, in changing the angular momentum in galaxies, one requires a large ensemble of simulated galaxies with a range of initial conditions. Large cosmological hydrodynamical simulations are well suited for this. These simulations follow the growth and evolution of the galaxy from high-redshift ( $z \sim 50$ ) to the present-day ( $z = 0$ ) and provide more realistic insights into the formation paths and rotational properties of galaxies as compared to previous idealized, binary merger simulations.

The success of such an approach has already been demonstrated by Naab et al. (2014), Welker et al. (2017), Remus et al. (2017), Penoyre et al. (2017), Choi & Yi (2017), Choi et al. (2018), Lagos et al. (2018a,b), and Martin et al. (2018). Naab et al. (2014) use cosmological hydrodynamical zoom-in simulations of 44 individual central galaxies, and link the assembly history of these galaxies to their stellar dynamical features. Their analysis of the stellar kinematic data is done in an identical way to the ATLAS<sup>3D</sup> kinematic observations (Cappellari et al. 2011). They find a good qualitative agreement between the simulated and observed kinematic measurements. By following the merger histories of galaxies, Naab et al. (2014) show that there are multiple formation paths for fast and slow rotating galaxies, emphasizing the importance of studying large ensembles of simulated galaxies.

Penoyre et al. (2017) use the ILLUSTRIS simulations to follow the dynamical evolution of thousands of galaxies. They show that after  $z = 1$ , the merger and star-formation histories of slow and fast rotator progenitors start to differ. In contrast to Naab et al. (2014) and Lagos et al. (2018b), they find no major difference between the effects of gas-rich

and gas-poor mergers. Minor mergers also appear to have little correlation with the spin of galaxies. Lagos et al. (2018a) use EAGLE (Schaye et al. 2015; Crain et al. 2015) to analyse the effect galaxy mergers (with different parameters) have on the specific angular momentum ( $j_\star$ ) of galaxies. They show that, on average, dry mergers reduce  $j_\star$  by  $\approx 30$  per cent, while wet mergers increase  $j_\star$  by  $\approx 10$  per cent. Choi et al. (2018) and Lagos et al. (2018b) focus on the impact of galaxy mass and environment on the spin-down of galaxies, using the HORIZON-AGN (Dubois et al. 2014) and EAGLE simulations respectively. Both studies agree with the observational results from Veale et al. (2017b), Brough et al. (2017), and Green et al. (2017) that galaxy stellar mass plays a more dominant role in changing the spin-parameter proxy ( $\lambda_R$ ) of galaxies than environment. For satellite early-type galaxies, non-merger-induced tidal perturbations also appear to play a bigger role than mergers in lowering the galaxy spin parameter (Choi et al. 2018).

These specific angular momentum and spin parameter evolution predictions, however, assume that other galaxy parameters and scaling relations at  $z \sim 0$  are also well-matched to observations. Most simulated galaxy populations appear to overlap in terms of their dynamics and shapes (e.g., Penoyre et al. 2017 using ILLUSTRIS; Lagos et al. 2018b using EAGLE and HYDRANGEA), but some mismatch between the observations and simulations is also present (Choi et al. 2018, using HORIZON-AGN). The validity of simulation predictions become doubtful if the main parameter that is being used to make the predictions matches well with observations, while other parameters do not. Thus, a detailed comparison between multiple observational properties of galaxies from observations and simulations is needed to support the idea that conclusion from simulations apply to the real Universe.

Integral field spectroscopic (IFS) observations are ideally suited for a comparison with simulations. IFS galaxy surveys provide a unique opportunity to compare resolved two-dimensional stellar dynamical measurements across a large range of galaxy stellar masses and morphologies. Furthermore, IFS samples are typically selected from larger surveys that contain a wealth of ancillary data including structural parameters, stellar masses, and large scale environmental estimates. Cosmological hydrodynamical simulations are now also capable of creating large samples of mock-galaxies with dynamical observations with high enough spatial resolution to resolve some of the inner dynamical structures of galaxies.

In this paper, we compare structural, resolved dynamical, and stellar population observations of mock galaxies from the EAGLE, HYDRANGEA (Bahé et al. 2017), HORIZON-AGN, and MAGNETICUM simulations to observations from the Sydney-AAO Multi-object Integral field spectrograph (SAMI) Galaxy Survey (Croom et al. 2012; Bryant et al. 2015), the ATLAS<sup>3D</sup> Survey (Cappellari et al. 2011), the CALIFA survey (Sánchez et al. 2012), and the MASSIVE survey (Ma et al. 2014). The paper is organized as follows: Section 2 and 3 respectively present the data from the observations and simulations. In Section 4 we compare several observational relationships between stellar mass, size and dynamical parameters with the predictions from the simulations. We review previous comparison studies in Section 5. The implications of these matches and mismatches are discussed and summarised in the Section 6.

Throughout the paper we assume a  $\Lambda$ CDM cosmology with  $\Omega_m=0.3$ ,  $\Omega_\Lambda = 0.7$ , and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Furthermore, we adopt a Chabrier (2003) stellar initial mass function (IMF).

## 2 OBSERVATIONAL DATA

### 2.1 SAMI Galaxy Survey

SAMI is a multi-object IFS mounted at the prime focus of the 3.9m Anglo Australian Telescope (AAT). It employs 13 revolutionary imaging fibre bundles, or *hexabundles* (Bland-Hawthorn et al. 2011; Bryant et al. 2011; Bryant & Bland-Hawthorn 2012; Bryant et al. 2014) that are manufactured from 61 individual fibres with  $1'6$  angle on sky. Each hexabundle covers a  $\sim 15''$  diameter region on the sky, has a maximal filling factor of 75%, and is deployable over a  $1^\circ$  diameter field of view. All 819 fibres, including 26 individual sky fibres, are fed into the AAOmega dual-beamed spectrograph (Saunders et al. 2004; Smith et al. 2004; Sharp et al. 2006).

The SAMI Galaxy Survey (Croom et al. 2012; Bryant et al. 2015) has finished observations, targeting over 3000 galaxies covering a broad range in galaxy stellar mass ( $M_* = 10^8 - 10^{12} M_\odot$ ) and galaxy environment (field, groups, and clusters) between redshift  $0.004 < z < 0.095$ . Here we use internal data release v0.10.1 that contains 2528 galaxies. SAMI's angular fibre size results in spatial resolutions of 1.6 kpc per fibre at  $z = 0.05$ . Field and group targets were selected from four volume-limited galaxy samples derived from cuts in stellar mass in the Galaxy and Mass Assembly Survey (GAMA) G09, G12 and G15 regions (Driver et al. 2011). GAMA is a major campaign that combines a large spectroscopic survey of  $\sim 300,000$  galaxies carried out using the AAOmega multi-object spectrograph on the AAT, with a large multi-wavelength photometric data set. SAMI Galaxy Survey cluster targets were obtained from eight high-density cluster regions sampled within radius  $R_{200}$  with the same stellar mass limit as for the GAMA fields (Owers et al. 2017).

For the SAMI Galaxy Survey, the 580V and 1000R grating are used in the blue (3750-5750Å) and red (6300-7400Å) arm of the spectrograph, respectively. This results in a resolution of  $R_{\text{blue}} \sim 1810$  at 4800Å, and  $R_{\text{red}} \sim 4260$  at 6850Å (van de Sande et al. 2017b). In order to create data cubes with  $0'5$  spaxel size, all observations are carried out using a six to seven position dither pattern (Sharp et al. 2015; Allen et al. 2015).

All reduced data-cubes and stellar kinematic data products in the GAMA fields are available on: <https://datacentral.org.au/>, as part of the first and second SAMI Galaxy Survey data release (Green et al. 2017; Scott et al. 2018).

#### 2.1.1 Ancillary Data

For galaxies in the GAMA fields, we use the aperture matched  $g$  and  $i$  photometry from the GAMA catalogue (Hill et al. 2011; Liske et al. 2015), measured from reprocessed SDSS Data Release Seven (York et al. 2000; Kelvin et al. 2012), to derive  $g - i$  colours. For the cluster environment,

photometry from the SDSS (York et al. 2000) and VLT Survey Telescope ATLAS imaging data are used (Shanks et al. 2013; Owers et al. 2017). From the rest-frame  $i$ -band absolute magnitude and  $g - i$  color, stellar masses are derived by using the color-mass relation as outlined in Taylor et al. (2011). A Chabrier (2003) stellar IMF and exponentially declining star formation histories are assumed in deriving the stellar masses. For more details see Bryant et al. (2015).

Effective radii, ellipticities, and positions angles are derived using the Multi-Gaussian Expansion (MGE; Emsellem et al. 1994; Cappellari 2002) technique and the code from Scott et al. (2013) on imaging from the GAMA-SDSS (Driver et al. 2011), SDSS (York et al. 2000), and VST (Shanks et al. 2013; Owers et al. 2017). We define  $R_e$  as the semi-major axis effective radius, and the ellipticity of the galaxy within one effective radius as  $\epsilon_e$ , measured from the best-fitting MGE model. For more details, we refer to D'Eugenio et al. (in prep).

We use visual morphological classifications that are based on the SDSS DR9 and VST  $gri$  colour images; late- and early-types are divided according to their shape, presence of spiral arms and/or signs of star formation. Pure bulges are then classified as ellipticals (E) and early-types with disks as S0s. Similarly, late-types with only a disk component are classified as late-spirals, while disk plus bulge late types are early-spirals (for more details see Cortese et al. 2016).

#### 2.1.2 Stellar Kinematics

The stellar kinematic measurements for the SAMI Galaxy Survey are described in detail in van de Sande et al. (2017b). In summary, we use the penalized pixel fitting code (pPXF; Cappellari & Emsellem 2004; Cappellari 2017) assuming a Gaussian line of sight velocity distribution (LOSVD). The red arm spectral resolution is convolved to match the instrumental resolution in the blue. Both blue and red are then rebinned and combined onto a logarithmic wavelength scale with constant velocity spacing ( $57.9 \text{ km s}^{-1}$ ). MILES stellar library (Sánchez-Blázquez et al. 2006) spectra are used for deriving a set of radially varying optimal templates using the SAMI annular binned spectra. For each individual spaxel, pPXF is allowed to use the optimal templates from the annular bin in which the spaxel is located as well as the optimal templates from neighbouring annular bins. The uncertainties on the LOSVD parameters are estimated from 150 simulated spectra.

We use the quality criteria for SAMI Galaxy Survey data as described in van de Sande et al. (2017b): signal-to-noise ( $S/N > 3 \text{ \AA}^{-1}$ ),  $\sigma_{\text{obs}} > \text{FWHM}_{\text{instr}}/2 \sim 35 \text{ km s}^{-1}$  where the FWHM is the full-width at half-maximum,  $V_{\text{error}} < 30 \text{ km s}^{-1}$  ( $Q_1$  from van de Sande et al. 2017b), and  $\sigma_{\text{error}} < \sigma_{\text{obs}} * 0.1 + 25 \text{ km s}^{-1}$  ( $Q_2$ ).

We visually inspect all 2528 SAMI kinematic maps, and flag and exclude 87 galaxies with irregular kinematic maps due to nearby objects or mergers that influence the stellar kinematics of the main object. Another 533 galaxies are excluded where the radius out to which we can accurately measure the stellar kinematics or  $R_e$  is less than the half-width at half-maximum of the PSF ( $\text{HWHM}_{\text{PSF}}$ ). Furthermore, we set the observational mass limit for stellar kinematic measurement at  $M_* = 5 \times 10^9 M_\odot$  or  $\log(M_*/M_\odot) = 9.7$ , simi-

lar to the simulations, which excludes another 332 galaxies. Finally, throughout the paper, we only use galaxies when where we can accurately derive  $V/\sigma$  out to one  $R_e$  (see Section 4.7). This brings the final number of galaxies from the SAMI Galaxy Survey to 1558.

We note that while the total number of targeted galaxies in the GAMA fields is significantly higher as compared to cluster targets, due to higher stellar mass limit and lower redshift range of the cluster sample, the stellar kinematic success rate in clusters is significantly higher. Hence, we end up with a relatively large fraction of cluster galaxies in the final SAMI stellar kinematic sample ( $\sim 47$  percent).

### 2.1.3 Stellar Population Age

We derive luminosity-weighted stellar population ages using 11 Lick indices in the SAMI blue spectral range following the method outlined in Scott et al. (2017). Lick indices are converted into single stellar population (SSP) equivalent age using stellar population synthesis models (Schiavon 2007) that predict Lick indices as a function of log Age, metallicity  $[Z/H]$ , and  $[\alpha/Fe]$ . We then determine the SSP that best reproduces the measured Lick indices using a  $\chi^2$  minimisation approach. Typical uncertainties are  $\pm 0.15$  dex in log Age. Because the stellar population parameters are derived from using individual Lick indices rather than full spectral fitting, our results are relatively insensitive to dust.

## 2.2 ATLAS<sup>3D</sup> Survey

We use a combined sample of 260 early-type galaxies from the SAURON survey (de Zeeuw et al. 2002) and ATLAS<sup>3D</sup> Survey (Cappellari et al. 2011) that were observed with the SAURON spectrograph (Bacon et al. 2001). The SAURON survey adopted an instrumental spectral resolution of 4.2Å FWHM ( $\sigma_{\text{instr}} = 105 \text{ km s}^{-1}$ ) and cover the wavelength range of 4800-5380Å. ATLAS<sup>3D</sup> galaxies were observed with a higher resolution of 3.9Å FWHM ( $\sigma_{\text{instr}} = 98 \text{ km s}^{-1}$ ). The data were Voronoi binned (Cappellari & Copin 2003) with a target signal-to-noise per bin of 40. The stellar kinematics were extracted using pPXF with stellar templates from the MILES stellar library (see Cappellari et al. 2011).

We use the publicly available unbinned data cubes (V1.0<sup>1</sup>) and the 2D Voronoi binned stellar kinematic maps (Emsellem et al. 2004; Cappellari et al. 2011)<sup>2</sup>. We exclude galaxy NGC 0936 from the sample because no unbinned data is available. Circularised size measurements are taken from Cappellari et al. (2011), corrected to semi-major axis effective radii using the global ellipticities from Krajnović et al. (2011) ( $R_e = R_{e,c}/\sqrt{1-\epsilon}$ ). We use ellipticities at one effective radius from Emsellem et al. (2011), and position angles from Krajnović et al. (2011). Stellar masses are calculated from the  $r$ -band luminosity and mass-to-light ratio as presented in Cappellari et al. (2013a,b), converted to a Chabrier (2003) IMF. Visual morphologies (T-type) are from Cappellari et al. (2011). We obtained luminosity-weighted stellar population ages from Table 3 of McDermid et al. (2015)

that are based on Lick indices measurements and single stellar population models from Schiavon (2007). As described in Appendix B, we subtract 0.23 dex from all ATLAS<sup>3D</sup> stellar ages to correct a median offset from the SAMI sample. Above a stellar mass limit of  $\log(M_\star/M_\odot) = 9.7$ , the ATLAS<sup>3D</sup> sample contains 244 galaxies.

## 2.3 CALIFA Survey

We use a sample of 294 CALIFA galaxies with a wide range in morphology and kinematic properties from Falcón-Barroso et al. (2017). The CALIFA IFS data were observed with the PMAS instrument (Roth et al. 2005), a  $74'' \times 64''$  hexagonal fibre spectrograph, mounted at the 3.5m telescope of the Calar Alto observatory. Similar to ATLAS<sup>3D</sup>, the CALIFA data are Voronoi binned to obtain spatial bins with an approximate S/N of 20 per pixel. Stellar kinematic measurements were derived from the Voronoi binned spectra with wavelength range between 3400-4750Å and spectral resolution of 2.3Å (V1200 grating; Husemann et al. 2013). Velocities  $V$  and velocity dispersions  $\sigma$  are extracted with the pPXF code in combination with 330 stellar templates selected from the Indo-U.S. spectral library (Valdes et al. 2004). The 2D kinematic maps are publicly available as part of the CALIFA DR3 (Sánchez et al. 2016b)<sup>3</sup>.

We use the stellar masses, visual morphologies, semi-major axis effective radii, position angles, and ellipticities at one effective radius as presented in Table 1 from Falcón-Barroso et al. (2017; see also Walcher et al. 2014). Luminosity-weighted stellar population ages are gathered from table C.2 of González Delgado et al. (2015). These ages are estimated from full spectral fitting using the spectral base GMe models that are a combination of SSP spectra provided by Vazdekis et al. (2010) and González Delgado et al. (2005). Measurements within one effective radius are not available, instead we derive the average of the "central" ([0]) and "at 1 half-light-radius" mean stellar ages. Finally, we note that CALIFA galaxies are selected based on angular isophotal diameter ( $45'' \leq D_{25} \leq 80''$  Walcher et al. 2014); this biases the sample towards galaxies that are more inclined and have higher ellipticities. Above a stellar mass limit of  $\log(M_\star/M_\odot) = 9.7$ , the ATLAS<sup>3D</sup> sample contains 257 galaxies.

There are six galaxies in the CALIFA survey that overlap with ATLAS<sup>3D</sup>. Similar to Falcón-Barroso et al. (2017), we find an excellent agreement between the dynamical measurements from both surveys. However, when comparing the stellar mass estimates, we notice that 5/6 CALIFA stellar masses are on average higher by  $\sim 0.22$  dex as compared to ATLAS<sup>3D</sup>. To investigate this offset further, we check for possible mismatches in the size-stellar mass plane and velocity dispersion-stellar mass plane. We find that the full CALIFA dataset is on average higher by  $\sim 0.2$  dex in stellar mass as compared to GAMA, SAMI, and ATLAS<sup>3D</sup> data in the size-mass diagram, and to SAMI and ATLAS<sup>3D</sup> in the  $\sigma_e$ -mass diagram. While this difference could be explained by a different IMF (e.g., Salpeter 1955 as compared to Chabrier 2003) the CALIFA DR3 documentation indicates that a Chabrier (2003) was used to derive the stellar masses. While

<sup>1</sup> <http://www-astro.physics.ox.ac.uk/atlas3d/>

<sup>2</sup> Unbinned stellar kinematic measurements are not available.

<sup>3</sup> <http://califa.caha.es>

other CALIFA stellar mass estimates exist (e.g., [González Delgado et al. 2015](#); [Sánchez et al. 2016a](#)), when comparing these stellar masses to the overlapping ATLAS<sup>3D</sup> galaxies, or with dynamical mass measurements (see Section 4.6), we find that they do not provide a better match. As the main idea behind using different surveys is to create a more homogeneous data-set, we therefore decide to use the [Falcón-Barroso et al. \(2017\)](#) stellar masses, but we subtract 0.2 dex to correct for the offset to GAMA, SAMI, and ATLAS<sup>3D</sup>.

## 2.4 MASSIVE Survey

The volume-limited MASSIVE IFS survey specifically targeted the  $\sim 100$  most massive early-type galaxies within a distance of 108 Mpc. IFS observations were done with the Mitchell fibre Spectrograph ([Hill et al. 2008](#)) with a  $107'' \times 107''$  field of view, on the 2.7m Harlan J. Smith Telescope at McDonald Observatory. The data are spatially combined using a circular binning scheme to reach a target S/N of 20. The wavelength range is 3650-5850Å and the average spectral resolution is 5Å (FWHM) with a dependence on wavelength and spatial position.

