[OII] emitters in MULTIDARK-GALAXIES and DEEP2

G. Favole,[1][*] V. Gonzalez-Perez,[2,3,1][†] D. Stoppacher,[4,5] Á. Orsi,[6] J. Comparat,[7] S. A. Cora,[8,9] C. A. Vega-Martínez,[10,11] A. R. H. Stevens,[12] C. Maraston,[1] D. Croton,[13] A. Knebe,[5,12,14] A. J. Benson,[15] A. D. Montero-Dorta,[16] N. Padilla,[17,18] F. Prada,[19] D. Thomas[1]

[1]*Institute of Cosmology and Gravitation, Portsmouth University, Burnaby Road, Portsmouth PO13FX, UK*
[2]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK.*
[3]*Energy Lancaster, Lancaster University, LA1 4YB, UK*
[4]*Instituto de Física Teórica (IFT) UAM/CSIC, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[5]*Departamento de Física Teórica, Módulo 15, Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*
[6]*Centro de Estudios de Física del Cosmos de Aragón, Plaza de San Juan 1, Teruel E-44001, Spain*
[7]*Max-Planck-Institut für extraterrestrische Physik (MPE), Giessenbachstrasse 1, D-85748 Garching bei München, Germany*
[8]*Instituto de Astrofísica de La Plata (CCT La Plata, CONICET, UNLP), Paseo del Bosque s/n, B1900FWA, La Plata, Argentina*
[9]*Facultad de Ciencias Astronómicas y Geofísicas, UNLP, Paseo del Bosque s/n, B1900FWA, La Plata, Argentina*
[10]*Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán 1305, La Serena, Chile*
[11]*Departamento de Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile*
[12]*International Centre for Radio Astronomy Research, The University of Western Australia, Crawley, WA6009, Australia*
[13]*Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, Victoria 3122, Australia*
[14]*Centro de Investigación Avanzada en Física Fundamental (CIAFF), Facultad de Ciencias, UAM, 28049 Madrid, Spain*
[15]*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[16]*Instituto de Física, Universidade de São Paulo, São Paulo, SP, Brazil*
[17]*Instituto de Astrofísica, Pontificia Universidad Católica, Av. Vicuña Mackenna 4860, Santiago, Chile*
[18] *Centro de Astro-Ingeniería, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Santiago, Chile*
[19]*Instituto de Astrofísica de Andalucía (IAA)/ CSIC, Granada, E-18008, Spain*

ABSTRACT

We use three semi-analytic models (SAMs) of galaxy formation and evolution run on the same $1h^{-1}$Gpc MultiDark Planck2 cosmological simulation to investigate the properties of [OII] emission line galaxies at redshift $z \sim 1$. We compare model predictions with different observational data sets, including DEEP2–FIREFLY galaxies with absolute magnitudes. We estimate the [OII] luminosity (L[OII]) of our model galaxies using the public code GET_EMLINES, which ideally assumes as input the instantaneous star formation rates (SFRs). This property is only available in one of the SAMs under consideration, while the others provide average SFRs, as most models do. We study the feasibility of inferring galaxies' L[OII] from average SFRs in post-processing. We find that the result is accurate for model galaxies with dust attenuated L[OII] $\lesssim 10^{42.2}$erg s^{-1} ($< 5\%$ discrepancy). The galaxy properties that correlate the most with the model L[OII] are the SFR and the observed-frame u and g broad-band magnitudes. Such correlations have r-values above 0.64 and a dispersion that varies with L[OII]. We fit these correlations with simple linear relations and use them as proxies for L[OII], together with an observational conversion that depends on SFR and metallicity. These proxies result in [OII] luminosity functions and halo occupation distributions with shapes that vary depending on both the model and the method used to derive L[OII]. The amplitude of the clustering of model galaxies with L[OII] $> 10^{40.4}$erg s^{-1} remains overall unchanged on scales above $1\,h^{-1}$Mpc, independently of the L[OII] computation.

Key words: galaxies: distances and redshifts — galaxies: haloes — galaxies: statistics — cosmology: observations — cosmology: theory — large-scale structure of Universe

1 INTRODUCTION

In the era of precision cosmology, surveys are starting to rely on star-forming galaxies to go further into early cosmic times, when dark energy is just starting to dominate the energy-matter budget of the Universe. Star-forming galaxies with strong nebular emission lines (ELGs) are among the preferred targets of the new generation of spectroscopic surveys as SDSS-IV/eBOSS (Dawson et al. 2016), DESI

[*] E-mail: ginevra.favole@port.ac.uk
[†] E-mail: violetagp@protonmail.com

(Schlegel et al. 2015), 4MOST[1], WFIRST[2], Subaru-PFS[3] and Euclid[4] (Laureijs et al. 2011; Sartoris et al. 2015), and will be used to trace the baryon acoustic oscillation (BAO; Eisenstein et al. 2005) scale and the growth of structure by measuring redshift-space distortions in the observed clustering (Alam et al. 2015; Mohammad et al. 2018; Orsi & Angulo 2018). Star-forming galaxies will also be fundamental to inform halo occupation distribution (HOD; Cooray & Sheth 2002; Berlind & Weinberg 2002; Kravtsov et al. 2004; Zheng et al. 2007) and (sub)halo abundance matching (SHAM; Conroy et al. 2006; Behroozi et al. 2010; Trujillo-Gomez et al. 2011; Nuza et al. 2013) models to generate fast mock galaxy catalogues useful for cosmological tests.