[Veale et al. \(2017b\)](#) use PPF in combination with the MILES stellar library to extract six-moment kinematics, i.e., they fit for  $V$ ,  $\sigma$ , and  $h_3 - h_6$ . Here,  $h_3 - h_6$  are the weights of Gauss-Hermite polynomials that are used to model the deviations from a Gaussian LOSVD. Stellar kinematic maps are not publicly available; instead we use the kinematic values as presented in Table 1 of [Veale et al. \(2018\)](#). This will be explained in more detail in Section 4.7. Stellar velocity dispersions and ellipticities are obtained from [Veale et al. \(2017b,a\)](#). Ellipticities are derived from a "super-coadd" isophote, not within one effective radius ([Ma et al. 2014](#)). Effective radii are from [Ma et al. \(2014\)](#) Table 3 (NASA-Sloan Atlas where available or 2MASS corrected using their Eq. 4).

Following [Ma et al. \(2014\)](#), we use absolute  $K$ -band magnitudes to estimate stellar masses. However, rather than deriving stellar masses from a relation based on Jeans Anisotropic Modelling (JAM) mass-to-light ratios ([Cappellari et al. 2013a](#)), we use stellar population model-based mass-to-light ratios ( $\log(M/L)_{\text{Salp}}$ ) from [Cappellari et al. \(2013b\)](#), converted to a [Chabrier \(2003\)](#) IMF:

$$\log_{10}(M_{\star}) = 10.39 - 0.46(M_K + 23). \quad (1)$$

This different choice as compared to [Ma et al. \(2014\)](#) is motivated by the need for a homogeneous sample where all stellar masses are calculated in the same way. Above a stellar mass limit of  $\log(M_{\star}/M_{\odot}) = 9.7$ , the total number of MASSIVE galaxies with stellar kinematic measurements is 85.

## 2.5 Summary of Observational Data

The combined data from the SAMI Galaxy Survey, ATLAS<sup>3D</sup>, CALIFA, and MASSIVE Survey yields a total of 2144 galaxies with stellar kinematic measurements above a stellar mass of  $\log(M_{\star}/M_{\odot}) > 9.7$ . Approximately  $\sim 40$  percent of the galaxies reside in high-density cluster environments, and  $\sim 66$  percent are visually classified to have early-type morphology (E and S0-type).

## 3 SIMULATION DATA

### 3.1 EAGLE and HYDRANGEA Simulations

The EAGLE project (Evolution and Assembly of GaLaxies and their Environments; [Schaye et al. 2015](#); [Crain et al. 2015](#); [McAlpine et al. 2016](#)), is a large set of cosmological hydrodynamic simulations that are publicly available<sup>4</sup>. In this paper, we use the reference model Ref-L100N1504 with a volume of (100 Mpc)<sup>3</sup> co-moving. EAGLE and HYDRANGEA adopt the Planck Collaboration XVI 2014 cosmological parameters ( $\Omega_m=0.307$ ,  $\Omega_{\Lambda} = 0.693$ ,  $H_0 = 67.77 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ). The dark matter particle mass is  $9.7 \times 10^6 M_{\odot}$ , the initial gas particle mass is  $1.81 \times 10^6 M_{\odot}$ , and the typical mass of a stellar particle is similar to the gas particle mass. The reference model was calibrated to match the  $z \sim 0.1$  stellar mass function and the observed relation between stellar mass and black-hole mass. The  $z \sim 0.1$  size-mass relation was also used as a guide to reject some stellar feedback models that led to too compact galaxies, despite reproducing the stellar mass function (see [Crain et al. 2015](#) for details).

To provide the best global environment match to the SAMI Galaxy Survey sample, we combine EAGLE with HYDRANGEA that consists of 24 cosmological zoom-in simulations of galaxy clusters and their environments ([Bahé et al. 2017](#)). HYDRANGEA is part of the larger Cluster-EAGLE project ([Barnes et al. 2017](#)). Cluster-EAGLE is similar to EAGLE but with different parameter values for the active galactic nuclei (AGN) feedback model, to make it more efficient. To reproduce the same ratio of field, group, and cluster galaxies as in observations, we select all HYDRANGEA galaxies that are in groups or clusters with mass greater than  $\log(M_{\star}/M_{\odot})_{\text{group}} > 13.85$ . We note that this group mass limit for HYDRANGEA is lower than the adopted mass for the SAMI Galaxy Survey cluster sample of  $\log(M_{\star}/M_{\odot})_{\text{group}} > 14.25$ . This lower limit was adopted to reach a fraction of  $\sim 40$  percent of galaxies that are in the highest-density environments, similar to observations.

We use the kinematic measurements as described in [Lagos et al. \(2018b\)](#), corrected to  $H_0 = 70.0 \text{ km s}^{-1} \text{ Mpc}^{-1}$  which was adopted for the observations. In summary, we extract effective radii, ellipticities, line-of-sight velocities and velocity dispersions, adopting techniques that closely match the observations. First, each galaxy is projected onto a 2D plane that is observed under two different inclinations: an edge-on view, and a random orientation (seen through the simulation  $z$ -axis) to mimic observations. In this 2D projection, we create a grid of pixels with size 1.5 kpc (proper), and construct an  $r$ -band luminosity weighted velocity distribution for each pixel. This LOSVD is fitted with a Gaussian function to estimate  $V$  and  $\sigma$ ; the rest-frame velocity is defined at the centre of the galaxy potential.

$r$ -band luminosities of star particles are derived by combining the age and metallicity of those star particles assuming a [Chabrier \(2003\)](#) IMF combined with a [Bruzual & Charlot \(2003\)](#) stellar population synthesis model. The  $r$ -band luminosities are then obtained by convolving this model with an SDSS  $r$ -band bandpass. We do not include the effect of dust on the  $r$ -band luminosities. While dust obscuration will have a larger impact on the measurements

<sup>4</sup> <http://icc.dur.ac.uk/Eagle/database.php>

in late-type galaxies than in early-types and depends on inclination, as the differences between mass and luminosity weighted quantities are relatively small (see Appendix A), we do not expect dust to change our results significantly. For a study on the impact of dust on galaxy colours in EAGLE, we refer to [Trayford et al. \(2017\)](#).

The effective radius is defined as  $R_e$ , the projected  $r$ -band half-luminosity radius, measured as the average from the  $xy$ ,  $xz$ , and  $yz$  projections. Ellipticities are calculated within one effective radius using the projected particle positions and particle luminosities using Eqs. 1 to 3 of [Lagos et al. \(2018b\)](#). The position angle of the major axis are defined in Eq. 4 of [Lagos et al. \(2018b\)](#). Stellar ages were calculated as an  $r$ -band luminosity weighted stellar age using all particles within  $r_{50}$ .

Following [Lagos et al. \(2018b\)](#), we adopt a lower mass limit of  $M_\star = 5 \times 10^9 M_\odot$ , to ensure that the simulated measurements converge to better than 10 percent. Example maps of the luminosity, velocity, and velocity dispersion are shown in Fig 1 of [Lagos et al. \(2018b\)](#). Above the adopted stellar mass limit, the combined EAGLE and HYDRANGEA sample contains 8,982 galaxies.

### 3.2 HORIZON-AGN Simulations

The cosmological hydrodynamic HORIZON-AGN simulations are described in detail in [Dubois et al. \(2014\)](#). In short, the simulation that we use here is run within a box with a volume of  $(142 \text{ Mpc})^3$  co-moving. HORIZON-AGN adopts a cosmology compatible with the Wilkinson Microwave Anisotropy Probe 7 cosmology ( $\Omega_m=0.272$ ,  $\Omega_\Lambda = 0.728$ ,  $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ; [Komatsu et al. 2011](#)). The dark matter particle mass is  $8 \times 10^7 M_\odot$ . The hydrodynamics are computed on a grid, adaptively refining to the local density following a quasi Lagrangian scheme ([Teyssier 2002](#)). Cells are 1kpc wide at maximal refinement level. The adopted resolution is such that the typical mass of a stellar particle is  $2 \times 10^6 M_\odot$ . The simulation implements the formation and evolution of black holes, and black holes can grow by gas accretion. The HORIZON-AGN project uses two modes of AGN feedback, while EAGLE adopts one. The tuning approach in HORIZON-AGN differs from EAGLE. Only the local black hole mass to stellar mass relation is tuned so as to ensure the predictive aspect of resulting statistical properties in the simulation.

Structural, stellar population, and stellar kinematic measurements are measured in a similar way as was done for the EAGLE simulation, corrected to  $H_0 = 70.0 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . We present a brief summary here, whereas a more detailed description will be presented in C. Welker et al. (in prep.). Ages and ellipticities are computed directly from the star particles with  $r$ -band luminosity weighting. Ellipticities are derived from the eigenvalues of the 2D luminosity weighted inertia tensor of the star particles (positions are projected on a plane orthogonal to the line of sight). All star particles within one half-luminosity radius are used in the calculation. All other quantities are computed on mock spaxels. For each galaxy, its star particles are projected on a plane orthogonal to the line of sight and sorted in a two dimensional spatial grid covering all star particles within one half luminosity radius, with a fixed pixel width of 1.5 kpc. In each pixel, we compute the average  $r$ -band luminosity, and we fit the aver-

age line-of-sight velocity and dispersion. We then compute the luminosity weighted average over all the pixels with more than 10 particles. The HORIZON-AGN sample contains 32,452 galaxies with stellar mass greater than  $M_\star = 5 \times 10^9 M_\odot$ .

### 3.3 MAGNETICUM Simulations

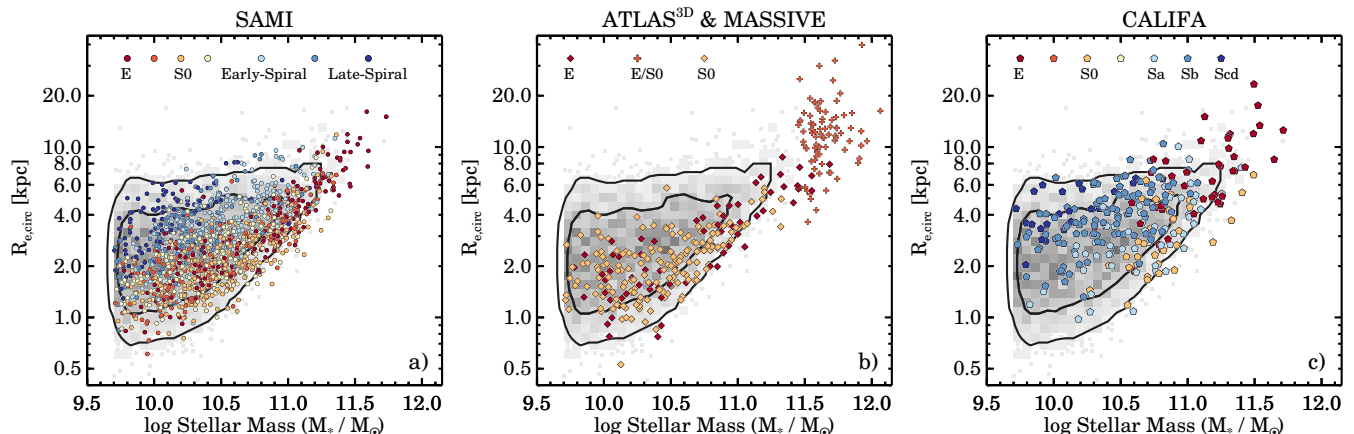
The third set of cosmological hydrodynamical simulations that we will use are the *Magneticum Pathfinder*<sup>5</sup> simulations, hereafter simply MAGNETICUM (see Dolag et al., in prep and [Hirschmann et al. 2014](#) for more details on the simulation). We use the data from the medium-sized cosmological box (Box 4) with a volume of  $(48 \text{ Mpc})^3$  co-moving at the ultra high resolution level. MAGNETICUM adopts a cosmology compatible with the Wilkinson Microwave Anisotropy Probe 7 cosmology ( $\Omega_m=0.272$ ,  $\Omega_\Lambda = 0.728$ ,  $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ; [Komatsu et al. 2011](#)). The dark matter and gas particles have masses of respectively  $3.7 \times 10^7 M_\odot$  and  $7.3 \times 10^6 M_\odot$ , and each gas particle can spawn up to four stellar particles.

Structural, stellar population, and stellar kinematic measurements are described in [Schulze et al. \(2018\)](#), here corrected to  $H_0 = 70.0 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . However, here we use luminosity weighted quantities as compared to the mass-weighted quantities in [Schulze et al. \(2018\)](#).  $r$ -band luminosities are derived using an identical method that was used for the EAGLE data. The effective radius of a galaxy is estimated from determining the radius of a sphere that contain half of the total  $r$ -band luminosity. Note that this approach is different from EAGLE and HORIZON-AGN where we used projected 2D sizes, and the MAGNETICUM could therefore be larger by approximately  $\sim 0.1 \text{ dex}$ . Kinematic maps are constructed from the mean projected velocity along a line-of-sight and the velocity dispersion is derived from the standard deviation of the particle velocities. Similar to some observational surveys (e.g., ATLAS<sup>3D</sup>, CALIFA), Voronoi tessellation is adopted to avoid low-numbers of particles along a line-of-sight. Ellipticities for a given projection are derived following the definition of [Cappellari et al. \(2007\)](#), but using an iterative process. First,  $\epsilon$  is calculated from all particles within a circular aperture radius of  $1.5R_e$ . Then this process is repeated using a new aperture with the previously determined ellipticity  $\epsilon$  containing the same stellar mass, until the estimate for  $\epsilon$  converges. This technique differs slightly from the approach taken for EAGLE and HORIZON-AGN, where no iterative approach was used. Stellar ages were calculated as an  $r$ -band luminosity weighted stellar age using all particles within  $r_{50}$ . Because the dark matter particle mass in MAGNETICUM is significantly higher than in EAGLE we adopt a higher mass limit of  $M_\star = 1 \times 10^{10} M_\odot$ . Above this stellar mass limit, the MAGNETICUM sample contains 2073 galaxies.

### 3.4 Summary of Data from Simulations

We have constructed a large sample of galaxy mock-observations from four major hydrodynamical cosmological simulations: EAGLE and HYDRANGEA, HORIZON-AGN, and MAGNETICUM. In all simulations, we have used measurement techniques as close to observational measurements

<sup>5</sup> [www.magneticum.org](http://www.magneticum.org)



**Figure 1.** Comparison of GAMA and IFS observational data in the size-mass diagram. We show the density of galaxies from the GAMA survey as grey squares, where darker grey means higher density of galaxies. The grey contours enclose 68 and 95 percent of the GAMA data. The individual symbols show the IFS data from the SAMI Galaxy Survey (left), ATLAS<sup>3D</sup> (middle; diamond symbols) and MASSIVE (middle; plus symbols), and CALIFA (right). IFS data are colour coded by their visual morphological classification as indicated by the legends. We recover the well-known morphology trend where below  $\log(M_*/M_\odot) \sim 11$  early-type galaxies tend to be smaller than late-type galaxies, whereas early-type galaxies start to match the sizes of late-type galaxies at the highest stellar masses ( $\log(M_*/M_\odot) > 11.3$ ).

techniques as possible, but minor differences still exist. Furthermore, we use simulations with different particle resolution and with different volume sizes: in EAGLE the dark particle mass is  $m_{\text{dm}} = 9.7 \times 10^6 M_\odot$  in a  $(100 \text{ Mpc})^3$  co-moving volume, for HORIZON-AGN we have  $m_{\text{dm}} = 8 \times 10^7 M_\odot$  and  $(142 \text{ Mpc})^3$  co-moving volume, and for MAGNETICUM  $m_{\text{dm}} = 3.7 \times 10^7 M_\odot$  in  $(48 \text{ Mpc})^3$  co-moving volume. We stress that these simulations adopt different philosophies for calibrating to and reproducing observational results. Where some are “made to match”, others are “made to bridge” (i.e., calibrated on larger-scale observed statistical properties, versus calibrated on ensemble average smaller-scale simulations). Therefore, our main aim of the paper is to identify key areas of success and tension, not to determine which simulations provides the closest or “best” match to observations.

## 4 COMPARING OBSERVATIONS AND SIMULATIONS

In the following section we will compare fundamental galaxy parameters obtained from simulations and observations using similar measuring techniques. We will look at structural parameters, such as effective radius and ellipticity, but also at dynamical properties like aperture velocity dispersion and  $(V/\sigma)_e$ , the average ratio of the velocity and velocity dispersion within one effective radius. The main goal of this section is to see where simulations and observations agree or disagree. Section 5 and 6 will be devoted to understanding where the differences are coming from and what lessons can be learned from this.

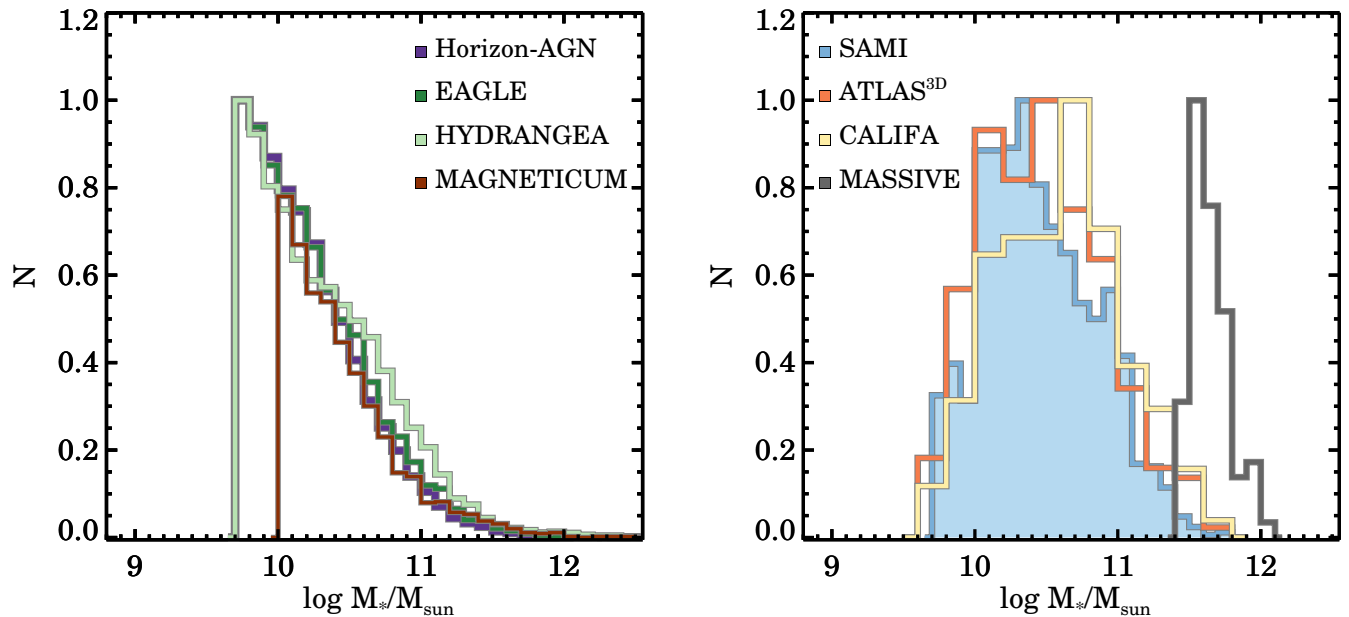
### 4.1 Observational Biases

Most observational selection effects are well understood and easily reproducible, but the combination of four different surveys with four different sets of selection criteria makes

the comparison with the simulations challenging. There is not a single, simple selection function that encompasses all the biases of the observations. The strongest observational bias is as a function of stellar mass: as galaxies decrease in stellar mass their total luminosity decreases, which makes it increasingly hard to obtain the targeted S/N per galaxy spaxel. Therefore, for the sake of simplicity, we only apply a stellar-mass selection to the simulated data. This will remove the strongest bias, but more subtle effects may arise from the other selection criteria (for details see Section 2). Note, for example that Cañas et al. (2018) find an offset between the stellar mass function from observations (Moustakas et al. 2013; Wright et al. 2017) and Horizon-AGN (see also Kaviraj et al. 2017). This implies that with our mass-matching technique, on average low-mass galaxies in HORIZON-AGN may be selected in lower-mass halos than in EAGLE. For MAGNETICUM a comparison of the stellar-to-halo mass relation was made in Teklu et al. (2017), and their simulated galaxies agree qualitatively with the different observations. However, possible biases might arise in MAGNETICUM due to the smaller volume box used that could lead to lower density environments being probed on average compared to EAGLE, HORIZON-AGN and the observations. As the evolution of the angular momentum in galaxies is sensitive to halo growth, this could add to the biases in the comparison between the simulated and observed dynamical measurements.

To reveal possible observational biases as a function of stellar mass, in Fig. 1 we present the size-mass relation of the observational samples compared to an unbiased sample from GAMA at  $z < 0.1$ . The data are colour coded by visual morphology. Note that the MASSIVE sample only contains ellipticals (E) and lenticulars (S0), however, no individual galaxy classifications were available to us.

A size-mass dependence on morphology is well known; at lower stellar mass ( $\log(M_*/M_\odot) < 10$ ) early-type galaxies tend to be smaller than late-type galaxies (e.g., Shen et al. 2003; Lange et al. 2016). The early-type size-mass relation



**Figure 2.** Normalised distribution of galaxy stellar masses from simulations (left panel) and observations (right panel). Simulation data from HORIZON-AGN are shown in purple, EAGLE in green, HYDRANGEA light-green, and MAGNETICUM in brown. Observational data from SAMI is shown in blue, ATLAS<sup>3D</sup> in orange, CALIFA in yellow, and MASSIVE in grey. Because the SAMI sample is the largest observational sample, for clarity the histogram is shown with a colour fill. Due to observational limitations and selection criteria, the number of galaxies below  $\log(M_{\star}/M_{\odot}) < 10$  rapidly decreases in the observational surveys. To overcome this bias as a function of stellar mass, data from simulations are mass-matched to the observational data (see Section 4.2).

is steeper as compared to the relation for late-type galaxies, and at high stellar masses ( $\log(M_{\star}/M_{\odot}) > 11$ ) early-type galaxies start to become larger. This trend with morphology is clearly visible in the SAMI Galaxy Survey data, in particular between  $10 < \log(M_{\star}/M_{\odot}) < 11$ . Similarly, because the ATLAS<sup>3D</sup> survey consists of early-types only, galaxies are on average smaller than the GAMA galaxies.

Below  $\log(M_{\star}/M_{\odot}) < 10.5$ , the fraction of late-type galaxies strongly increases. Below  $\log(M_{\star}/M_{\odot}) < 10$  we find a bias towards more compact galaxies, as there are few SAMI galaxies above the  $1\text{-}\sigma$  GAMA contour line around 4kpc. The CALIFA sample becomes strongly dominated by late-type galaxies at  $\log(M_{\star}/M_{\odot}) < 10.5$  and CALIFA galaxies are on average slightly larger than the GAMA comparison sample at this stellar mass. The sizes of the MASSIVE galaxies appear to lie well above the different surveys. We will discuss this further in Section 4.6.

While individual surveys show small morphological biases, we can diminish the effect of these biases by combining the data from the SAMI Galaxy Survey, ATLAS<sup>3D</sup>, and CALIFA into one sample. We note that combining the different surveys does not completely remove the sample biases, but lessens the biases of the individual surveys.

#### 4.2 Mass Distribution and Sample Matching

We now describe the mass selection of simulated galaxies in order to match the observational sample. We start by showing the normalised mass distributions of the simulated and observational data in Fig. 2. Note that due to the different stellar mass range of the MAGNETICUM sample, we have normalised the stellar mass distribution of MAGNETICUM to

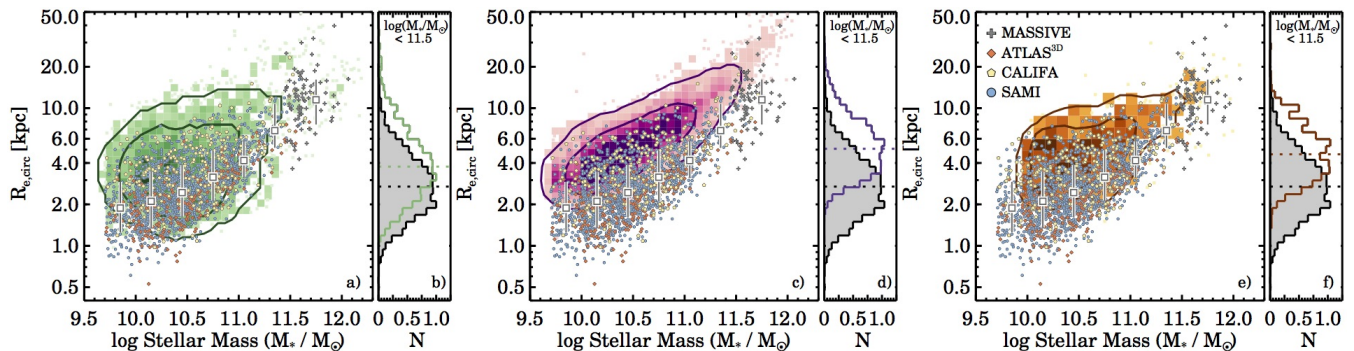
have a peak of 0.78, rather than 1.0 that was adopted for EAGLE, HYDRANGEA, and HORIZON-AGN. This way, we can better compare the shape of the distribution between the different simulations.

We find a close match between the mass distribution of EAGLE, HORIZON-AGN, and MAGNETICUM (left panel), although the HORIZON-AGN distribution is slightly higher at low stellar mass ( $\log(M_{\star}/M_{\odot}) < 10.5$ ). The clear difference between the EAGLE and HYDRANGEA stellar mass distribution is due to the HYDRANGEA cluster environment. For the observations (right panel), we see that the shape of the distributions for SAMI and ATLAS<sup>3D</sup> closely match, with a median stellar mass of  $\log(M_{\star}/M_{\odot})=10.4$ . The CALIFA distribution is skewed towards slightly higher stellar mass (median  $\log(M_{\star}/M_{\odot})=10.6$ ), and the MASSIVE survey is a clear but intended outlier, with a narrow distribution at very high-stellar mass (median  $\log(M_{\star}/M_{\odot})=11.6$ ).

As the simulated and observed stellar mass functions differ in shape, we cannot directly compare parameters from the simulations and observations without introducing a bias caused by trends with stellar mass. Therefore, to remove this bias, we will perform a mass-matching by randomly selecting simulated galaxies as a function of stellar mass. The number of galaxies in the EAGLE, HYDRANGEA, and HORIZON-AGN simulations exceeds the number of galaxies in the observed sample. Thus, we can select many more galaxies from these simulations than there are in the observed sample. However, in MAGNETICUM, the number of galaxies is less than in the observed sample, and in order to mass-match the sample, we have to select an even lower number of galaxies.

Despite the differences in model parameters between EAGLE and HYDRANGEA, we combine the two models because





**Figure 3.** Comparison of simulated and observational data in the size-mass diagram. In panel a), we show the density of galaxies from the combined EAGLE and HYDRANGEA simulations as green squares, where darker green means higher density of galaxies. The green contours enclose 68 and 95 percent of the simulated data. Observational data from the SAMI Galaxy Survey is shown as blue circles, ATLAS<sup>3D</sup> as orange diamonds, CALIFA as yellow pentagons, and MASSIVE as grey pluses. The grey and white squares show the median of the observed sample in mass bins, and the vertical lines show the 16th and 84th percentile of the observed distribution. In panel b) we show the effective radius-distribution of EAGLE<sup>+</sup> (green) and observational data (black), where the dotted lines show the median in effective-radius. Data from the MASSIVE survey is not included in this panel (see Section 2). The EAGLE<sup>+</sup> size-mass relation matches reasonably well with observational data, although the median size of simulated galaxies is larger by 42 percent as compared to our observed sample (median of 3.82 kpc versus 2.67 kpc, respectively). For HORIZON-AGN (panel c-d), the size-mass relation is offset from observational data; at all stellar masses simulated galaxies are too large. The median size is 5.03 kpc as compared to 2.67 kpc for the observations, but the offset is similar at all stellar masses. The MAGNETICUM size-mass relation (panel e-f) has a similar slope and spread, but the average galaxy size is too large (median of 4.68 kpc versus 2.67 kpc in observations), except for the most massive galaxies.

it was shown in Lagos et al. (2018b) that both models have a similar span in the  $\lambda_R$  (spin-parameter proxy) - ellipticity space, and the differences were due to stellar mass sampling. In observations, the ratio of field and group galaxies compared to cluster galaxies in the observations is approximately 40 percent. As mentioned in Section 3.1, in order to match this number in the EAGLE and HYDRANGEA simulations, we only select HYDRANGEA galaxies that are in groups or clusters with mass greater than  $\log(M_\star/M_\odot)_{\text{group}} > 13.85$ . From now on, we will refer to this joined EAGLE and HYDRANGEA sample as EAGLE<sup>+</sup>.

For the mass-matching, we first determine the number of galaxies in both the simulated and observational datasets, in stellar mass bins with a 0.15 dex width, starting at  $\log(M_\star/M_\odot) = 9.7$ . In every mass-bin we then compare the number of simulated versus observed galaxies, and calculate the ratio between the two. For example, in the stellar mass bin of  $10.0 < \log(M_\star/M_\odot) < 10.15$ , we find 1227 EAGLE<sup>+</sup> galaxies, whereas the combined observational dataset contains 279 galaxies. Thus the ratio of simulated to observed galaxies is a factor of 4.40. Over the entire stellar mass range, we find that the lowest ratio of simulated to observed galaxies is 1.96 in the stellar mass range of  $10.9 < \log(M_\star/M_\odot) < 11.05$  (320 and 163 galaxies in EAGLE<sup>+</sup> and observations, respectively). Similarly, for HORIZON-AGN the lowest ratio is 4.56 at  $11.2 < \log(M_\star/M_\odot) < 11.35$ , and for MAGNETICUM the lowest ratio is 0.51 at  $10.9 < \log(M_\star/M_\odot) < 11.05$ .

The lowest ratio then sets the number of galaxies that we can randomly sample in each mass bin from the simulations. That is, for *every* mass bin in the EAGLE<sup>+</sup> (HORIZON-AGN, MAGNETICUM) simulation, we randomly select 1.96 (4.56, 0.51) times as many galaxies as there are observed. This way, the mass-distribution of the simulations will be identical to the observations, while also maintaining the largest number of simulated galaxies to which to compare

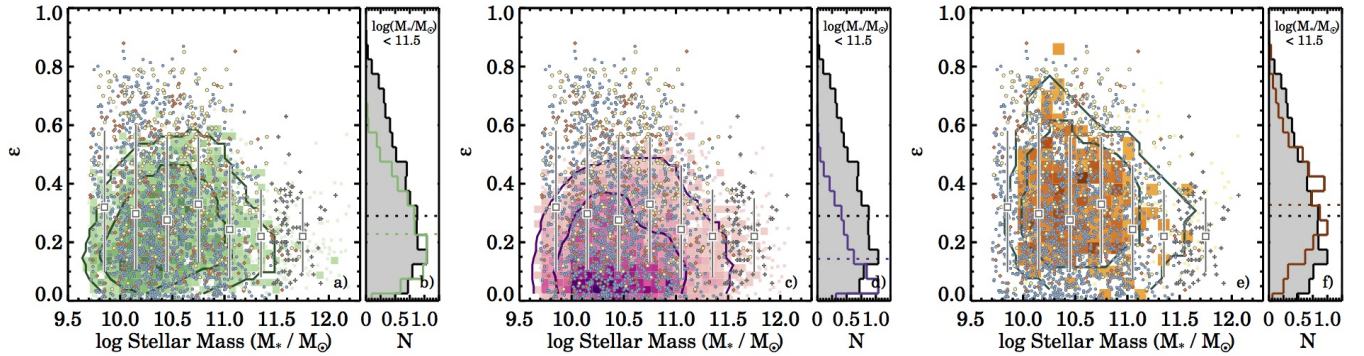
to. As we only do a single random draw, the choice of the random seed may potentially impact our results when the sample to draw from gets small (e.g., towards high stellar mass). We check and confirm that by using different random seeds none of our conclusions change.

We set an upper limit for the mass-matching at  $\log(M_\star/M_\odot) = 11.5$ , where the number of simulated galaxies is low as compared to the observations. Above a stellar mass of  $\log(M_\star/M_\odot) = 11.5$ , there are 89 galaxies in the combined surveys including MASSIVE, while there are 109 in EAGLE<sup>+</sup>, 166 in HORIZON-AGN, and 38 in MAGNETICUM. If we were to include the MASSIVE sample in the mass matching, the overall sampling factor would be  $\sim 2$  or less. Thus, to keep the sampling factor as high as possible for the combined observed sample we limit the mass matching to  $\log(M_\star/M_\odot) < 11.5$ . Effectively, this means that we combine the SAMI, ATLAS<sup>3D</sup>, and CALIFA measurements into one sample.

Combining the different surveys also helps to homogenise coverage for several observed parameters (see e.g., Fig. 1). We note that our observational sample is biased towards early-type galaxies. Mass-matching will not correct for possible morphological differences between the observed and simulated samples, but is beyond the scope of this paper to try and correct for such second order effects. Finally, because the MASSIVE Survey sample only contains galaxies with stellar mass  $\log(M_\star/M_\odot) > 11.5$ , we will describe the comparison with the MASSIVE sample separately in the following sections.

### 4.3 Size-Mass

We start the comparison between observations and simulations with one of the most fundamental galaxy relations, between the stellar mass and effective radius (Fig 3). Due to



**Figure 4.** Comparison of simulated and observational data in the ellipticity-mass diagram. Symbols are the same as in Fig. 3. We find a good match between the ellipticities from EAGLE<sup>+</sup> and observations, although there are too few flattened high  $\epsilon$  galaxies in EAGLE<sup>+</sup>. Simulated HORIZON-AGN galaxies are too round by a large factor, except for the highest stellar masses. Almost no EAGLE<sup>+</sup> galaxies with  $\epsilon > 0.6$  or HORIZON-AGN galaxies with  $\epsilon > 0.5$  exist. At low stellar masses, MAGNETICUM provides a significantly better match at high  $\epsilon$  as compared to the other simulations, yet the number of observed round galaxies  $\epsilon > 0.2$  is lower as compared to the observed sample.

the availability of high-spatial resolution *Hubble Space Telescope* imaging campaigns (e.g., COSMOS, CANDELS, etc; Capak et al. 2007; Koekemoer et al. 2007; Grogin et al. 2011; Koekemoer et al. 2011), the size-mass relation is extremely well studied out to high-redshift and offers invaluable insight into how galaxies grow over time (e.g., van der Wel et al. 2014, and references therein).

The match between EAGLE<sup>+</sup> and observations is good (Fig. 3a), but EAGLE<sup>+</sup> galaxies have sizes that are on average too large. While the EAGLE simulations are not directly calibrated to match disk-type galaxies, models were rejected that produced galaxies that were far too small (Crain et al. 2015), so a good match here is not entirely unexpected (see Schaye et al. 2015). However, the calibration was performed with mass-weighted sizes, which are typically smaller than luminosity-weighted (see Appendix A). The addition of the HYDRANGEA data increases the number of early-type galaxies, and thus adds more data to the bottom-right side of the size-mass relation. Nonetheless, at fixed stellar mass, we observe the spread in the EAGLE<sup>+</sup> sizes to be larger as compared to the observed sample. In Fig. 3b, we find an offset between the median effective radius of the simulations versus the observations (3.82 kpc versus 2.67 kpc, respectively). From bootstrapping the distributions a thousand times, we find that the uncertainty on the medians are small (1-2 percent). The number of observed galaxies is largest between  $10 < \log(M_*/M_\odot) < 10.5$ , which is also where the observational bias towards early-type galaxies is the largest. Thus part of the mismatch here may be ascribed to a morphological observational bias. At the highest stellar masses ( $\log(M_*/M_\odot) > 11.5$ ), the match to the MASSIVE data is excellent.

In Fig. 3c-d we present the size-mass relation from the HORIZON-AGN simulations. We find a good qualitative match between the shape and slope of the size-mass relation, but the simulated size-mass relation is offset towards larger radii as compared to observations by a factor of 1.88 (median effective radius of the simulations versus the observations is 5.03 kpc versus 2.67 kpc, respectively). The spread in size at fixed stellar masses is significantly smaller as compared to EAGLE, but also slightly smaller as compared to observations. The offset towards larger radii is not likely caused

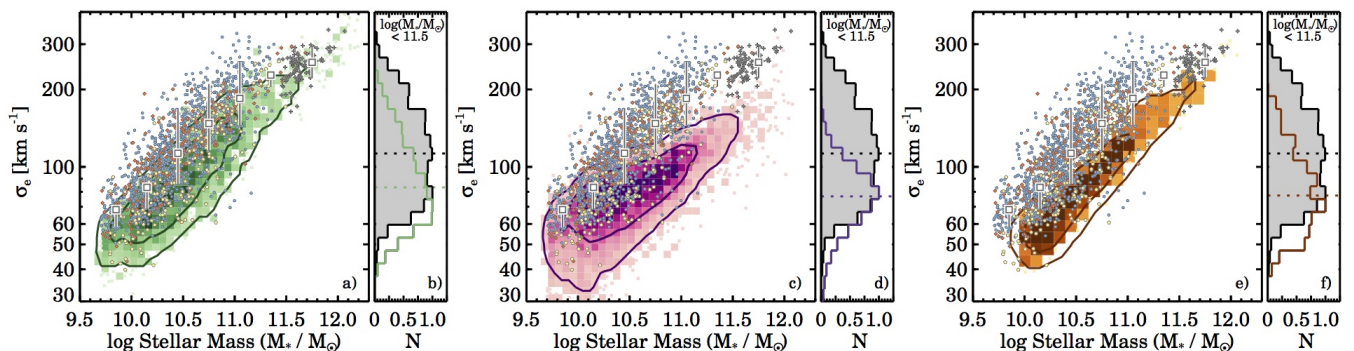
by a morphological bias in the simulations towards disk-type galaxies. The offset is similar at all stellar masses, and in HORIZON-AGN the size-mass relation for early-type lies above the relation for late-type galaxies at all stellar masses (Dubois et al. 2016). The shape of the size-mass relation is surprisingly similar for early-type and late-type galaxies in HORIZON-AGN. We will discuss this offset further in Section 5.