At $z \sim 1$ and for optical detectors, the samples of star-forming ELGs are dominated by [O II] emitters. Therefore, measuring and modelling the relations between redshift and the physical properties of these galaxies – such as [O II] luminosity with star formation rate (SFR) – is crucial for capitalising on the science that can be addressed from [O II] data. In this work, we aim to do exactly this, ultimately allowing us to build robust galaxy clustering predictions for near-future [O II] data sets dominated by star-forming galaxies.

Modelling emission lines requires, at least, a certain knowledge of both the gas and the star formation history of a given galaxy. [O II] emission is particularly difficult to predict, as it critically depends on local properties, such as dust attenuation, the structure of the H II regions and their ionisation fields. For this reason, [O II] traces star formation and metallicity in a non-trivial way (e.g., Kewley et al. 2004; Dickey et al. 2016). Previous works on [O II] emitters have shown that semi-analytic models of galaxy formation (SAMs) are ideal laboratories for studying the physical properties of these galaxies, since they can reproduce the observed [O II] luminosity function (LF) at $z \sim 1$ (Orsi et al. 2014; Comparat et al. 2015, 2016; Hirschmann et al. 2017). Gonzalez-Perez et al. (2018) explored how [O II] emitters are distributed in the dark matter haloes. They found typical host halo masses in agreement with the results from Favole et al. (2016a), which were based on a modified SHAM technique combining observational data with the MultiDark Planck dark matter N-body simulation (MDPL; Klypin et al. 2016).

For this project, we use the MULTIDARK-GALAXIES mock products, which are publicly available at https://www.cosmosim.org. These catalogues were produced using 3 different SAMs to populate the snapshots of the MultiDark2 (MDPL2; Klypin et al. 2016) dark matter cosmological simulation, over the redshift range $0 < z < 8$ (Knebe et al. 2018). MDPL2 is one of the largest dark matter simulation boxes with a volume of $1\,h^{-3}\mathrm{Gpc}^3$ and 3048^3 particles with mass resolution of $1.51 \times 10^9\,h^{-1}\mathrm{M}\odot$. The models used in the production of these catalogues were: SAG (Gargiulo et al. 2015; Muñoz Arancibia et al. 2015; Cora et al. 2018), SAGE (Croton et al. 2016) and GALACTICUS (Benson 2012).

In this work, we explore the limitations of estimating the [O II] luminosity in post-processing using different approaches, assessing how this quantity correlates with other galactic properties within the studied SAMs. The results from model galaxies are compared with observations from

DEEP2 (Newman et al. 2013). The DEEP2 spectra have been fitted using FIREFLY (Wilkinson et al. 2017; Comparat et al. 2017) to extract physical properties for these galaxies. Our final DEEP2 data set and SAM galaxy catalogues including emission line properties are made publicly available (see Sec. 6). In this analysis, we assume a Planck Collaboration et al. (2015) cosmology with $\Omega_\mathrm{m} = 0.6929$, $\Omega_\Lambda = 0.3071$, $h = 0.6777$.

The paper is organised as follows: in Section 2, we describe the semi-analytic models considered in our study, the DEEP2 observational data set and the FIREFLY code for spectral fitting. We compare the model SFR and stellar mass functions with current observations. In Section 3, we describe how we calculate the [O II] emission line luminosity in the SAMs using the publicly available code GET_EMLINES by Orsi et al. (2014) with instantaneous SFR and cold gas metallicity as inputs. We analyse the impact of using average rather instantaneous SFR in this calculation to be used in those SAMs that do not provide instantaneous quantities. We compare the derived [O II] luminosity functions with current observations. In Section 4, we explore the correlations between $L[\mathrm{O\,II}]$ and several galactic properties to establish model proxies for the [O II] luminosity. We provide scaling relations among these quantities that can be used in models without an emission line estimate. We further test these proxies by checking the consistency of the evolution of their [O II] luminosity functions and clustering signal with observations and direct predictions from SAMs. Section 5 summarises our findings.

2 DATA

2.1 Semi-analytic models

Semi-analytic models of galaxy formation (White & Frenk 1991; Kauffmann et al. 1993) encapsulate the key mechanisms that contribute to form galaxies in a set of coupled differential equations, allowing one to populate the dark matter haloes in cosmological N-body simulations with relative haste (see e.g., Baugh 2006; Benson 2010; Somerville & Davé 2015). In the last two decades, a huge effort has been made to improve these models and account for the physical processes that shape galaxy formation and evolution, such as gas cooling (e.g., De Lucia et al. 2010; Monaco et al. 2014; Hou et al. 2017), gas accretion (e.g., Guo et al. 2011; Henriques et al. 2013; Hirschmann et al. 2016), star formation (e.g., Lagos et al. 2011), stellar winds (e.g., Lagos et al. 2013), stellar evolution (e.g., Tonini et al. 2009; Henriques et al. 2011; Gonzalez-Perez et al. 2014), AGN feedback (e.g., Bower et al. 2006; Croton et al. 2006) or environmental processes (e.g., Weinmann et al. 2006; Font et al. 2008; Stevens & Brown 2017; Cora et al. 2018). Typically, SAMs do not attempt to resolve the scales on which these key astrophysical processes take place, but rather they describe their effects globally. Inevitably, this leads to free parameters in the models that require calibration; in essence, these compensate for the lack of understanding of certain processes and also for not resolving the relevant small scales.