We investigate the size-mass relation from the MAGNETICUM simulations in Fig. 3e-f. Similar to both EAGLE<sup>+</sup> and HORIZON-AGN the shape and slope of the size-mass relation are in good agreement with the observations, but simulated MAGNETICUM galaxies are on average too large below  $\log(M_*/M_\odot) < 11.5$ . The spread of the distribution is similar to the observed sample. We find an excellent agreement between the MAGNETICUM and the MASSIVE sample (i.e., at  $\log(M_*/M_\odot) > 11.5$ ). In Fig. 3f, the median offset between the effective radius of the simulations versus the observations is 4.68 kpc versus 2.67 kpc, respectively.

In conclusion, the size-mass relation is qualitatively well-reproduced by all simulations, but on average simulated galaxies are too large. We checked whether our conclusion would change if we use the full GAMA dataset (see Figure 1), rather than the combined IFS sample, but the results remain the same. The spread in sizes at fixed stellar masses varies between different simulation, from larger to slightly smaller as compared to observations.

#### 4.4 Ellipticity-Mass

In Fig. 4 we compare the relation between the observed ellipticity and stellar mass. The overall shape of the distribution, where the most massive galaxies become increasingly round, is well-recovered by EAGLE<sup>+</sup>, HORIZON-AGN, and MAGNETICUM. For both observations and simulations we find that above  $\log(M_*/M_\odot) \sim 11$  flat galaxies start to disappear. However, above  $\log(M_*/M_\odot) > 11.5$ , galaxies in both EAGLE<sup>+</sup> and HORIZON-AGN appear to be rounder (median  $\epsilon = 0.18, 0.17$ ) as compared to the MASSIVE galaxies (median  $\epsilon = 0.23$ ), whereas MAGNETICUM produces a significant number of extremely massive galaxies with  $\epsilon > 0.25$ ). However, as the ellipticities for MASSIVE galaxies are not de-



**Figure 5.** Comparison of simulated and observational data in the stellar velocity dispersion-mass diagram. Symbols are the same as in Figure 3. While the shape of the EAGLE<sup>+</sup>  $\sigma$ -mass is the same as observation (panel a), the median velocity dispersion is lower: 81.9 km s<sup>-1</sup> versus 112.6 km s<sup>-1</sup> in the observational data (panel b). The HORIZON-AGN relation has a different coefficient and does not extend to high-velocity dispersions (panel c). The median velocity dispersion in HORIZON-AGN is lower than in the observations (panel d, 77.0 km s<sup>-1</sup> versus 112.6 km s<sup>-1</sup> respectively). The shape of the MAGNETICUM  $\sigma$ -mass is the same as observations (panel e), but the relation is offset to lower velocity dispersion (77.2 km s<sup>-1</sup> versus 112.6 km s<sup>-1</sup> in the observational data panel f).

rived within one effective radius but from a global isophote, this offset between EAGLE<sup>+</sup> HORIZON-AGN and MASSIVE, could possibly be attributed to a difference in measuring technique.

On average, the EAGLE<sup>+</sup> ellipticities are slightly lower (rounder) than our observed sample. The offset is most pronounced below  $\log(M_*/M_\odot) < 11$ , where there are almost no galaxies with  $\epsilon > 0.6$ . The median ellipticity is  $\epsilon = 0.24$  for EAGLE<sup>+</sup> versus  $\epsilon = 0.29$  in the observed sample. For HORIZON-AGN the offset is more dramatic, the median ellipticity  $\epsilon = 0.14$  which is a factor 2.0 lower than observed. MAGNETICUM is the only simulation that produces flattened galaxies with ellipticities as high as detected in the observations (median  $\epsilon = 0.34$ ). However, the MAGNETICUM has a notable lower number of extremely round galaxies at all stellar masses.

Lagos et al. (2018b) showed that the missing high-ellipticity galaxies in EAGLE was not due to the resolution of the simulation, as the analysis of the higher resolution EAGLE runs did not improve the sampling of the high ellipticity range. The authors therefore concluded that this is a limitation of the ISM model and cooling adopted in EAGLE. The ISM model includes a temperature floor of 8,000K, i.e., gas is not allowed to cool below this temperature. This temperature limit corresponds to a Jeans length of approximately 1kpc physical, and thus, disks cannot be thinner than this. Our Milky Way and other local disks, however, display thinner disks with typical scale heights of 300-700pc (e.g., Kregel et al. 2002; Bland-Hawthorn & Gerhard 2016). Thus, an important improvement needed in simulations to reproduce the very high ellipticity galaxies, is to realistically model the ISM to produce more realistic vertical structure of disks. Because HORIZON-AGN employs a similar ISM model, we expect the reasoning above to also apply to this simulation. Nonetheless, it should be noted that HORIZON-AGN has a slightly lower spatial resolution than EAGLE<sup>+</sup> and that, due to the method used to compute the hydrodynamics, this spatial limit represents a sharp scale cut where it is no longer possible to derive gradients in the gas properties, at fixed positions. This is also the minimal scale on which feedback energy and momentum from supernovae and AGN

feedback can be released (usually several times that to distinguish anisotropic and isotropic modes of AGN feedback for instance). Such processes therefore limit the formation of thin, highly rotating discs, including at high stellar mass in HORIZON-AGN.

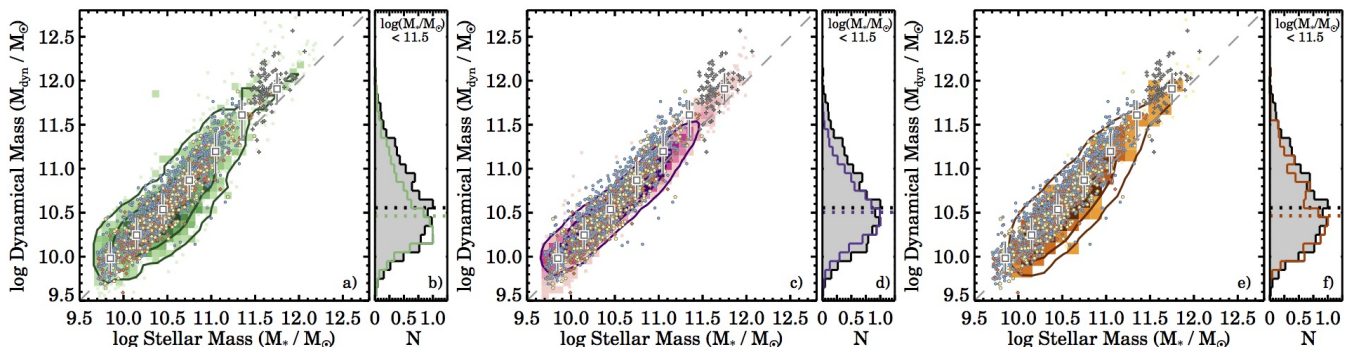
#### 4.5 Velocity Dispersion-Mass

We will now include dynamical information in the comparison. The first step is to look at aperture velocity dispersions that are also accessible using single fibre spectroscopy. However, by using IFS observations we have the advantage that we can precisely define our aperture to match the effective isophote of the galaxies and thus remove any aperture-related uncertainties that might arise from a fixed fibre size. We use a flux-weighted sum of all velocity dispersion measurements within one effective radius:

$$\sigma_e = \frac{\sum_{i=0}^{N_{spix}} F_i \sigma_i}{\sum_{i=0}^{N_{spix}} F_i}. \quad (2)$$

Here, the subscript  $i$  refers to the position of a spaxel within the ellipse with semi-major axis  $R_e$  and ellipticity  $\epsilon_e$ ,  $F_i$  the flux of the  $i^{th}$  spaxel,  $\sigma_i$  the velocity dispersion in km s<sup>-1</sup>. Not all SAMI and ATLAS<sup>3D</sup> measurements have coverage out to one effective radius, and we only include galaxies with a minimum fill factor of 85 per cent.

Note that this approach differs from a single-fibre velocity dispersion measurement. To match a single-fibre measurement, we first need to sum all spectra within an aperture and then measure the stellar velocity dispersion from that spectrum. This single-fibre approach does not take the velocity gradients of a galaxy into account, and is also not easily reproduced from the simulated data. Hence, we use Eq. 2 here. We note that using SAMI Galaxy Survey data we find that velocity dispersion measurements from Eq. 2 are on average 10.8 per cent smaller than the single-fibre approach  $\sigma_e$  values. For the MASSIVE survey data the velocity dispersion maps are not available. Instead, we use the



**Figure 6.** Comparison of simulated and observational data in the dynamical-stellar mass diagram. Symbols are the same as in Figure 3. The combination of size and velocity dispersion largely reconciles the discrepancy between observations and simulations that were seen in Fig. 3 and Fig. 5.

tabulated data from [Veale et al. \(2017b\)](#) that are calculated in a similar way as described here.

We present the velocity dispersion ( $\sigma_e$ ) - stellar mass relation in Fig. 5. The observational data show a clear increasing trend between  $\sigma$  and stellar mass, but there is a noticeable turnover around  $\log(M_*/M_\odot) \sim 11.3$ . This turnover will be further discussed in the next Section 4.6, where we compare dynamical and stellar mass estimates.

The shape of the EAGLE<sup>+</sup> data matches the observed relation well, but the simulated data are lower by 0.15 dex at fixed stellar mass. At the highest stellar masses, there is an excellent agreement between the MASSIVE and EAGLE<sup>+</sup> velocity dispersions. From Fig. 5b we find that the width of the distribution of  $\sigma_e$  values from the EAGLE<sup>+</sup> simulations is similar, but slightly more skewed towards lower  $\sigma_e$  values as compared to observations (0.18 dex versus 0.16 dex in simulations and observations respectively). At fixed stellar mass, between  $10.4 < \log(M_*/M_\odot) < 10.6$  the scatter in EAGLE<sup>+</sup> is slightly smaller than observations (0.09 dex versus 0.12 dex, respectively).

One might naively expect that this offset could be caused by observational biases due to seeing that are known to increase the stellar velocity dispersion as the absorption lines get artificially broadened by the velocity profile of the galaxy. However, as the SAMI and ATLAS<sup>3D</sup> data are similar, this indicates that seeing is not a likely cause for the higher velocity dispersions in the data as compared to observations.

The closest match between HORIZON-AGN and observations is at low stellar masses, but overall the slope of the HORIZON-AGN relation is shallower and the velocity dispersions are too low (Fig. 5c). Below  $\log(M_*/M_\odot) = 10.5$  the offset is approximately 0.12 dex whereas, at high stellar masses, the velocity dispersions are too low by 0.23 dex. From Fig. 5d we also find that the width of the distribution is smaller in HORIZON-AGN as compared to observations (0.09 dex at  $10.4 < \log(M_*/M_\odot) < 10.6$ ). However, this is due to a different slope in the relation from the simulations, and the fact that HORIZON-AGN has few high-dispersion galaxies.

In Fig. 5e we show the results from the MAGNETICUM simulations. The shape of the MAGNETICUM  $\sigma_e$ -mass relation matches the observations well. However, the spread in  $\sigma_e$  at fixed stellar mass ( $10.4 < \log(M_*/M_\odot) < 10.6$ ) is significantly smaller as compared to observations (0.05 dex versus

0.12 dex, respectively). Similar to EAGLE<sup>+</sup> and HORIZON-AGN, the MAGNETICUM velocity dispersions are also offset to lower values as compared to the observed  $\sigma_e$ -mass relation. We note that none of the simulations show the strong change in the slope of the  $\sigma_e$ -mass relation around a stellar mass of  $\log(M_*/M_\odot) \sim 11.3$ .

#### 4.6 Dynamical-Stellar Mass

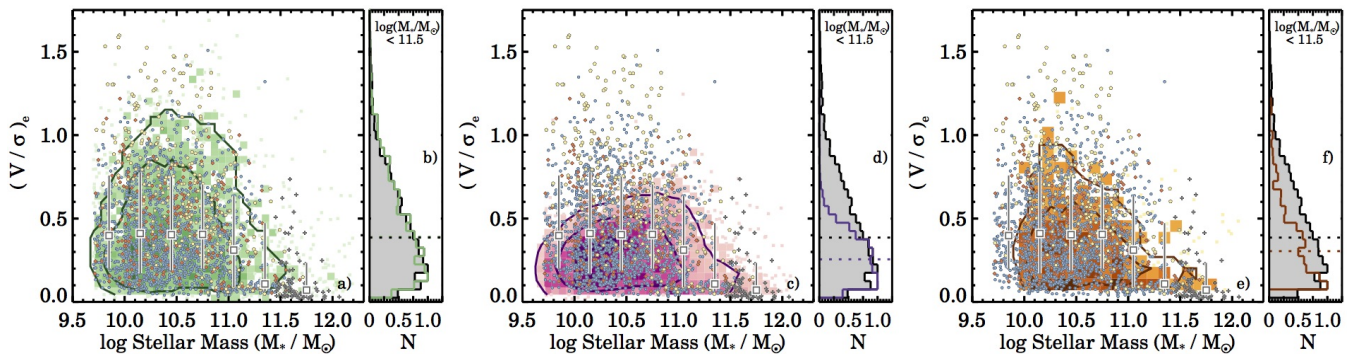
We now calculate dynamical masses from the size and velocity dispersion measurements using the Virial theorem, to reconcile the offsets that we found between the observational and simulated size-mass and  $\sigma_e$ -mass relations. Secondly, in the observational data we also noted that at the highest stellar masses, the size-mass relation becomes steeper, whereas the velocity dispersion-mass relation becomes less steep. As this turnover coincides with the addition of the MASSIVE data, one might worry that there are remaining observational biases between the different samples. To address these issues, we will now compare dynamical and stellar mass estimates.

Dynamical masses are derived from the velocity dispersions and circularised effective radius using the following expression:

$$M_{\text{dyn}} = \frac{\beta \sigma_e^2 R_{\text{e,circ}}}{G}. \quad (3)$$

We use a  $\beta(n) = 5$  that is commonly used for early-type only samples ([Cappellari et al. 2006](#)).  $\beta$  can also be derived from an analytic expression as a function of the Sérsic index, as described by [Cappellari et al. \(2006\)](#), but this is beyond the scope of the paper. Thus, there is a caveat that the dynamical mass estimates will be more uncertain for late-type than early-type galaxies due to the larger rotational component in late-type galaxies.

In Fig. 6 we compare the observational and simulated data in the  $M_{\text{dyn}} - M_*$  plane. First, we note that the large majority of observational data lie above the one-to-one relation, but that their offset from this relation increases with increasing stellar mass. This increase can be explained by increasing non-homology, difference in rotational support, dark matter fraction, and possibly a varying IMF (e.g., [van Dokkum & Conroy 2010](#); [Cappellari et al. 2012](#)). While the exact nature of this offset is beyond the scope of this paper, we do note



**Figure 7.** Comparison of simulated and observational data in the stellar mass- $V/\sigma$  diagram. Symbols are the same as in Figure 3. There is a remarkably good match between EAGLE<sup>+</sup> and the simulations for  $(V/\sigma)_e$  both as a function of stellar mass and overall distribution. While the trend with stellar mass is correct for HORIZON-AGN, the  $(V/\sigma)_e$  values are too low at all stellar mass except  $\log(M_*/M_\odot) > 11.5$ . MAGNETICUM galaxies show a stronger decline with increasing stellar mass as compared to observations, and the overall  $(V/\sigma)_e$  values are lower except at the highest stellar masses.

that there is no sudden change in the  $M_{\text{dyn}} - M_\star$  relation around a stellar mass of  $\log(M_\star/M_\odot) \sim 11.3$ . By comparing dynamical and stellar masses we can now conclude that there are no measurement biases in  $R_e$  and  $\sigma_e$  from the difference observational samples. Instead, above a stellar mass of  $\log(M_\star/M_\odot) \sim 11.3$ , massive-round slow rotating galaxies formed by dry-major mergers start to dominate the galaxy population (for a recent review on this topic see Cappellari 2016, but also Robotham et al. 2014).

The combination of size and velocity dispersion also largely reconciles the discrepancy between observations and simulations. This is important to note, as we will continue to compare both the rotational and random motions of stars in the next section. EAGLE<sup>+</sup> shows the largest spread in the  $M_{\text{dyn}} - M_\star$  relation (Fig. 6a), which is unsurprising as this simulation also showed the largest spread in the size-mass and  $\sigma_e$ -mass relation. The distribution of dynamical masses in Fig. 6b,d,f is lower by approximately  $-0.1$  dex in the simulations as compared to observations. While the exact cause for this offset is unclear, as mentioned above, there are many factors that could contribute to this relatively minor offset.

#### 4.7 $V/\sigma$ -Mass

In the following section we will look at the  $(V/\sigma)_e$  measurements.  $V/\sigma$  describes the average ratio of the velocity and velocity dispersion within one effective radius. We adopt the following definition by Cappellari et al. (2007):

$$\left(\frac{V}{\sigma}\right)^2 \equiv \frac{\langle V^2 \rangle}{\langle \sigma^2 \rangle} = \frac{\sum_{i=0}^{N_{\text{spix}}} F_i V_i^2}{\sum_{i=0}^{N_{\text{spix}}} F_i \sigma_i^2}. \quad (4)$$

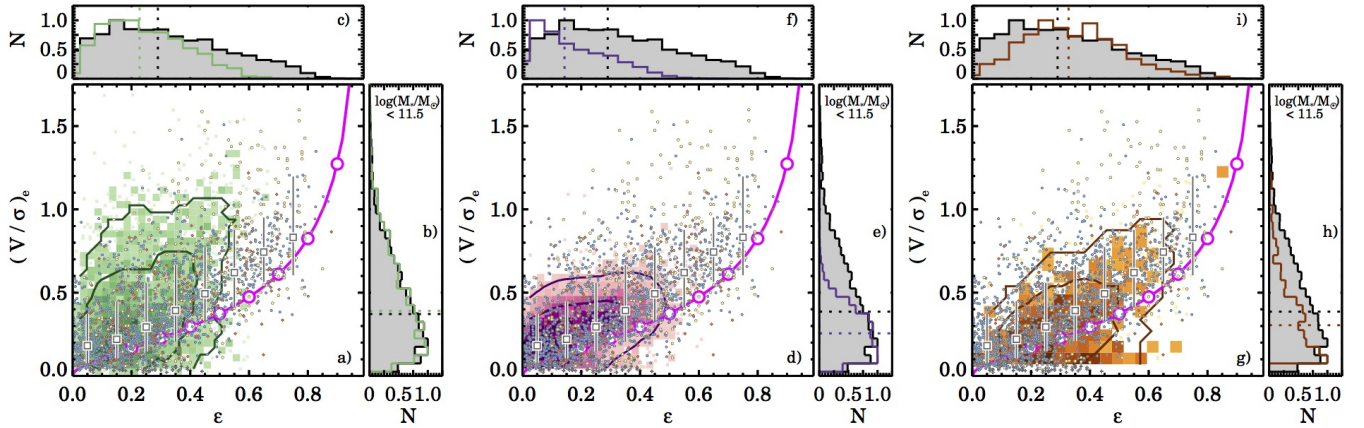
Here, the subscript  $i$  refers to the position of a spaxel within the ellipse,  $F_i$  the flux of the  $i^{\text{th}}$  spaxel,  $V_i$  is the stellar velocity in  $\text{km s}^{-1}$ ,  $\sigma_i$  the velocity dispersion in  $\text{km s}^{-1}$ . For the SAMI Galaxy Survey data, we use the unbinned flux, velocity, and velocity dispersion maps as described in Section 2.1.2. We take the sum over all spaxels  $N_{\text{spix}}$  that meet the quality cut  $Q_1$  and  $Q_2$  (Section 2.1.2) within an ellipse with semi-major axis  $R_e$  and axis ratio  $b/a$ . For the ATLAS<sup>3D</sup> data, the unbinned flux maps are combined with

the Voronoi binned stellar kinematic data as described in Section 2.2. The approach for CALIFA data is similar to ATLAS<sup>3D</sup> with the exception of the flux maps that are also Voronoi binned. The kinematic maps for the MASSIVE survey are not publicly available. Instead we convert the spin-proxy measurements  $\lambda_{R_e}$  to  $V/\sigma$  values using Eq. 3 from van de Sande et al. (2017a) assuming  $\kappa = 1.06$ . Note that these measurements are obtained within circularised apertures, but as most MASSIVE galaxies are round with little rotation, the impact of this choice is relatively small. For the simulated EAGLE<sup>+</sup>, HORIZON-AGN, and MAGNETICUM data, the approach for measuring  $(V/\sigma)_e$  is identical to that of the SAMI Galaxy Survey data.