In this study, we use the results from three semi-analytic models of galaxy formation: SAG (Cora 2006; Gargiulo et al. 2015; Muñoz Arancibia et al. 2015; Cora et al. 2018), SAGE (Croton et al. 2006, 2016) and Galacticus (Benson 2012). The three SAMs considered have been run on the same MultiDark2 $1\,h^{-1}\mathrm{Gpc}$ cosmological simulation with Planck

cosmology (Klypin et al. 2016) to produce mock galaxy catalogues[5].

The complete description of the first data release of the MULTIDARK-GALAXIES products including SAG, SAGE and GALACTICUS mock catalogues can be find in Knebe et al. (2018). All these catalogues lack [O II] luminosity estimates. A version of Galacticus does have an emission line calculation (Merson et al. 2018), but it has not been applied to the MultiDark models.

2.1.1 SAG

We consider a modified version of the Semi-Analytical Galaxies (SAG; Cora 2006; Lagos et al. 2008; Gargiulo et al. 2015; Muñoz Arancibia et al. 2015; Collacchioni et al. 2018; Cora et al. 2018) code, which involves a detailed chemical model and implements an improved treatment of environmental effects (ram-pressure of both hot and cold gas phases and tidal stripping of gaseous and stellar components). It also includes the modelling of the strong galaxy emission lines in the optical and far-infrared range as described in Orsi et al. (2014). The free parameters of the model have been tuned by applying the Particle Swarm Optimisation technique (PSO; Ruiz et al. 2015) and using as constraints the stellar mass function at $z = 0$ and 2 (data compilations from Henriques et al. 2015), the SFR function at $z = 0.14$ (Gruppioni et al. 2015), the fraction of mass in cold gas as a function of stellar mass (Boselli et al. 2014), and the black hole–bulge mass relation (McConnell & Ma 2013; Kormendy & Ho 2013).

2.1.2 SAGE

The Semi-Analytic Galaxy Evolution[6] code (SAGE; Croton et al. 2006, 2016) is a modular and customisable SAM. The updated physics includes gas accretion, ejection due to feedback, a new gas cooling–radio mode AGN heating cycle, AGN feedback in the quasar mode, galaxy mergers, disruption, and the build-up of intra-cluster stars.

SAGE was calibrated to reproduce several statistical features and scaling relations of galaxies at $z = 0$, including the stellar mass function, tightly matching the observational uncertainty range (Baldry et al. 2008), the black hole-bulge mass relation, the stellar mass-gas metallicity relation, and the Baryonic Tully-Fisher relation (Tully & Fisher 1977).
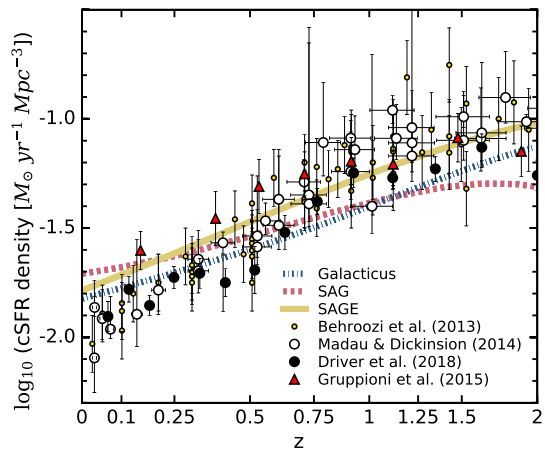
2.1.3 GALACTICUS

GALACTICUS[7] (Benson 2012) has much in common with the previous two models, in terms of modularity, the range of physical processes included and the type of quantities that it can predict. Although this model has not been re-tuned to this simulation, the original calibration was performed using analytically built merger trees assuming a WMAP7 cosmology (Benson 2012). The original model reproduced reasonably well the observed stellar mass function at $z \sim$ 0.07 (Li & White 2009) and the HI mass function at $z \sim 0$ (Martin et al. 2010).

---

[5] publicly available at https://www.cosmosim.org and http://skiesanduniverses.org/
[6] http://www.asvo.org.au/
[7] https://bitbucket.org/galacticusdev/galacticus/wiki/Home

Figure 1. Cosmic star formation rate density of SAG, SAGE and GALACTICUS MULTIDARK-GALAXIES as a function of redshift, compared to four independent compilations of data sets from Behroozi et al. (2013) (this was corrected to a Chabrier et al. (2014) IMF by the same authors), Madau & Dickinson (2014), Driver et al. (2018) and Gruppioni et al. (2015). The error bars are the 1σ dispersion around each point. We show this result only up to $z \sim 2$, which is the maximum redshift of interest for our study.

2.1.4 Model comparison

For a full comparison between the SAG, SAGE and GALACTICUS semi-analytic models adopted in this work, we refer the reader to Knebe et al. (2018). The main differences between them are: i) the calibrations; ii) the treatment of mergers; iii) galaxies without a host halo, "orphans", are not allowed in SAGE, while they can happen, due to mass stripping, within GALACTICUS and SAG; and iv) the metal enrichment models, with GALACTICUS and SAGE assuming an instantaneous recycling approximation and SAG implementing a more complete chemical model (Cora 2006; Collacchioni et al. 2018).