Not all SAMI and ATLAS<sup>3D</sup> measurements have coverage out to one effective radius. For all galaxies where the fill factor of good spaxels within one effective radius is less than 95 percent, we apply an aperture correction as outlined in van de Sande et al. (2017a). A total of 266 SAMI galaxies are aperture corrected (17 percent), 136 for ATLAS<sup>3D</sup> (56 percent), 21 for CALIFA (8 percent), but none for MASSIVE.

We present all  $(V/\sigma)_e$  measurements in Fig. 7. First, we comment on the variation in  $(V/\sigma)_e$  between the different observational surveys. While the ATLAS<sup>3D</sup> and SAMI distributions show an excellent agreement, with slightly more high  $(V/\sigma)_e$  late-type galaxies in SAMI, the clear outlier towards high  $(V/\sigma)_e$  is CALIFA. As mentioned in Section 2.3, the isophotal diameter selection of CALIFA creates a bias towards edge-on galaxies, an effect which is also clearly visible in  $(V/\sigma)_e$ . Due to its high stellar mass, the MASSIVE sample contains the largest amount of slow-rotating galaxies with low  $(V/\sigma)_e$ .

There is an excellent match between the observed  $(V/\sigma)_e$  measurements and the EAGLE<sup>+</sup> simulations (Fig. 7a-b), whereas HORIZON-AGN galaxies on average have too low  $(V/\sigma)_e$  values (Fig. 7c-d). In particular there are almost no galaxies that have  $(V/\sigma)_e > 0.7$ . Note that pure disks typically have  $(V/\sigma)_e > 0.8$ , and thus appear not to be produced in the HORIZON-AGN simulations. The median  $(V/\sigma)_e$  values of MAGNETICUM galaxies are lower as compared to observations, but in between the range covered by EAGLE<sup>+</sup> and HORIZON-AGN.



**Figure 8.** Comparison of simulated and observational data in the  $V/\sigma$ - $\epsilon$  diagram. Symbols are the same as in Figure 3. The magenta line shows the theoretical prediction for an edge-on view of axisymmetric galaxies. On the line, from the bottom; up, the open circles correspond to the locations of galaxies with different intrinsic ellipticities  $\epsilon_{\text{intr}}=0.2-0.9$  (see Cappellari et al. 2007; Emsellem et al. 2011). Galaxies that are observed under different viewing angles follow a downward line from edge-on (magenta line) to face-on (towards zero ellipticity). As compared to observations, the shape of the distribution of EAGLE<sup>+</sup> data is steeper, that is few simulated galaxies with  $\epsilon_e > 0.6$  and  $(V/\sigma)_e > 0.5$  exist. For HORIZON-AGN the shape of the distribution appears to be correct, but both  $\epsilon_e$  and  $(V/\sigma)_e$  are significantly lower as compared to observations. The MAGNETICUM data closely matches the shape of the observed distribution but at fixed  $\epsilon_e$  galaxies with high  $(V/\sigma)_e$  are missing.

All simulations reproduce the observed trend with stellar mass, albeit with small differences. EAGLE<sup>+</sup> and HORIZON-AGN show a moderate increase in  $(V/\sigma)_e$  from  $\log(M_*/M_\odot) = 9.5$  to  $\log(M_*/M_\odot) = 10.75$ , with a strong decline in  $(V/\sigma)_e$  at  $\log(M_*/M_\odot) > 10.75$ , whereas MAGNETICUM galaxies have lower  $(V/\sigma)_e$  with increasing stellar mass across the entire mass range. At the highest stellar masses ( $\log(M_*/M_\odot) > 11.5$ ), we find that the majority of MASSIVE galaxies are slow rotators with  $(V/\sigma)_e < 0.2$  (median  $(V/\sigma)_e=0.07$ ). In comparison, EAGLE<sup>+</sup> galaxies at the same stellar mass have a considerably higher spin with a median  $(V/\sigma)_e=0.20$ , and there are several galaxies with  $(V/\sigma)_e > 0.5$ . This was also noticed by Lagos et al. (2018b) who showed that the higher spins in the most massive galaxies are due to AGN feedback being insufficient to quench star formation in the central galaxies of halos with masses  $\gtrsim 10^{14} M_\odot$  to the levels observed in brightest cluster galaxies. These galaxies end up having two times too much stellar mass and SFRs compared to observations. HORIZON-AGN galaxies at this mass range show more consistent values of  $(V/\sigma)_e$  as compared to the observations (median  $(V/\sigma)_e=0.11$ ), whereas MAGNETICUM has higher values and is more similar to EAGLE<sup>+</sup> (median  $(V/\sigma)_e=0.21$ ) above  $\log(M_*/M_\odot) > 11.5$ .

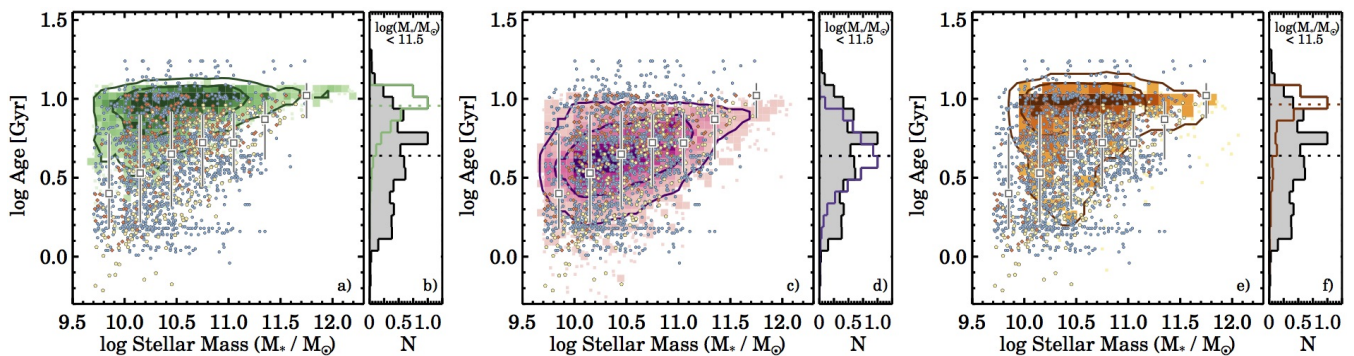
#### 4.8 $V/\sigma$ -Ellipticity

In Fig. 8 we combine the ellipticity and  $(V/\sigma)_e$  measurements, which allows us to study the dynamical properties of galaxies in relation to their intrinsic shape and inclination (see e.g., Binney 2005; Cappellari et al. 2007). It is common practise to show tracks of different inclination and intrinsic ellipticity for oblate axisymmetric rotators, however, to avoid over-crowding of data within the figures we only show model tracks for galaxies observed edge-on with varying intrinsic ellipticity (magenta-line).

We previously found that the EAGLE<sup>+</sup> galaxies are on average rounder, but with a similar distribution in  $(V/\sigma)_e$ . From Fig. 8a-c we conclude that the  $(V/\sigma)_e$ - $\epsilon_e$  relation is steeper as compared to the observations; there are almost no simulated galaxies with  $\epsilon_e > 0.6$  and  $(V/\sigma)_e > 0.5$ . The EAGLE<sup>+</sup> simulations recover the right ratio of velocity versus velocity dispersion, yet at fixed  $V/\sigma$  galaxies are too round. Before, we also showed that the aperture velocity dispersions from the simulations were slightly too low (Fig. 5). However, as the effective radii were also too large, the fact that the  $V/\sigma$  values are correct here, does not directly imply that the average rotational velocities are too high. Whereas the shape of the HORIZON-AGN relation resembles the observational trend, we find that galaxies are both too round with too low  $V/\sigma$  (Fig. 8d-f). Interestingly, we find a slow rotating horizontal band of galaxies with  $(V/\sigma)_e < 0.15$  where the ellipticity extent almost out to  $\epsilon_e = 0.4$  that is very similar to the data from the MASSIVE Survey but to a lesser extent in the other surveys. This region of the  $V/\sigma - \epsilon$  diagram is typically occupied with galaxies that are triaxial in shape. In Fig. 8g-i we find that galaxies in the MAGNETICUM closely match the observational data and the predictions for oblate axisymmetric rotators (magenta line), while also recovering the region assigned to slow rotators. However, at fixed ellipticity, galaxies with high  $(V/\sigma)_e$  values are absent, causing the overall distribution of  $(V/\sigma)_e$  to be lower than observations.

#### 4.9 Age-Mass

For the second-last set of comparisons we switch to looking at stellar population properties, but limit ourselves to stellar population age. This is partly motivated by the recent SAMI Galaxy Survey result that intrinsic ellipticity and stellar age are correlated, independent of stellar mass or morphology (van de Sande et al. 2018). These next two sections



**Figure 9.** Comparison of simulated and observational data in the age-mass diagram. Symbols are the same as in Figure 3. For all three simulations we find a weak trend with stellar mass, where lower mass galaxies are younger. However, the median age for EAGLE<sup>+</sup> and MAGNETICUM is significantly higher as compared to observations. We find a large observed spread in age, whereas both EAGLE<sup>+</sup> and MAGNETICUM have a small spread in stellar population age. While HORIZON-AGN also shows a smaller spread in age as compared to observations, it is considerably more consistent as compared to the other two simulations.

are therefore devoted to investigating whether this same relation exists within the simulations.

In Fig. 9 we first present stellar population age as a function of stellar mass. The SAMI and CALIFA sample have a large spread in age  $-0.2 < \log \text{Age}[\text{Gyr}] < 1.2$ , mostly due to the variety in morphologies from late-spirals to ellipticals. ATLAS<sup>3D</sup> galaxies on average have older stellar populations, which is not surprising given that the sample consists of early-type galaxies only.

In all three simulations there is a mild trend with stellar mass; on average galaxies are younger at lower stellar mass. In the EAGLE<sup>+</sup> simulations we find no young galaxies with ages below  $\log \text{Age} < 0.4$  ( $\sim 2.5$  Gyr), and the spread of the age distribution is also significantly smaller. This could possibly be a consequence of EAGLE having a consistently higher fraction of passive low-mass satellite galaxies as compared to observations (Y. Bahé et al., in prep). The median observed log stellar age is 4.36 Gyr, whereas we find median age of 8.96 Gyr in EAGLE<sup>+</sup>. The median age of HORIZON-AGN is nearly identical to the observations, but similar to EAGLE<sup>+</sup> we find a smaller spread in the age distribution. The age distribution of MAGNETICUM is very similar to EAGLE<sup>+</sup> albeit with a few more young galaxies. However, the mismatch of MAGNETICUM ages with the observations is large (median age 9.25 Gyr).

There is a caveat that the observed luminosity-weighted SSP-equivalent ages are likely to be biased to young ages for all galaxies that have experienced recent star formation. While both observed and simulated ages are luminosity weighted, we expect the spectral modelling and choice of stellar population model to have a relatively large impact (e.g., Sánchez et al. 2016a). We note that the observed stellar ages scatter well above the age of the Universe, which we regard as a caution to not over-interpret these results. In addition, no extinction effects were included in the  $r$ -band luminosity weighting in the simulations.

#### 4.10 Intrinsic Ellipticity - Age

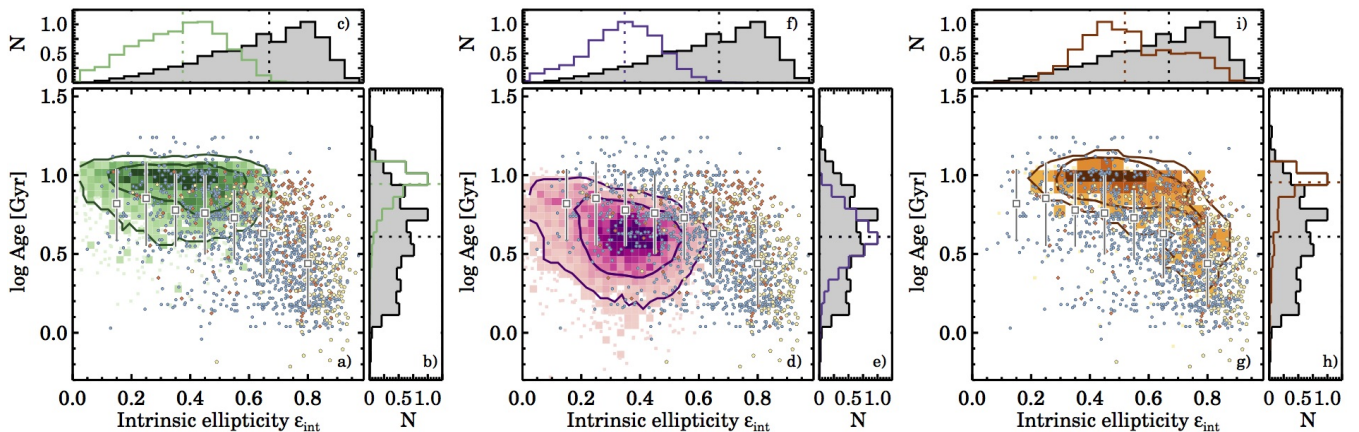
For the final comparison, we explore the relation between the intrinsic ellipticity or edge-on ellipticity of galaxies and their mean stellar population age. In van de Sande et al.

(2018), we showed that the characteristic stellar age follows the intrinsic ellipticity of galaxies remarkably well. As the shape and kinematic properties of the galaxies change from dynamically cold, and intrinsically flat into dynamically-hotter, pressure-supported, thicker, oblate spheroids, we find that the mean age of the stellar population increases. This trend is still observed when galaxies are separated in bins of low and high-stellar mass, or early-type and late-type.

We repeat the analysis of van de Sande et al. (2018) with the full sample, with the exception of MASSIVE galaxies for which we do not have stellar age measurements. For details on the technique to derive intrinsic ellipticities we refer to their work, but we highlight the main assumptions here. First, we assume that all galaxies are axisymmetric, and mildly anisotropic:  $\beta_z = 0.6 \pm 0.1 \times \epsilon_{\text{intr}}$  (Cappellari et al. 2007) where  $\beta_z$  is the anisotropy and  $\epsilon_{\text{intr}}$  is the intrinsic ellipticity. Secondly, any galaxy that falls below the magenta line in Fig. 8 is excluded from the sample because they are outside the model range. These are typically the galaxies with the oldest stellar populations. Third, near-perfectly round galaxies ( $\epsilon_e < 0.025$ ) are also excluded, because the model predictions are highly degenerate and the relative measurement uncertainties on  $\epsilon_e$  are large.

In the simulations we do not have to rely on any of the de-projection assumptions for deriving intrinsic ellipticities. Instead, we can use the edge-on projection of all galaxies and measure the intrinsic ellipticity from those. Here, the edge-on projection angle is computed using the total stellar angular momentum of all stellar particles in the galaxy. Similar to the observed sample, we exclude all EAGLE<sup>+</sup>, HORIZON-AGN, and MAGNETICUM galaxies below the magenta line in Fig. 8 so that both simulated and observed samples are selected in the same way.

We note that we could have derived  $\epsilon_{\text{intr}}$  using an identical method that is used for the observations. However, we already concluded in Section 4.8 that the simulated data occupy different regions of the  $V/\sigma - \epsilon$  diagram as compared to observations. The observed intrinsic ellipticity distribution derived from  $(V/\sigma)_e$  and  $\epsilon_e$  is similar to intrinsic ellipticity distributions derived from inverting the distributions of apparent ellipticities and kinematic misalignments (Weijmans et al. 2014; Foster et al. 2017). However, as the simulated



**Figure 10.** Comparison of simulated and observational data in the age-intrinsic ellipticity diagram. Symbols are the same as in Figure 3. The intrinsic ellipticities from the simulations are derived from the edge-on projection of galaxies, whereas the observational intrinsic ellipticities have been derived using theoretical model predictions of axisymmetric oblate rotators. The trend where younger galaxies are also more flattened is poorly recovered in EAGLE<sup>+</sup>, slightly more pronounced in HORIZON-AGN, but clearly detected in MAGNETICUM. However, at fixed intrinsic ellipticity, all simulations show offsets in age. In both EAGLE<sup>+</sup> and HORIZON-AGN there is a deficit of flattened galaxies with  $\epsilon_{\text{intr}} > 0.7$ , while MAGNETICUM galaxies extend to  $\epsilon_{\text{intr}} \sim 0.7$ .

data does not overlap with the observations presented here in either  $V/\sigma$  or  $\epsilon$ , using the same approach will not result in accurate intrinsic ellipticity. Therefore, we use the edge-on projection in the simulations to estimate intrinsic ellipticities.

In Fig. 10 we show the relation between age and intrinsic ellipticity from observations and simulations. There is a good agreement between the SAMI, ATLAS<sup>3D</sup> and the CALIFA trends, but as expected the CALIFA sample contains more galaxies with intrinsic ellipticity that are highly flattened. We find that the shape of the EAGLE<sup>+</sup> age-intrinsic ellipticity distribution is mildly consistent with the observations, although the EAGLE<sup>+</sup> distribution is offset to older ages. The distribution from the EAGLE<sup>+</sup> simulations is also not as extended as the observed relation (median  $\epsilon_{\text{intr}}$  of 0.37 and 0.67 for EAGLE<sup>+</sup> and observation, respectively), and the scatter in age is also considerably smaller. The regime where the observational ages significantly start to decrease at  $\epsilon_{\text{intr}} > 0.6$  is not covered by EAGLE<sup>+</sup> thus we cannot conclude whether there is a trend between age and intrinsic ellipticity. HORIZON-AGN galaxies show a qualitatively consistent trend with the observations where younger galaxies are more flattened (lower  $\epsilon_{\text{intr}}$ ). Yet, while HORIZON-AGN has relatively young stellar ages, the galaxies are surprisingly round (median  $\epsilon_{\text{intr}}$  of 0.35). We do find a clear trend between age and  $\epsilon_{\text{intr}}$  in the MAGNETICUM simulations, albeit with an offset towards older ages. MAGNETICUM galaxies extend significantly beyond both EAGLE<sup>+</sup> and HORIZON-AGN in terms of intrinsic ellipticity. Flattened galaxies with high intrinsic ellipticity ( $\epsilon_{\text{intr}} > 0.7$ ) that are abundant in the observed samples are only seen in MAGNETICUM, but the overall  $\epsilon_{\text{intr}}$  distribution in MAGNETICUM is offset to lower values (median  $\epsilon_{\text{intr}}$  of 0.53).

Naively, some of the discrepancies can be understood by the simple lack of high ellipticity galaxies or from the lack of galaxies with young stellar ages. In the first scenario one would expect that future simulations in which high ellipticity galaxies exist, would naturally extend the relations

seen here for EAGLE<sup>+</sup> and HORIZON-AGN to the high ellipticity population. However, this may not be so simple, as the changes needed to reproduce the vertical structure of disks, and therefore the high-ellipticity galaxies, are major (i.e., higher resolution, more realistic ISM models, and a corresponding stellar feedback that acts on smaller scales). Thus, it is not guaranteed that the properties of the rounder, massive ellipticals and lenticulars - that the current generation of simulations reproduce well - will still be reproduced by the next generation of simulations.

## 5 PREVIOUS COMPARISON STUDIES

We have now compared seven different structural, dynamical, and stellar population parameters from observations and mass-matched simulations at  $z < 0.1$ . Before summarising our findings and discussing their implications, here we first briefly review the literature from the last decade and then review previous work that used the same simulations in comparison to observations.

### 5.1 Results from Simulations in the Last Decade

The last decade has seen significant progress in the area of cosmological hydrodynamical simulations, with the first large cosmological volumes, with high enough resolution to study the internal structure of galaxies being possible (e.g., Devriendt et al. 2010; Dubois et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Pillepich et al. 2018). These simulations have been designed to overcome the problem of catastrophic loss of angular momentum and over-cooling problem that was observed in simulations prior to 2010 (Steinmetz & Navarro 1999; Navarro & Steinmetz 2000). The catastrophic loss of angular momentum problem refers generally to the issue of simulated galaxies being too compact and having too low specific angular momentum, which affected Smoothed



Particle Hydrodynamics (SPH) codes mostly (e.g., GADGET, Springel et al. 2001). This problem that was mainly solved by implementing angular momentum conserving SPH schemes. On the other hand, Adaptive Mesh Refinement (AMR) codes (e.g., RAMSES, Teyssier 2002) did not suffer from the same angular momentum problem, but only from the over-cooling problem (Lackner et al. 2012; Dubois et al. 2013). In AMR, massive galaxies tended to be too large and have prominent disks, in contradiction with the observations of massive galaxies. Thus, in general both SPH and AMR codes generated galaxies were too massive compared to observations due to the efficient cooling and star formation, and in addition too compact in the case of SPH. The angular momentum catastrophe and over-cooling problem were solved by a combination of improved spatial resolution and numerical techniques of the simulations and the inclusion of efficient feedback (e.g., Kaufmann et al. 2007; Zavala et al. 2008; Governato et al. 2010; Guedes et al. 2011; DeFelippis et al. 2017).

Feedback was necessary to prevent large amounts of gas from cooling onto the galaxies and therefore fuelling star formation, to avoid the build-up of stellar mass in the centres of galaxies, and for removing low angular momentum gas from galaxies. For example, Remus et al. (2017) showed that the properties of early-type galaxies can only be reproduced with AGN. Without AGN, early-type galaxies are too compact with dark matter fractions that are too low. Because of the catastrophic angular momentum problem, the comparison of kinematic measurements of galaxies and simulations was mostly done in a qualitative manner, or in idealised simulations tackling specific physical processes (such as galaxy mergers; e.g., Cox et al. 2006; Taranu et al. 2013; Naab et al. 2014).

There has been a considerable amount of work over the last years comparing the specific angular momentum of galaxies with observations of late- and early-type galaxies in the local Universe (e.g., Pedrosa & Tissera 2015; Teklu et al. 2015; Genel et al. 2015; Zavala et al. 2016; Lagos et al. 2017; Wang et al. 2018), whereas other studies compare  $\lambda_R$ , the proxy for the stellar spin parameter, from observations in early-type galaxies (Emsellem et al. 2007, 2011) with simulations (e.g., Naab et al. 2014; Penoyre et al. 2017; Choi & Yi 2017; Lagos et al. 2018b; Schulze et al. 2018). Some comparisons between observations and simulations have also been presented for galaxies at high redshift (Swinbank et al. 2017). All this work has shown that simulations are now getting close to reproducing observations, with disks typically having high spin and high specific angular momentum, whereas bulges have  $\approx 5 - 8$  times less specific angular momentum at fixed stellar mass.

## 5.2 Comparing Observations and Simulations with EAGLE, HORIZON-AGN, and MAGNETICUM

We now review previous work that used EAGLE, HORIZON-AGN, or MAGNETICUM, for a comparison to observations. The size-mass relation for EAGLE is presented in Schaye et al. (2015), where the simulations are compared to SDSS results from Shen et al. (2003) for disk-like galaxies, i.e., where Sérsic  $n < 2.5$ . The agreement they find is excellent, albeit by construction, as the models that produced galaxies that were

too small were rejected (Crain et al. 2015). This calibration was derived by using the size-mass relation of star-forming galaxies. The difference between their models and observations is  $\lesssim 0.1$  dex over the full stellar mass range. The size-mass relation in EAGLE is also studied in Furlong et al. (2017), at different redshifts and separating star-forming and passive galaxies. They too find a good agreement to the Shen et al. (2003) results. We note that Furlong et al. (2017) apply a constant factor of 1.4 to convert the Shen et al. (2003) circularised radii to semi-major axis values of both star forming and passive galaxies. They find systematically larger radii ( $\sim 0.2$  dex) as compared to Schaye et al. (2015), but this mismatch in radii is attributed to different measuring techniques. Nonetheless, Furlong et al. (2017) conclude that the EAGLE size-mass relation shows an excellent agreement with observations.

Instead of using the Shen et al. (2003) size-mass relations, Lange et al. (2016) use reprocessed SDSS imaging as part of the GAMA Survey as their observational reference. They find an excellent agreement of spheroidal galaxies down to  $\log(M_\star/M_\odot) \sim 10$ , but below this mass EAGLE galaxies are too large. Disk galaxy sizes are found to be larger in the GAMA data. This is partly because disk-only components are derived for GAMA data (from bulge/disk decomposition), which are then compared with global sizes from EAGLE.

In general, the main conclusions from these papers are slightly different to the results presented here; we find that the size-mass from EAGLE is offset to larger sizes as compared to the observational data. An important caveat here is that previous work used observational data with optical luminosity effective radii, while for EAGLE stellar mass half-mass radii were used. This can lead to a systematic offsets when one compares half-light versus half-mass radii, as half-light radii tend to be larger than the stellar mass-based ones (see Appendix A). Thus, a significant improvement for future simulations is to make more realistic comparisons with the measured sizes of galaxies by taking into account the way they have been measured.

The HORIZON-AGN size-mass relation is compared to observational results in Dubois et al. (2016). Their conclusion is that the sizes at  $z = 0.25$  are in good agreement with observational results from van der Wel et al. (2014), who use deep *Hubble Space Telescope* data to investigate the size-mass evolution of galaxies. For early-type galaxies, the HORIZON-AGN size-mass relation is higher in size by  $\sim 0.2$  dex as compared to late-type galaxies at a stellar mass of  $\log(M_\star/M_\odot) < 11$ . This is inconsistent with the observational result from van der Wel et al. (2014), and with the results presented here. While we do not split the sample into late- and early-type, we find that HORIZON-AGN galaxies are on average too large at all stellar masses as compared at  $z < 0.1$ , which appears to be in contrast with the results from Dubois et al. (2016). The most likely reason for late-type and early-type to have a lower size-mass relation in Dubois et al. (2016) is because the effective radii are mass-weighted quantities and not  $r$ -band luminosity weighted. As luminosity weighting typically increases the sizes of late-type galaxies more than early-type galaxies, this could explain the apparent inconsistency.

Remus et al. (2017) compare the MAGNETICUM size-mass relation for early-type galaxies to results from SDSS (Shen et al. 2003) and GAMA (Baldry et al. 2012) and find

that AGN feedback is essential to establish the observed size-mass relations. [Schulze et al. \(2018\)](#) also investigate the MAGNETICUM size-mass relation, but now for both early-type and late-type galaxies to observations from ATLAS<sup>3D</sup>, CALIFA, SLUGGS ([Forbes et al. 2017](#)), and GAMA ([Baldry et al. 2012](#); [Lange et al. 2015](#)). Note that both [Remus et al. \(2017\)](#) and [Schulze et al. \(2018\)](#) use mass-weighted radii from MAGNETICUM in the comparison but luminosity-weighted sizes from observations. [Schulze et al. \(2018\)](#) conclude that the simulated MAGNETICUM sizes are consistent with GAMA observations, but that they are larger than the sizes from IFS observations. This is contradictory to our results. When using luminosity-weighted radii and a mass-matched sample, we find that at fixed stellar mass, MAGNETICUM sizes are overestimated. Similar as for the previous EAGLE comparisons, this can be attributed to the fact that luminosity-weighted sizes on average are larger than mass-weighted sizes (see Appendix A).

The dynamical properties of simulated galaxies in EAGLE, HORIZON-AGN, and MAGNETICUM have been studied in several papers. [Chisari et al. \(2015\)](#) use HORIZON-AGN and present stellar rotational properties of galaxies as a function of stellar mass. As their definition of  $V/\sigma$  is different from the commonly used definition used in observations, their results were not compared to observations. Instead of measuring velocity and velocity dispersion from an LOSVD,  $V$  and  $\sigma$  are derived from decomposing the spin properties, i.e., the total angular momentum, of individual star particles. Thus  $V/\sigma$  is an average derived property using all individual star particles. The trend in  $V/\sigma$  with stellar mass is similar as found here, but due to the different definition of  $V/\sigma$  we can only make a qualitative comparison.

Using the same data, [Dubois et al. \(2016\)](#) investigate the impact of AGN on the  $V/\sigma$  measurements and find that without AGN feedback their simulation produces too many fast-rotating galaxies above a stellar mass of  $\log(M_\star/M_\odot) > 10.5$ . [Choi et al. \(2018\)](#) also use the HORIZON-AGN simulations with an approach similar to ours, i.e., they extract galaxy parameters from the simulations with measuring techniques that closely match observations. They use the spin parameter proxy  $\lambda_R$ , which is nearly identical to  $V/\sigma$ , and investigate the evolution of early-type galaxies. Similar to our results, they find that the majority of galaxies in HORIZON-AGN have low spin ( $\lambda_{R_e} \lesssim 0.3$ ) and are rounder ( $\epsilon < 0.2$ ) as compared to observed samples. Nonetheless, they successfully recover the observed trend that the most massive galaxies have the lowest spin.

MAGNETICUM early-type galaxies are compared to SAMI, ATLAS<sup>3D</sup>, CALIFA, and SLUGGS in the  $\lambda_{R_e} - \epsilon$  plane in [Schulze et al. \(2018\)](#). Their analysis also includes a cumulative distribution comparison. They conclude that 70 percent of their galaxies are classified as fast-rotators and 30 percent as slow-rotators using the [Emsellem et al. \(2011\)](#) criterion. While the number of fast-rotators is lower as compared to ATLAS<sup>3D</sup> (86 percent, [Emsellem et al. 2011](#)), these numbers are qualitatively similar to what we find here. When comparing  $(V/\sigma)_e$  from MAGNETICUM to the full observed sample, without a cut in morphology, we find that the MAGNETICUM galaxies have moderately lower  $(V/\sigma)_e$  values.

The dynamical properties of galaxies in the EAGLE simulations are studied by [Lagos et al. \(2017, 2018a,b\)](#). They find an excellent match between the specific angular mo-

mentum of galaxies ( $j_\star$ ) in EAGLE, ATLAS<sup>3D</sup>, and SAMI as presented in [Cortese et al. \(2016\)](#). Using both EAGLE and HYDRANGEA, [Lagos et al. \(2018b\)](#) focus on the impact of galaxy mass and environment on the spin-down of galaxies using the spin parameter proxy. When comparing observations and simulations in the  $\lambda_{R_e} - \epsilon_e$  plane, they find similar results as presented here. Because the focus in [Lagos et al. \(2018b\)](#) was placed on massive, slowly rotating galaxies, our comparison here extends to significantly higher spin parameter values and high ellipticity galaxies.

The comparisons presented here of ellipticity, velocity dispersions, and age, have not been analysed in this much detail before. Thus we have the ability to unveil new and important shortcomings of the simulations. Ellipticity has been explored in combination with  $V/\sigma$  and  $\lambda_R$  (e.g., [Naab et al. 2014](#); [Penoyre et al. 2017](#); [Choi et al. 2018](#); [Lagos et al. 2018b](#); [Schulze et al. 2018](#)), but here we have taken an additional step of comparing ellipticity distributions using mass-matched samples. The offset towards thicker disks in EAGLE was also noticed by [Trayford et al. \(2017\)](#) who found that the optical extinction is much smaller than observed for edge-on galaxies, but consistent for face-on objects. Their interpretation was that EAGLE disks are too thick. Stellar ages are a key aspect where we can advance our understanding of galaxy formation by informing the upcoming simulations of areas in which more development is needed. There have been several papers looking at stellar population properties in simulations, in particular colour (e.g., [Trayford et al. 2015](#); [Correa et al. 2017](#)). These results show that the colour of simulated galaxies match the observed loci of the blue-cloud and red-sequence, though not the width of the colour distributions. This agreement may be inconsistent with our findings that the average age of low-redshift EAGLE galaxies is too high. However, colours and ages, though correlated, are not equivalent to each other and therefore the discrepancies here may come from additional effects, such as metallicity, dust, or systematic differences in the way ages and environments are measured in the simulation versus observations.

Our work pushes the boundaries of the comparisons done to date in the sense that we explore a very wide range of galaxy properties. Simultaneously, we take special care in the way we are calculating all galaxy properties to mimic the observations. The detail in which we compare to simulations, both qualitative and quantitative, and using both the median and the spread of the distributions, is also unprecedented. This has been largely feasible because simulations are now able to predict galaxy properties that match better with observations than ever before.

## 6 SUMMARY AND DISCUSSION

This study presents a detailed comparison of multiple galaxy properties from simulations and IFS observations using as close to identical measurement techniques as possible. We compare results from the EAGLE, HYDRANGEA, HORIZON-AGN, and MAGNETICUM simulation projects to observational results from the SAMI Galaxy Survey, CALIFA, ATLAS<sup>3D</sup> and MASSIVE Surveys. The main goal of this work is to take advantage of the recent wealth of 2D IFS data and compare these to cosmological simulations, with the aim of identifying key areas of success and tension. By investigat-

ing whether simulations simultaneously reproduce the structural, resolved dynamical and stellar population properties of galaxies, any area of tension will inform us where the simulations require revision. Furthermore, the advantage of using more than one cosmological hydrodynamical simulation allows us to compare the results from different simulations with each other.

Our measurements from the simulations closely match the observational techniques. In particular for the dynamical measurements, this involves constructing line-of-sight velocity distributions, from which the velocity and velocity dispersion were extracted. We combine four different IFS surveys to create a more homogeneous observational sample where individual survey biases are lessened. A mass-matching technique is used to remove the difference between the observed and simulated stellar mass function. With this method, we sample the same fraction of galaxies from both observations and simulations as a function of stellar mass.

For the EAGLE and HYDRANGEA simulations, here combined and referred to as EAGLE<sup>+</sup>, we find a good match with the observations in the size-mass plane, although there is an offset to larger effective radii. The trend of observed ellipticity with stellar mass is in good agreement, but the thinnest/flattest galaxies are not present in the simulations. At fixed stellar mass, the aperture velocity dispersions are too low. Yet when we combine size and velocity dispersions into dynamical masses, and compare these to stellar masses, there is an excellent agreement with observations. The ratio of velocity and velocity dispersion  $V/\sigma$  match remarkably well with observations. However, not all regions of the  $V/\sigma$ - $\epsilon_e$  space that are covered by observations are matched by EAGLE<sup>+</sup>. The median stellar age of galaxies in EAGLE<sup>+</sup> is higher than in observations, and galaxies with young stellar populations are absent. We do not recover the recently discovered observational trend between intrinsic ellipticity and age in EAGLE<sup>+</sup>. This is can be attributed to the fact that both young and intrinsically flat galaxies are missing, thus we cannot conclude whether a relation between intrinsic ellipticity and age exists in EAGLE<sup>+</sup>. We find the best match between observations and the EAGLE<sup>+</sup> simulations in the low to mid stellar mass regime below  $\log(M_\star/M_\odot) < 11$ . At higher stellar masses, EAGLE<sup>+</sup> galaxies tend to retain or build-up too much spin.

The HORIZON-AGN simulations provide a qualitatively decent match to the observations, i.e., many of the trends with stellar mass are recovered. Quantitatively, there are several parameters where the median is significantly lower or higher, or where the range in HORIZON-AGN is smaller as compared to observations. While the size-mass relation has the right slope and scatter, we find that simulated galaxies are offset towards larger size. As a function of stellar mass, galaxies in HORIZON-AGN become rounder, but the median ellipticity is too low at all stellar masses, except above  $\log(M_\star/M_\odot) > 11.5$ . Aperture velocity dispersions are too low as compared to observations, and the slope of the  $\sigma$ -mass relation is slightly too shallow. When comparing dynamical masses to stellar masses we find trends similar to observations, but with less scatter at fixed stellar mass. At the highest stellar mass galaxies ( $\log(M_\star/M_\odot) > 11.3$ ), we find a good match between the ratio of velocity and velocity dispersion  $V/\sigma$ . For galaxies with stellar mass  $\log(M_\star/M_\odot) < 11$ , fast-rotating galaxies with  $(V/\sigma)_e > 0.75$  appear to be ab-

sent. HORIZON-AGN recovers the  $(V/\sigma)_e$ - $\epsilon_e$  relation across the range that it resolves, but a large fraction of the  $(V/\sigma)_e$ - $\epsilon_e$  space covered by observations is not occupied by HORIZON-AGN galaxies. We find an excellent match in the median age of galaxies, but the spread in age is smaller in the simulations. We recover a weak trend between intrinsic ellipticity and age in HORIZON-AGN, but for a given stellar age, we find that HORIZON-AGN have intrinsic ellipticity that are too low as compared to observations. The best match between observations and simulation is observed at high-stellar masses ( $\log(M_\star/M_\odot) > 11.3$ ). At all other masses galaxies are typically too big, too round, with values of  $(V/\sigma)_e$  that are too low.