Here we also recall some results from Knebe et al. (2018) that are important for interpreting the outcomes of our analysis and a further study of global properties can be found in Appendix C. As we impose a minimum limit of $M_\star > 10^{8.87}\,M_\odot$ and SFR $> 0\,\mathrm{yr}^{-1}M_\odot$ to the three SAMs of interest, some of our model results will be slightly different from those presented in Knebe et al. (2018). The cuts above have been chosen taking into account the resolution limit of the MultiDark cosmological simulation (see Knebe et al. 2018). At $z \sim 1$, the limit on SFR excludes about 4% of the entire SAG population, 17% of galaxies in SAGE, and no galaxies from GALACTICUS.

Fig. 1 shows the redshift evolution of the MULTIDARK-GALAXIES cosmic star formation rate (SFR) density compared to a compilation of observations including estimates of the cosmic SFR from narrow-band (Hα), broad-band (UV-IR), and radio (1.4 GHz) surveys by Behroozi et al. (2013), and more recent results by Madau & Dickinson (2014), Gruppioni et al. (2015) and Driver et al. (2018). The observational data sets are consistent, despite being affected by different systematic errors. Fig. 1 only extends to $z \sim 2$, as higher redshifts are not of interest for this study. All the SAMs agree with the observations within our redshift range of interest $0.6 < z < 1.2$. Beyond $z = 2$, SAG and GALACTICUS model galaxies maintain a good agreement with the data out to $z \sim 8.5$, while SAGE overpredicts the SFR density at $z \gtrsim 4$ (see Knebe et al. 2018).
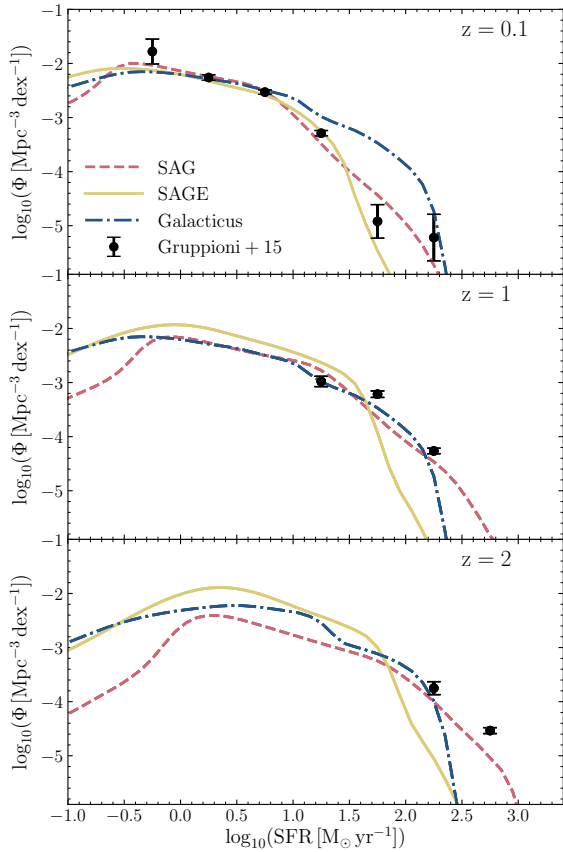
Figure 2. MULTIDARK-GALAXIES average SFR function evolution at $z \lesssim 2$ (lines) compared to the Herschel/PEP and HerMES observations (Gruppioni et al. 2015, filled circles).

In SAMs, galaxy properties are obtained by solving coupled differential equations in a certain number of steps in which the time interval between snapshots of the underlying DM simulation is divided. In this context, we define the "instantaneous SFR" as the star formation rate computed using the mass of stars formed over the last step before the output. The "average SFR" is instead the SFR obtained by considering the average contribution from all the steps. The SAG model subdivides the time between snapshots in 25. This timescale typically corresponds to $\sim$10-25 Myrs at $z \sim 1$, which is the timescale physically relevant for the [O II] emission. SAGE and GALACTICUS split time in 10 steps.

Fig. 2 displays the average SFR functions of the MULTIDARK-GALAXIES at different redshifts compared to the Herschel data from the PEP and HerMES surveys (Gruppioni et al. 2015). We find good agreement for SAG model galaxies over the whole SFR and z ranges considered. GALACTICUS is consistent with the measurements at SFR $\lesssim 10^{2.5}\,\mathrm{yr}^{-1}\mathrm{M}_\odot$, while SAGE agrees with the data up to $10^2\,\mathrm{yr}^{-1}\mathrm{M}_\odot$. At higher SFRs, SAGE under-predicts the number of star-forming galaxies by ~ 2 dex.

In Fig. 3, we show the evolution of the MULTIDARK-GALAXIES stellar mass function compared to, from top to bottom, the SDSS-GALEX observations at $z = 0.1$ (Moustakas et al. 2013), the PRIMUS measurements at $0.50 < z < 0.65$ (Moustakas et al. 2013), the BOSS CMASS observations at $0.5 < z < 0.6$ (Maraston et al. 2013), the DEEP2-FF data at $0.9 < z < 1.1$, and the ZFOURGE/CANDELS star-forming galaxies at $1.5 < z < 2.5$ (Tomczak et al. 2014). The BOSS CMASS mass function drops in the low-mass end
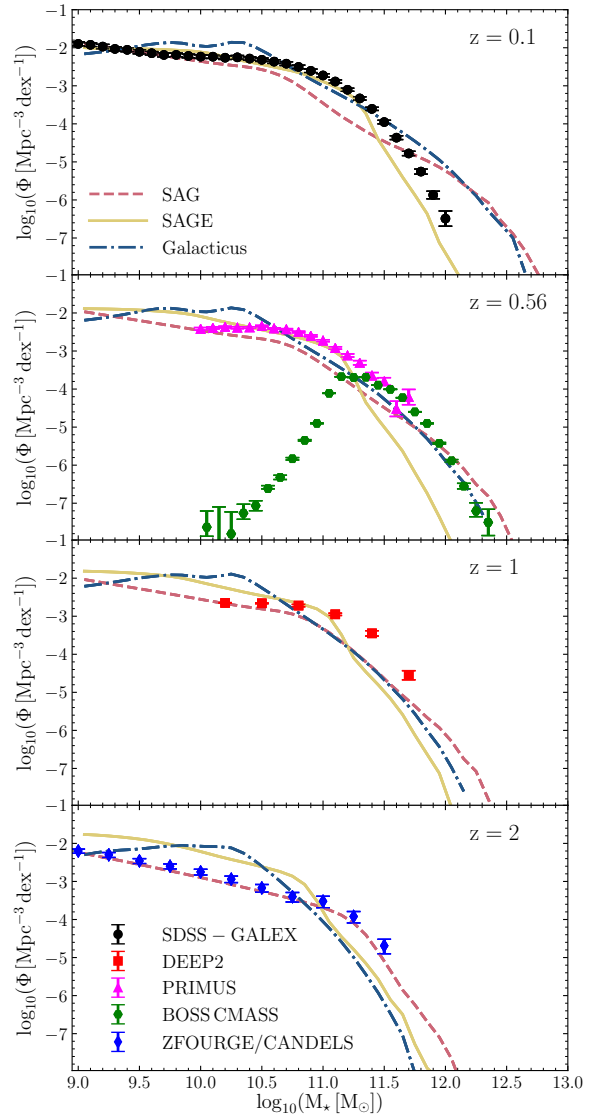
Figure 3. Stellar mass function evolution of our model galaxies (lines colour-coded as in the legend) compared to the SDSS-GALEX $z = 0.1$ (Moustakas et al. 2013, black points) observations, the PRIMUS results at $0.50 < z < 0.65$ (Moustakas et al. 2013, magenta triangles), the BOSS CMASS measurements at $0.5 < z < 0.6$ not corrected from incompleteness (Maraston et al. 2013, green hexagons), the DEEP2-FF data at $0.9 < z < 1.1$ (red squares), and the ZFOURGE/CANDELS observations at $1.5 < z < 2.5$ (Tomczak et al. 2014, blue diamonds). Note that the BOSS data drop due to the selection of luminous, red, massive galaxies for this sample.

due to the incompleteness effect generated by the CMASS colour cuts specifically designed to select luminous, red, massive galaxies (Maraston et al. 2013). Note that the stellar mass functions shown in Fig. 3 are not the same as those from Knebe et al. (2018) due to the SFR> 0 cut we apply to the SAMs. The systematic errors on DEEP2 observations at $z \sim 1$ are expected to differ from those of SDSS galaxies at lower redshifts.

It is not surprising that the agreement between SAG and ZFOURGE/CANDELS data is especially good because this model was calibrated against these observations. SAGE and GALACTICUS over-predict the number of galaxies with $\log(\mathrm{M}_\star\,[\mathrm{M}_\odot]) \lesssim 11$, and this excess is enhanced at higher

redshift (from ~ 0.1 dex at $z = 0.1$ to ~ 0.4 dex at $z = 2$). SAGE under-estimates the number of galaxies more massive than $10^{11} M_\odot$ at all redshifts. We deem the MULTIDARK-GALAXIES to be in sufficient agreement with observations in terms of their stellar-mass and SFR evolution such that we can draw meaningful predictions from the models that rely on these properties.

2.2 DEEP2 galaxies

We are interested in exploring the relationship between L[O II] and different galactic properties. For comparison, we use an observational data set, the DEEP2–FIREFLY (DEEP2-FF, hereafter) galaxy sample, which allows us to test whether the model galaxies cover similar ranges of parameters, once the adequate selection functions are implemented.

The DEEP2 survey obtained spectra of about 50,000 galaxies brighter than $R \sim 24.1$, in four separate fields covering $\sim 2.8 \, \text{deg}^2$ (Newman et al. 2013). The redshift measurement for each object in the DEEP2 DR4 database was inspected by eye and assigned an integer quality code $-2 < Q < 4$ based on the determined accuracy of the redshift value.[8] For this work, we consider galaxies with $Q > 2$, corresponding to secure redshifts, within the range $0.001 < z_{\text{best}} < 1.7$.

We adopt the DEEP2 flux-calibrated spectra generated by Comparat et al. (2016).[9] We also use the extended photometric catalogues developed by Matthews et al. (2013),[10] which supplement the DEEP2 photometric catalogues with (u, g, r, i, z) photometry from the Sloan Digital Sky Survey (SDSS). By applying the cuts specified above and taking into account the cross-match between the mentioned catalogues, the spectra of 33 838 galaxies from the original DEEP2 DR4 catalogue are used in this study. These spectra are fitted using stellar population models to extract quantities such as stellar masses, stellar metallicities, star formation rates, and ages. In particular, the DEEP2 SFR values are computed by fitting stellar population models to the spectral continuum, where the emission lines are masked for the fit. Thus, this constitutes an independent estimate from an [O II]-based SFR.