For the MAGNETICUM simulations, we find a good qualitative and quantitative match for many of the trends with stellar mass and (intrinsic) ellipticity. A complication in all the comparisons of the MAGNETICUM data with observations is the relatively small volume of the cosmological box. Nonetheless, the size-mass relation is well reproduced, although the median size of galaxies below  $\log(M_\star/M_\odot) < 11.5$  is too large. Flat observed galaxies (high  $\epsilon$ ) are produced in MAGNETICUM, but there is a deficit of extremely round galaxies. Galaxies decrease in  $\epsilon$  with increasing stellar mass, but the onset of this decline appears to at lower stellar mass as compared to observations. Velocity dispersions are too low at all stellar masses, but dynamical mass estimates agree well with observations.  $V/\sigma$  measurements in MAGNETICUM are moderately too low at fixed stellar mass and the declining trend of  $V/\sigma$  with stellar mass starts at lower stellar mass as compared to observations. We find a good match in  $(V/\sigma)_e$ - $\epsilon_e$  space, i.e., the simulated galaxies follow the same trend as the observations, whilst also producing slow-rotating galaxies. However, at fixed ellipticity high  $(V/\sigma)_e$  galaxies values are missing, and the overall distribution of  $(V/\sigma)_e$  is lower as compared to observations, i.e., the relative number of fast spinning galaxies is still too low. The derived mean stellar age in MAGNETICUM is always higher than the observed age distribution. We find a clear confirmation of the relation between age and intrinsic ellipticity in MAGNETICUM albeit with an offset towards older ages. We find the best match between observations and the MAGNETICUM simulations in the mid-to-high stellar mass regime above  $\log(M_\star/M_\odot) > 10.75$ . While there is not a strong trend between how well the simulations match the observations and stellar mass, it is clear that higher particle resolution is needed to test MAGNETICUM predictions at low stellar mass.

The implications of these results can be separated into two stellar mass regimes. Without comparing simulations to each other, we find that the EAGLE<sup>+</sup> simulations reproduce several of the relations at low and intermediate stellar masses ( $\log(M_\star/M_\odot) < 11$ ) better than at high stellar mass. The HORIZON-AGN simulations reproduce the properties of the most massive galaxies ( $\log(M_\star/M_\odot) > 11.3$ ) considerably better as compared to low-mass HORIZON-AGN galaxies. Similarly, MAGNETICUM also appears to perform better at high-stellar mass than low stellar mass.

As shown in Dubois et al. (2016), the biggest impact of their two-mode AGN-feedback in HORIZON-AGN is at the massive end of the galaxy mass function. In EAGLE the treatment of AGN feedback is implemented in a different way, but given that the most massive galaxies typically retain too much spin, this suggests that the feedback is not effec-

tive enough. This agrees with the fact that EAGLE<sup>+</sup> galaxies in clusters are about a factor of 2 too massive and have star formation rates that are higher than observed in brightest cluster galaxies (Bahé et al. 2017; Barnes et al. 2017). In Lagos et al. (2018b) this was seen as the fraction of slow rotators being too low at the highest stellar masses compared to observations. In HORIZON-AGN, at  $\log(M_{\star}/M_{\odot}) < 11$ , where AGN feedback is less dominant, the simulations do not provide a good quantitative match to the observations.

Cosmological hydrodynamical simulations are rich tools to disentangle the relative importance of the physical processes involved in galaxy formation and evolution. The comparison with IFS measurements has shown that some fundamental parameters that describe a galaxy are not well reproduced in simulations. Moreover, the areas of discrepancy and agreement are different in the various simulations. Thus, there are a number of caveats which concern the validity of using these mismatching parameters to study the evolution of galaxies. Our comparison has highlighted several areas where the simulations might be improved, among which are the limiting particle resolution and the ISM model.

Based on our results, we consider the next major challenge for cosmological hydrodynamical simulations of galaxy formation is to produce Milky Way analogue galaxies that have both dynamically *and* chemically distinct thick and thin disks (e.g., Navarro et al. 2018). In this task, it will be essential to reproduce realistic vertical structure in galaxy disks. The next major observational step will be to gather larger statistical samples of resolved stellar kinematic measurements (e.g., Hector, Bryant et al. 2016) and to go beyond  $z \sim 0.1$ . With ever growing IFS surveys on galaxies near and far, the next generation of simulations will have even better samples to which to calibrate.

## ACKNOWLEDGEMENTS

We thank Joop Schaye, Klaus Dolag, Matthieu Schaller, and Tom Theuns for useful discussions and constructive comments.

The SAMI Galaxy Survey is based on observations made at the Anglo-Australian Telescope. The Sydney-AAO Multi-object Integral-field spectrograph (SAMI) was developed jointly by the University of Sydney and the Australian Astronomical Observatory, and funded by ARC grants FF0776384 (Bland-Hawthorn) and LE130100198. JvdS is funded under Bland-Hawthorn’s ARC Laureate Fellowship (FL140100278). CL has received funding from a Discovery Early Career Researcher Award (DE150100618) and by the ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. RSR and FS acknowledge support from the DAAD through funds from the Federal Ministry of Education and Research (BMBF). SB acknowledges the funding support from the Australian Research Council through a Future Fellowship (FT140101166). JJB acknowledges support of an Australian Research Council Future Fellowship (FT180100231). LC is the recipient of an Australian Research Council Future Fellowship (FT180100066) funded by the Australian Government. NS acknowledges support of a University of Sydney Postdoctoral Research Fellowship. Support for AMM is provided by NASA through Hubble

Fellowship grant #HST-HF2-51377 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555.

The SAMI input catalogue is based on data taken from the Sloan Digital Sky Survey, the GAMA Survey and the VST ATLAS Survey. The SAMI Galaxy Survey is supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013, the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020, and other participating institutions. The SAMI Galaxy Survey website is <http://sami-survey.org/>.

The MAGNETICUM Pathfinder simulations were partially performed at the Leibniz-Rechenzentrum with CPU time assigned to the Project ‘pr86re’. This work was supported by the DFG Cluster of Excellence ‘Origin and Structure of the Universe’. We are especially grateful for the support by M. Petkova through the Computational Center for Particle and Astrophysics (C2PAP).

This study uses data provided by the Calar Alto Legacy Integral Field Area (CALIFA) survey (<http://califa.caha.es/>). Based on observations collected at the Centro Astronómico Hispano Alemán (CAHA) at Calar Alto, operated jointly by the Max-Planck-Institut für Astronomie and the Instituto de Astrofísica de Andalucía (CSIC).

## REFERENCES

- Allen J. T., et al., 2015, *MNRAS*, **446**, 1567  
 Bacon R., et al., 2001, *MNRAS*, **326**, 23  
 Bahé Y. M., et al., 2017, *MNRAS*, **470**, 4186  
 Baldry I. K., et al., 2012, *MNRAS*, **421**, 621  
 Barnes D. J., et al., 2017, *MNRAS*, **471**, 1088  
 Bendo G. J., Barnes J. E., 2000, *MNRAS*, **316**, 315  
 Binney J., 2005, *MNRAS*, **363**, 937  
 Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, **54**, 529  
 Bland-Hawthorn J., et al., 2011, *Optics Express*, **19**, 2649  
 Bois M., et al., 2010, *MNRAS*, **406**, 2405  
 Bois M., et al., 2011, *MNRAS*, **416**, 1654  
 Brough S., et al., 2017, *ApJ*, **844**, 59  
 Bruzual G., Charlot S., 2003, *MNRAS*, **344**, 1000  
 Bryant J. J., Bland-Hawthorn J., 2012, in *Ground-based and Airborne Instrumentation for Astronomy IV*. p. 84466K, [doi:10.1117/12.925127](https://doi.org/10.1117/12.925127)  
 Bryant J. J., O’Byrne J. W., Bland-Hawthorn J., Leon-Saval S. G., 2011, *MNRAS*, **415**, 2173  
 Bryant J. J., Bland-Hawthorn J., Fogarty L. M. R., Lawrence J. S., Croom S. M., 2014, *MNRAS*, **438**, 869  
 Bryant J. J., et al., 2015, *MNRAS*, **447**, 2857  
 Bryant J. J., et al., 2016, preprint, ([arXiv:1608.03921](https://arxiv.org/abs/1608.03921))  
 Cañas R., Elahi P. J., Welker C., Lagos C. d. P., Power C., Dubois Y., Pichon C., 2018, preprint, ([arXiv:1806.11417](https://arxiv.org/abs/1806.11417))  
 Capak P., et al., 2007, *ApJS*, **172**, 99  
 Cappellari M., 2002, *MNRAS*, **333**, 400  
 Cappellari M., 2016, *ARA&A*, **54**, 597  
 Cappellari M., 2017, *MNRAS*, **466**, 798  
 Cappellari M., Copin Y., 2003, *MNRAS*, **342**, 345  
 Cappellari M., Emsellem E., 2004, *PASP*, **116**, 138  
 Cappellari M., et al., 2006, *MNRAS*, **366**, 1126  
 Cappellari M., et al., 2007, *MNRAS*, **379**, 418  
 Cappellari M., et al., 2011, *MNRAS*, **413**, 813

- Cappellari M., et al., 2012, *Nature*, 484, 485
- Cappellari M., et al., 2013a, *MNRAS*, 432, 1709
- Cappellari M., et al., 2013b, *MNRAS*, 432, 1862
- Chabrier G., 2003, *PASP*, 115, 763
- Chisari N., et al., 2015, *MNRAS*, 454, 2736
- Choi H., Yi S. K., 2017, *ApJ*, 837, 68
- Choi H., Yi S. K., Dubois Y., Kimm T., Devriendt J. E. G., Pichon C., 2018, *ApJ*, 856, 114
- Correa C. A., Schaye J., Clauwens B., Bower R. G., Crain R. A., Schaller M., Theuns T., Thob A. C. R., 2017, *MNRAS*, 472, L45
- Cortese L., et al., 2016, *MNRAS*,
- Cox T. J., Dutta S. N., Di Matteo T., Hernquist L., Hopkins P. F., Robertson B., Springel V., 2006, *ApJ*, 650, 791
- Crain R. A., et al., 2015, *MNRAS*, 450, 1937
- Croom S. M., et al., 2012, *MNRAS*, 421, 872
- Davies R. L., Efstathiou G., Fall S. M., Illingworth G., Schechter P. L., 1983, *ApJ*, 266, 41
- DeFelippis D., Genel S., Bryan G. L., Fall S. M., 2017, *ApJ*, 841, 16
- Devriendt J., et al., 2010, *MNRAS*, 403, L84
- Driver S. P., et al., 2011, *MNRAS*, 413, 971
- Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, *MNRAS*, 433, 3297
- Dubois Y., et al., 2014, *MNRAS*, 444, 1453
- Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, *MNRAS*, 463, 3948
- Emsellem E., Monnet G., Bacon R., 1994, *A&A*, 285
- Emsellem E., et al., 2004, *MNRAS*, 352, 721
- Emsellem E., et al., 2007, *MNRAS*, 379, 401
- Emsellem E., et al., 2011, *MNRAS*, 414, 888
- Falcón-Barroso J., et al., 2017, *A&A*, 597, A48
- Forbes D. A., Sinpetru L., Savorgnan G., Romanowsky A. J., Usher C., Brodie J., 2017, *MNRAS*, 464, 4611
- Foster C., et al., 2017, *MNRAS*, 472, 966
- Furlong M., et al., 2017, *MNRAS*, 465, 722
- Genel S., Fall S. M., Hernquist L., Vogelsberger M., Snyder G. F., Rodriguez-Gomez V., Sijacki D., Springel V., 2015, *ApJ*, 804, L40
- González Delgado R. M., Cerviño M., Martins L. P., Leitherer C., Hauschildt P. H., 2005, *MNRAS*, 357, 945
- González Delgado R. M., et al., 2015, *A&A*, 581, A103
- Governato F., et al., 2010, *Nature*, 463, 203
- Green A. W., et al., 2017, preprint, ([arXiv:1707.08402](https://arxiv.org/abs/1707.08402))
- Green A. W., et al., 2018, *MNRAS*, 475, 716
- Grogin N. A., et al., 2011, *ApJS*, 197, 35
- Guedes J., Callegari S., Madau P., Mayer L., 2011, *ApJ*, 742, 76
- Hill G. J., et al., 2008, in Kodama T., Yamada T., Aoki K., eds, *Astronomical Society of the Pacific Conference Series Vol. 399, Panoramic Views of Galaxy Formation and Evolution*. p. 115 ([arXiv:0806.0183](https://arxiv.org/abs/0806.0183))
- Hill D. T., et al., 2011, *MNRAS*, 412, 765
- Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, *MNRAS*, 442, 2304
- Husemann B., et al., 2013, *A&A*, 549, A87
- Illingworth G., 1977, *ApJ*, 218, L43
- Jesseit R., Cappellari M., Naab T., Emsellem E., Burkert A., 2009, *MNRAS*, 397, 1202
- Kaufmann T., Mayer L., Wadsley J., Stadel J., Moore B., 2007, *MNRAS*, 375, 53
- Kaviraj S., et al., 2017, *MNRAS*, 467, 4739
- Kelvin L. S., et al., 2012, *MNRAS*, 421, 1007
- Koekemoer A. M., et al., 2007, *ApJS*, 172, 196
- Koekemoer A. M., et al., 2011, *ApJS*, 197, 36
- Komatsu E., et al., 2011, *ApJS*, 192, 18
- Krajnović D., et al., 2011, *MNRAS*, 414, 2923
- Kregel M., van der Kruit P. C., de Grijs R., 2002, *MNRAS*, 334, 646
- Lackner C. N., Cen R., Ostriker J. P., Joung M. R., 2012, *MNRAS*, 425, 641
- Lagos C. d. P., Theuns T., Stevens A. R. H., Cortese L., Padilla N. D., Davis T. A., Contreras S., Croton D., 2017, *MNRAS*, 464, 3850
- Lagos C. d. P., et al., 2018a, *MNRAS*, 473, 4956
- Lagos C. d. P., Schaye J., Bahé Y., Van de Sande J., Kay S. T., Barnes D., Davis T. A., Dalla Vecchia C., 2018b, *MNRAS*, 476, 4327
- Lange R., et al., 2015, *MNRAS*, 447, 2603
- Lange R., et al., 2016, *MNRAS*, 462, 1470
- Liske J., et al., 2015, *MNRAS*, 452, 2087
- Ma C.-P., Greene J. E., McConnell N., Janish R., Blakeslee J. P., Thomas J., Murphy J. D., 2014, *ApJ*, 795, 158
- Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Pichon C., 2018, *MNRAS*,
- McAlpine S., et al., 2016, *Astronomy and Computing*, 15, 72
- McDermid R. M., et al., 2015, *MNRAS*, 448, 3484
- Moustakas J., et al., 2013, *ApJ*, 767, 50
- Naab T., et al., 2014, *MNRAS*, 444, 3357
- Navarro J. F., Steinmetz M., 2000, *ApJ*, 538, 477
- Navarro J. F., et al., 2018, *MNRAS*, 476, 3648
- Owers M. S., et al., 2017, *MNRAS*, 468, 1824
- Pedrosa S. E., Tissera P. B., 2015, *A&A*, 584, A43
- Penoyre Z., Moster B. P., Sijacki D., Genel S., 2017, *MNRAS*, 468, 3883
- Pillepich A., et al., 2018, *MNRAS*, 475, 648
- Planck Collaboration et al., 2014, *A&A*, 571, A31
- Remus R.-S., Dolag K., Naab T., Burkert A., Hirschmann M., Hoffmann T. L., Johansson P. H., 2017, *MNRAS*, 464, 3742
- Robotham A. S. G., et al., 2014, *MNRAS*, 444, 3986
- Roth M. M., et al., 2005, *PASP*, 117, 620
- Salpeter E. E., 1955, *ApJ*, 121, 161
- Sánchez-Blázquez P., et al., 2006, *MNRAS*, 371, 703
- Sánchez S. F., et al., 2012, *A&A*, 538, A8
- Sánchez S. F., et al., 2016a, *Rev. Mex. Astron. Astrofis.*, 52, 171
- Sánchez S. F., et al., 2016b, *A&A*, 594, A36
- Saunders W., et al., 2004, in Moorwood A. F. M., Iye M., eds, *Proc. SPIE Vol. 5492, Ground-based Instrumentation for Astronomy*. pp 389–400, [doi:10.1117/12.550871](https://doi.org/10.1117/12.550871)
- Schaye J., et al., 2015, *MNRAS*, 446, 521
- Schiavon R. P., 2007, *ApJS*, 171, 146
- Schulze F., Remus R.-S., Dolag K., Burkert A., Emsellem E., van de Ven G., 2018, *MNRAS*, 480, 4636
- Scott N., et al., 2013, *MNRAS*, 432, 1894
- Scott N., et al., 2017, *MNRAS*, 472, 2833
- Scott N., et al., 2018, preprint, ([arXiv:1808.03365](https://arxiv.org/abs/1808.03365))
- Shanks T., et al., 2013, *The Messenger*, 154, 38
- Sharp R., et al., 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62690G ([arXiv:astro-ph/0606137](https://arxiv.org/abs/astro-ph/0606137)), [doi:10.1117/12.671022](https://doi.org/10.1117/12.671022)
- Sharp R., et al., 2015, *MNRAS*, 446, 1551
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *MNRAS*, 343, 978
- Smith G. A., et al., 2004, in Moorwood A. F. M., Iye M., eds, *Proc. SPIE Vol. 5492, Ground-based Instrumentation for Astronomy*. pp 410–420, [doi:10.1117/12.551013](https://doi.org/10.1117/12.551013)
- Springel V., Yoshida N., White S. D. M., 2001, *New Astron.*, 6, 79
- Steinmetz M., Navarro J. F., 1999, *ApJ*, 513, 555
- Swinbank A. M., et al., 2017, *MNRAS*,
- Szomoru D., Franx M., van Dokkum P. G., Trenti M., Illingworth G. D., Labbé I., Oesch P., 2013, *ApJ*, 763, 73
- Taranu D. S., Dubinski J., Yee H. K. C., 2013, *ApJ*, 778, 61
- Taylor E. N., et al., 2011, *MNRAS*, 418, 1587
- Teklu A. F., Remus R.-S., Dolag K., Beck A. M., Burkert A., Schmidt A. S., Schulze F., Steinborn L. K., 2015, *ApJ*, 812,

29

- Teklu A. F., Remus R.-S., Dolag K., Burkert A., 2017, *MNRAS*, **472**, 4769
- Teyssier R., 2002, *A&A*, **385**, 337
- Trayford J. W., et al., 2015, *MNRAS*, **452**, 2879
- Trayford J. W., et al., 2017, *MNRAS*, **470**, 771
- Valdes F., Gupta R., Rose J. A., Singh H. P., Bell D. J., 2004, *ApJS*, **152**, 251
- Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, *MNRAS*, **404**, 1639
- Veale M., Ma C.-P., Greene J. E., Thomas J., Blakeslee J., McConnell N., Walsh J., Ito J., 2017a, preprint, ([arXiv:1703.08573](https://arxiv.org/abs/1703.08573))
- Veale M., et al., 2017b, *MNRAS*, **464**, 356
- Veale M., Ma C.-P., Greene J. E., Thomas J., Blakeslee J. P., Walsh J. L., Ito J., 2018, *MNRAS*, **473**, 5446
- Vogelsberger M., et al., 2014, *MNRAS*, **444**, 1518
- Walcher C. J., et al., 2014, *A&A*, **569**, A1
- Wang L., et al., 2018, [ArXiv:1808.05564](https://arxiv.org/abs/1808.05564),
- Weijmans A.-M., et al., 2014, *MNRAS*, **444**, 3340
- Welker C., Dubois Y., Devriendt J., Pichon C., Kaviraj S., Peirani S., 2017, *MNRAS*, **465**, 1241
- Wright A. H., et al., 2017, *MNRAS*, **470**, 283
- York D. G., et al., 2000, *AJ*, **120**, 1579
- Zavala J., Okamoto T., Frenk C. S., 2008, *MNRAS*, **387**, 364
- Zavala J., et al., 2016, *MNRAS*, **460**, 4466
- Zibetti S., Charlot S., Rix H.-W., 2009, *MNRAS*, **400**, 1181
- de Zeeuw P. T., et al., 2002, *MNRAS*, **329**, 513
- van Dokkum P. G., Conroy C., 2010, *Nature*, **468**, 940
- van de Sande J., et al., 2017a, *MNRAS*, **472**, 1272
- van de Sande J., et al., 2017b, *ApJ*, **835**, 104
- van de Sande J., et al., 2018, *Nature Astronomy*,
- van der Wel A., et al., 2014, *ApJ*, **788**, 28