The spectral fit is performed using the FIREFLY[11] code (Wilkinson et al. 2017; Comparat et al. 2017) in which no priors, other than the assumed model described immediately below, are applied. FIREFLY treats dust attenuation in a novel way, by rectifying the continuum before the fit; for full details see Wilkinson et al. (2017) and Comparat et al. (2017). The FIREFLY fit is performed for spectral templates with ages below 20 Gyr and metallicities in the range $0.001 < Z < 3$. The maximum age found for the DEEP2-FF sample is 10.18 Gyr. It is noteworthy to remark that FIREFLY does not interpolate between the ages of the templates used in the spectral fitting. For this study, we adopt spectral templates from Maraston & Strömbäck (2011), assuming a Chabrier (2003) IMF, same as in the MULTIDARK-GALAXIES, and the ELODIE stellar library. This latter covers the wavelength range $3900 - 6800$ Å with

a 0.55 Å sampling at 5500 Å, i.e. at a resolution $R = 10,000$ (Prugniel et al. 2007).

The DEEP2 survey used the DEIMOS spectrograph at Keck, which covers approximately the wavelength range $6500 - 9300$ Å with a resolution ~ 6000 (Faber et al. 2003). The discrepancy in wavelength coverage results in a lack of fits at low redshifts for this survey.

The FIREFLY fits to the DEEP2 spectra described above are available at http://www.icg.port.ac.uk/Firefly/ (340 MB). Another fit to the DEEP2 spectra has been performed by Comparat et al. (2017) assuming slightly different age and metallicity ranges, and using a previous version of FIREFLY that did not take into account the presence of mass loss in the stellar population models. Here we refer to "stellar mass" as the sum of the mass of living stars and the mass locked in stellar remnants (i.e., white dwarfs, neutron star and black holes).

2.2.1 Broad-band absolute magnitudes

The DEEP2-FF galaxy catalogue also provides SDSS (u, g, r, i, z) apparent magnitudes. In order to compare these observations with the MULTIDARK-GALAXIES absolute magnitudes, we have $(k + e)$ corrected them (where "e" stands for evolution). To this end, we have produced an evolving set of simple stellar populations (SSP; Maraston & Strömbäck 2011) with ages, metallicities, and redshifts matching those used for the FIREFLY runs described above. In particular, we produce a table of possible evolutionary paths that provides the observed-frame properties of the given SSPs in the SDSS filters and allows us to determine the k-correction in those filters without any approximation. Hereafter, we will call it "MS table". This table calculates intrinsic magnitudes. The DEEP2 data have been corrected from interstellar dust attenuation by applying Calzetti et al. (2000) extinction law.

These SDSS observed-frame properties are computed by red-shifting the model SEDs to a fixed grid of redshifts from $z = 3.5$ down to $z = 0.$, with $\Delta z = 0.1$, and applying cosmological dimming using the Flexible-k-and-evolutionary-correction algorithm (FLAKE, Maraston, in prep.). We interpolate between the redshifts when needed. Such a technique has been widely used in the literature (e.g., Maraston et al. 2013; Etherington et al. 2017) and can be generalised to any arbitrary set of filters.

From each SSP model in the MS table above we extract the $(k + e)$ correction as:

$$(k + e)_j = M_j(z) - m_j = M_j(z) - M_j(z = 0), \quad (1)$$

where $M_j(z)$ are the galaxy SDSS $j = (u, g, r, i, z)$ absolute magnitudes at redshift z and m are the observed magnitudes, i.e. the absolute magnitudes at $z = 0$.

The FIREFLY spectral fitting code finds the best fit to a galaxy by weighting different SSPs and adding them together. It turns out that the best FIREFLY fits to the DEEP2 galaxy sample have only two SSP components. Thus, the DEEP2-FF galaxy sample can be cross-matched with the components of the MS table, by using a linear combination of the two SSP components of each FIREFLY (FF) best fit:

$$\text{SSP}^{\text{MS}} = w_0 \, \text{SSP}_0^{\text{FF}} + w_1 \, \text{SSP}_1^{\text{FF}}, \quad (2)$$

with $w_0 + w_1 = 1$. Then, each DEEP2-FF galaxy is assigned a $(k + e)$ correction that is the weighted, linear combination of the corrections from each SSP component:

$$(k + e)_j = (k + e)_j^0 \, w_0 + (k + e)_j^1 \, w_1. \quad (3)$$

2.2.2   *The DEEP2–FIREFLY galaxy sample*

For our analysis, we focus on DEEP2-FF galaxies within the redshift range $0.9 < z < 1.1$. We consider the sum of the [O II] 3727Å and 3729Å line fluxes as the [O II] doublet. Here we impose a flux limit of $F[O\,II] > 5\,\sigma_{F_{[OII]}}$ (where $\sigma_{F_{[OII]}}$ is the flux error) to guarantee robust flux estimates, and a minimum stellar mass uncertainty of $[\log_{10}(M_\star^{1\sigma\,up}) - \log_{10}(M_\star^{1\sigma\,low})]/2 < 0.4$. In the previous expression, $M_\star^{1\sigma\,up,low}$ represents the FIREFLY stellar mass within $\pm 1\,\sigma$ from the mean value of the distribution.

After applying the cuts described above, our final sample includes 4478 emitters with minimum [O II] flux of $2.45 \times 10^{-19}\,erg\,s^{-1}\,cm^{-2}$, mean $L[O\,II] \sim 10^{41.6}\,erg\,s^{-1}$, $M_\star \sim 10^{10.3}\,M_\odot$, age $\sim 10^{9.2}\,yr$, and mean cold gas metallicity $Z_{cold} \sim 0.72$. Fig. 5 shows the distribution of $L[O\,II]$ as a function of SFR. The observed sample only populates a narrow range of SFR, and this affects the comparison with the model galaxies, which have SFRs lower than the minimum value of the DEEP2-FF sample. Other properties from this data set can be seen in Fig. 8 and in Appendix C. We assume the dust attenuation of the nebular emission lines to be the same as for the continuum. Thus, we also correct the $L[O\,II]$ from interstellar dust attenuation by applying Calzetti et al. (2000) extinction law, as we have detailed above for the broad-band magnitudes.

For the analysis, we select both observed and models galaxies using a more conservative flux cut, $F[O\,II] > 5 \times 10^{-18}\,erg\,s^{-1}\,cm^{-2}$. This corresponds to $L[O\,II] \sim 10^{40.4}\,erg\,s^{-1}$ at $z = 1$ in Planck cosmology (Planck Collaboration et al. 2015), and roughly mimics the observational limitations (see also Gonzalez-Perez et al. 2018). This cut reduces the sparse, faint tail of the observed distribution (there are only 4 DEEP2 galaxies with flux lower than $5 \times 10^{-18}\,erg\,s^{-1}\,cm^{-2}$) and allows us to obtain much narrower SAM constraints.

As shown in Fig. 4, most galaxies with $L[O\,II] < 10^{40.5}\,erg\,s^{-1}$ have been removed from the DEEP2-FF sample, compared to the original DEEP2 population. Despite this, the DEEP2-FF galaxy sample is statistically representative of the original DEEP2 population. In fact, the cumulative distribution functions of these two samples, approximated by splines, differ by less than 5%, according to a Kolmogorov–Smirnov test.

Fig. 4 shows the distribution of the dust-attenuated $L[O\,II]$ computed with GET_EMLINES (see §3.1) from instantaneous and average SFRs, for SAG model galaxies at $z = 0.94$ and with $M_\star > 10^{8.87}\,M_\odot$ and SFR$> 0\,yr^{-1}\,M_\odot$ (see Sec. 2.1.4). These model $L[O\,II]$ distributions are statistically different from the DEEP2-FF one. However, they have similar mean values: $\langle L[O\,II]\rangle \sim 10^{41}\,erg\,s^{-1}$, $\langle age\rangle \sim 10^{9.2}\,yr$, $\langle M_\star\rangle \sim 10^{9.47}\,M_\odot$.

In order to draw a sample of model galaxies consistent with DEEP2-FF observations, we select SAG galaxies with a $L[O\,II]$ distribution following the spline fit to the DEEP2-FF distribution, as shown in Fig. 4. We perform such a drawing for SAG [O II] luminosities computed both from instantaneous and average SFR. The $L[O\,II]$ of these new selections have mean values consistent with those from the DEEP2-FF sample. Meanwhile, the ages, $\langle age\rangle \sim 10^9\,yr$, and the stellar masses, $\langle M_\star\rangle \sim 10^{9.86}\,M_\odot$, are lower than the observed ones.

In Appendix A, the DEEP2-FF sample is directly compared to the SAG model galaxies selected following the DEEP2-FF $L[O\,II]$ distribution. These SAG model subsets have brighter $L[O\,II]$, M_u and M_g values, slightly lower
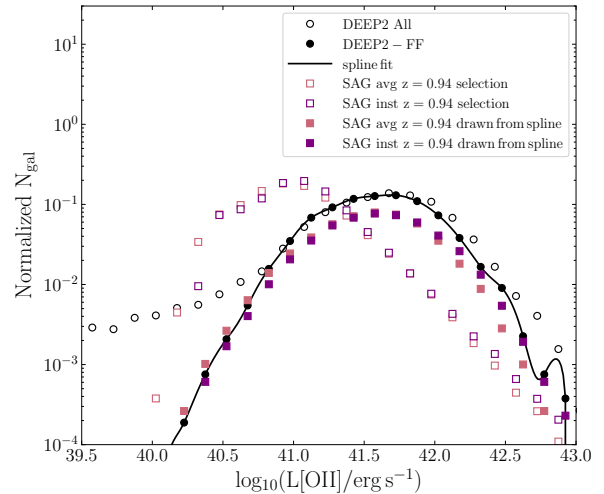
Figure 4. Attenuated $L[O\,II]$ distribution of the original DEEP2 sample (black, empty dots) compared to the DEEP2-FF selection (black, filled dots), which we fit with a spline function. The area under the curves is normalized to unity. We compare these results with the SAG model galaxies selected with $M_\star > 10^{8.87}\,M_\odot$ and SFR$> 0\,yr^{-1}\,M_\odot$ (empty squares; see Sec. 2.1.4), and with the SAG galaxies randomly drawn from the DEEP2-FF spline distribution (filled squares). The [O II] luminosity values in the model galaxies are calculated using the GET_EMLINES code, inputing either the instantaneous (purple) or average (salmon) SFR. All the details about these quantities and the calculations are given in Sec. 3.1.

ages, higher stellar masses and span higher SFR values compared to the SAG selection at $M_\star > 10^{8.87}\,M_\odot$ and SFR$> 0\,yr^{-1}\,M_\odot$.