<sup>1</sup>*Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW, 2006, Australia*

<sup>2</sup>*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*

<sup>3</sup>*International Centre for Radio Astronomy Research, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia*

<sup>4</sup>*Universitäts-Sternwarte München, Scheinerstr. 1, D-81679 München, Germany*

<sup>5</sup>*Max Planck Institute for Extraterrestrial Physics, Giessenbachstraße 1, D-85748 Garching, Germany*

<sup>6</sup>*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

<sup>7</sup>*School of Physics, University of New South Wales, NSW 2052, Australia*

<sup>8</sup>*Australian Astronomical Optics, AAO-USydney, School of Physics, University of Sydney, NSW 2006, Australia*

<sup>9</sup>*University of Oxford, Astrophysics, Keble Road, Oxford OX1 3RH, UK*

<sup>10</sup>*CNRS and UPMC Univ. Paris 06, UMR 7095, Institut d'Astrophysique de Paris, 98bis boulevard Arago, 75014 Paris, France*

<sup>11</sup>*Australian Astronomical Optics, Faculty of Science and Engineering, Macquarie University, 105 Delhi Rd, North Ryde, NSW 2113, Australia*

<sup>12</sup>*Atlassian, 341 George St, Sydney, NSW 2000, Australia*

<sup>13</sup>*Ritter Astrophysical Research Center University of Toledo Toledo, OH 43606, USA*

<sup>14</sup>*Research School of Astronomy and Astrophysics, Australian National University, Canberra ACT 2611, Australia*

<sup>15</sup>*Hubble Fellow*

<sup>16</sup>*Korea Institute for Advanced Study (KIAS), 85 Hoegiro, Dongdaemun-gu, Seoul, 02455, Republic of Korea*

<sup>17</sup>*SOFIA Science Center, USRA, NASA Ames Research Center, Building N232, M/S 232-12, P.O. Box 1, Moffett Field, CA 94035-0001, USA*

<sup>18</sup>*Instituto de Astronomía, Universidad Nacional Autónoma de México, A. P. 70-264, C.P. 04510 México, D.F. Mexico*

<sup>19</sup>*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia*

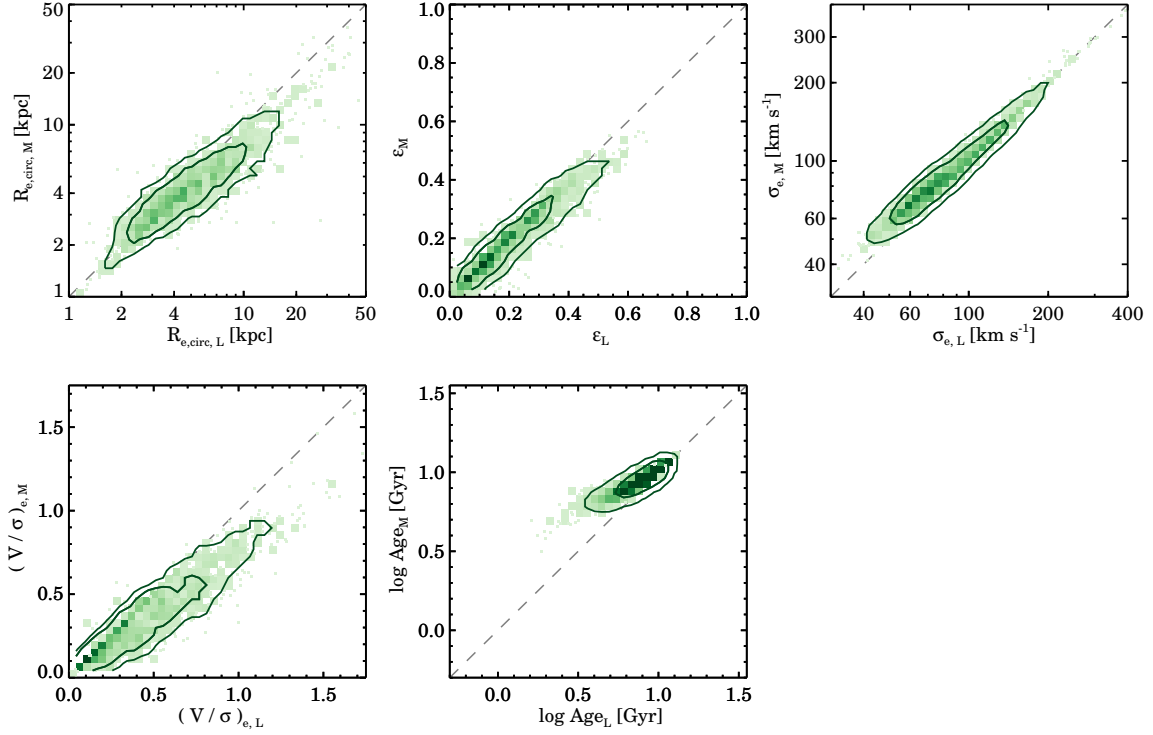
## APPENDIX A: LUMINOSITY AND MASS WEIGHTED QUANTITIES

To provide the best-match between the simulations and IFS observations, in this paper all mock-observations are luminosity weighted. In this appendix, we study the impact of using  $r$ -band luminosity-weighted versus stellar mass-weighted quantities. In Fig. 1-2 we compare these two different weighting approaches for the five core measurements used in this work: effective radius  $R_e$ , ellipticity  $\epsilon_e$ , velocity dispersion  $\sigma_e$ ,  $(V/\sigma)_e$ , and the mean stellar population age.

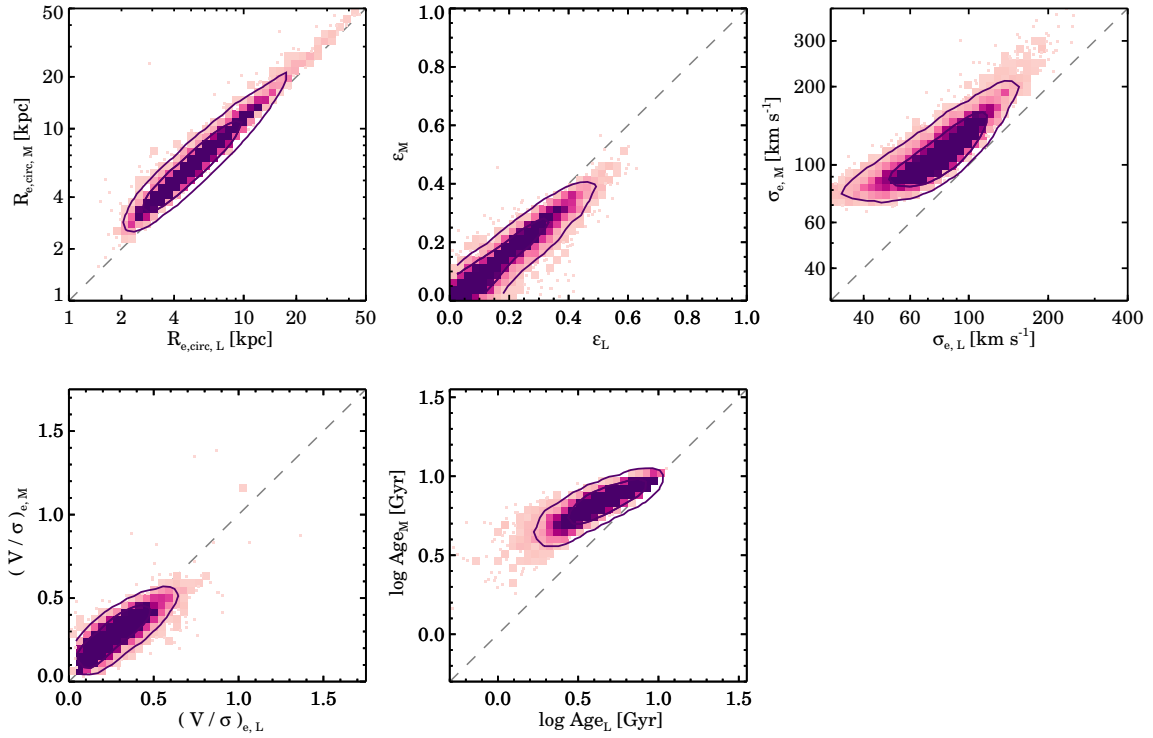
For the EAGLE simulations (Fig. 1), we find trends consistent with observations. Mass-weighted effective radii are smaller than mass-weighted radii, as expected for galaxies with negative colour gradients where the cores are relatively red (e.g., Zibetti et al. 2009; Szomoru et al. 2013). Luminosity-weighted ellipticities are larger, i.e., galaxies appear to be flatter. This is consistent with a simple morphological picture, where for a relatively red bulge and bluer disk, when randomly observed, the luminosity-weighting increases the ellipticity because more weight is given to the disk component. Similarly, because the velocity dispersion of the disk component is typically much lower than the bulge, a luminosity-weighted velocity dispersion is expected to be lower than a mass-weighted velocity dispersion. We find that this trend is recovered in the EAGLE simulations. Furthermore, the lower velocity dispersion will also cause the  $V/\sigma$  values to be higher when using luminosity weighting as compared to mass weighting. Finally, luminosity weighted stellar ages in EAGLE are younger, consistent with observational results (e.g., McDermid et al. 2015; González Delgado et al. 2015).

In HORIZON-AGN (Fig. 2), most mock-observed quantities follow the expectations from the observations, with the exception of the size measurements. The luminosity-weighted effective radii in HORIZON-AGN are smaller than the mass-weighted quantities. We note that a burst of central star formation can make a galaxy more compact when using luminosity-weighted sizes. However, from the fact that galaxies of all sizes appear to follow the same trend, we consider this scenario to be unlikely. As the effect is small, we consider a further investigation of this result beyond the scope of the paper. We also note that while the HORIZON-AGN results for the velocity dispersion are quantitatively the same as in EAGLE, the difference between luminosity and mass-weighted dispersion is considerably stronger in HORIZON-AGN.

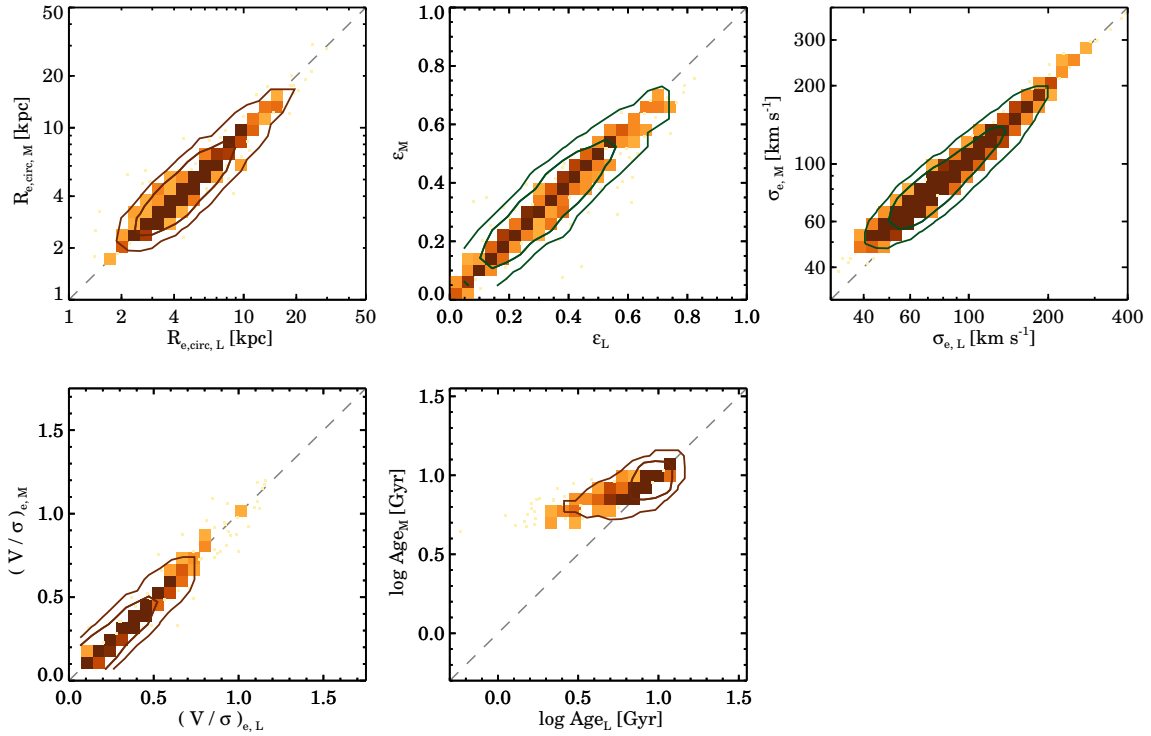
Lastly, all mock-observations from the MAGNETICUM



**Figure 1.** Comparison of simulated luminosity and mass-weighted quantities in EAGLE. The assumption of using luminosity-weighted measurements rather than mass-weighting, results in larger effective radii, higher or more flattened ellipticities, lower velocity dispersions, higher  $(V/\sigma)_e$ , and younger mean ages.



**Figure 2.** Comparison of simulated luminosity and mass-weighted quantities in HORIZON-AGN. Similar to EAGLE, we find that with luminosity-weighted ellipticities are higher, velocity dispersions are considerably lower,  $(V/\sigma)_e$  values are higher, and ages are younger. However, the luminosity-weighted effective radii are smaller than the mass-weighted radii, in contrast with observational results.



**Figure 3.** Comparison of simulated luminosity and mass-weighted quantities in MAGNETICUM. Similar to EAGLE, we find that with luminosity-weighted sizes and ellipticities are higher, velocity dispersions are considerably lower,  $(V/\sigma)_e$  values are slightly higher, and ages are younger.

simulation follow the expected observational trends. The results are the same as for EAGLE: luminosity-weighted sizes are larger than the mass-weighted size, ellipticities are more flattened, velocity dispersions are considerably lower,  $(V/\sigma)_e$  values are higher, and ages are younger. However, the difference in using mass-weighted versus luminosity-weighted quantities is small for all but mean stellar age.

## APPENDIX B: STELLAR POPULATION AGE IN SAMI, CALIFA, AND ATLAS<sup>3D</sup>

As described in Section 2, mean stellar ages have been measured by Scott et al. (2017) for the SAMI Galaxy Survey data, McDermid et al. (2015) for ATLAS<sup>3D</sup> data. Both use a method where lick indices are converted into single stellar population. For the CALIFA survey, González Delgado et al. (2015) adopt a full spectral fitting method. Besides the different methods, the wavelength regions from which the stellar population parameters have been derived are also different: 3750–5750Å for SAMI, 4800–5380Å for ATLAS<sup>3D</sup> and 3745–7300Å for CALIFA. Because of the different wavelength regions covered by the different surveys we expect differences in the mean age estimates as blue wavelengths are more sensitive to young stellar populations. Therefore, in this appendix we compare the mean stellar age distribution of galaxies from the SAMI Galaxy Survey, ATLAS<sup>3D</sup>, and CALIFA.

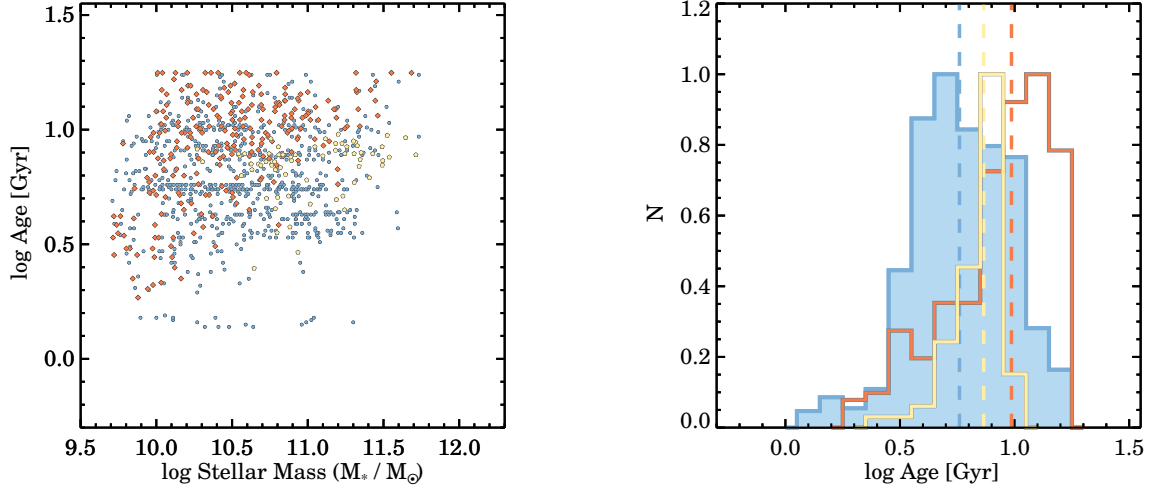
In Fig. B1 left-panel, we compare the mean  $\log Age$  as a function of stellar mass for the three different surveys for early-type galaxies only (E and S0). We find that ATLAS<sup>3D</sup> galaxies are clearly offset to older ages, but that this is not

caused by a difference in the stellar mass distribution of the different samples. At fixed stellar mass, ATLAS<sup>3D</sup> galaxies have higher age estimates as compared to SAMI and CALIFA. Note that there are almost no galaxies in the CALIFA Survey with ages older than 10Gyr, while there are a considerable number in SAMI and ATLAS<sup>3D</sup>. This is most likely caused by the different fitting methods.

We show the distribution of  $\log Age$  in the right-panel of Fig. B1. The SAMI distribution shows the largest spread in age, but also has the lowest median  $\log Age$  in Gyr of 0.76. The CALIFA age distribution is considerably narrower, with a slightly higher median age ( $\log age [\text{Gyr}] = 0.87$ ), whereas the ATLAS<sup>3D</sup> age distribution is noticeably offset and skewed towards older ages (median  $\log age [\text{Gyr}] = 0.99$ ). Thus, in order to create a homogeneous observational dataset, we subtract 0.23 dex from the ATLAS<sup>3D</sup> ages. We do not apply an offset to the CALIFA sample because we do not find a significant offset in  $\log Age$  for the entire SAMI and CALIFA sample without a morphological cut (median  $\log Age$  in Gyr of 0.63 and 0.61, respectively).

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





**Figure B1.** Comparison of the mean stellar ages from the SAMI Galaxy Survey (blue), ATLAS<sup>3D</sup> (orange), and CALIFA (yellow). At fixed stellar mass, ATLAS<sup>3D</sup> age estimates are higher with a median  $\log Age$  [Gyr] of 0.99 as compared to SAMI (0.76) and CALIFA (0.87).