The main focus of this paper is to test the validity of different approaches for modelling emission lines in large galaxy samples with volumes comparable to the observable Universe. In this context, the comparison to the DEEP2-FF sample is meant to be a rough guide to the expected location of observed galaxies in different parameter spaces.

3   [O II] EMITTERS IN THE SAMS

The physics of [O II] emission lines is difficult to model, as it depends on local processes, such as dust extinction, and the inner structure and the ionising fields of the H II nebula in which they are embedded. Different approaches have been used to model the [O II] emission line: (i) assume a relation between $L[O\,II]$ and SFR and, possibly, metallicity as it happens in observations (Kennicutt 1998; Kewley et al. 2004; Moustakas et al. 2006; Jouvel et al. 2009; Sobral et al. 2012; Talia et al. 2015; Valentino et al. 2017); (ii) assume an average H II region for a range of metallicities (Gonzalez-Perez et al. 2018); (iii) couple a photoionisation model with a galaxy evolution one (Hirschmann et al. 2012; Orsi et al. 2014). We address method (i) in Section 4 and method (iii) here.

None of the MULTIDARK-GALAXIES catalogues studied in this work provides direct $L[O\,II]$ estimates. Therefore, we couple the SAMs with the GET_EMLINES model (Orsi et al. 2014), which encapsulates the results from the MAPPINGS-III photoionisation code (Groves et al. 2004; Allen et al. 2008). Here, the ionisation parameter of gas in galaxies is directly related to their cold gas metallicity, obtaining a

reasonable agreement with the observed Hα, [O II] λ3727, [O III] λ5007 luminosity functions, and the Baldwin, Phillips & Terlevich (BPT; Baldwin et al. 1981) diagram for local star-forming galaxies. Ideally, the GET_EMLINES methodology requires as input the cold gas metallicity and the instantaneous SFR. This latter quantity, however, is not usually output by SAMs. The instantaneous SFR is preferred to a time-averaged equivalent, as the latter can include contributions from stellar populations older than those responsible for generating the nebular emission in star-forming galaxies.

SAG is the only semi-analytic model providing both instantaneous and average SFR values, while SAGE and GALACTICUS only provide average SFRs. In the next section, we describe in detail the GET_EMLINES algorithm to be used in the L[O II] calculation for a semi-analytic model. Because SAMs do not usually output the instantaneous SFR, which is needed as default input for the GET_EMLINES code, we test the usage of the average SFR and how this affects different galactic properties.

3.1 The GET_EMLINES code

We now describe step by step how we have implemented the GET_EMLINES nebular emission code to obtain [O II] luminosities for the MULTIDARK-GALAXIES. This methodology is based on the photoionisation code MAPPINGS-III (Groves et al. 2004; Allen et al. 2008), which relates the ionisation parameter of gas in galaxies, q, to their cold gas metallicity $Z_{\rm cold}$ as:

$$q(Z) = q_0 \left(\frac{Z_{\rm cold}}{Z_0} \right)^{-\gamma}, \quad (4)$$

where q_0 is the ionisation parameter of a galaxy that has cold gas metallicity Z_0 and γ is the exponent of the power law. Following Orsi et al. (2014), from the pre-computed H II region model grid of Levesque et al. (2010), we assume $q_0 = 2.8 \times 10^7 \, {\rm cm \, s^{-1}}$, $Z_0 = 0.012$ and $\gamma = 1.3$ for all the analysed galaxy models. This specific combination of values was presented in Orsi et al. (2014), and it has ionization parameter values that bracket the range spanned by the MAPPINGS-III grid for the bulk of the galaxy population studied in that work. The q_0 and γ parameters above were found to produce model Hα, [O II] λ3727 (to indicate the doublet), [O III] λ5007 luminosity functions and a model BTP (Baldwin et al. 1981) diagram for local star-forming galaxies in good agreement with observations.

The GET_EMLINES code has been calibrated to reproduce a range of luminosity functions at different redshifts and local line ratios diagrams, and it has been tested against observations up to $z = 5$ (Orsi et al. 2014). A different combination of q_0 and γ changes the L[O II] results in a complicated way. For instance, higher parameter values produce a lower number density of bright emitters, which translates into a substantial difference in the lower peak of the L[O II]-SFR bimodality shown in Fig. 5. Changing the q_0 and γ parameters would require to recalibrate the GET_EMLINES model, and this goes beyond the scope of this work.

The cold gas metallicity is defined as the ratio between the cold gas mass in metals and the cold gas mass (e.g., Yates 2014), considering both bulge and disc components, when available:

$$Z_{\rm cold} = \frac{M_{Z{\rm cold}}}{M_{\rm cold}}. \quad (5)$$

Another fundamental quantity needed to derive the

[O II] line luminosity is the hydrogen ionising photon rate defined as:

$$Q_{\rm H^0} = \int_0^{\lambda_0} \frac{\lambda L_\lambda}{hc} d\lambda, \quad (6)$$

where L_λ is the galaxy composite SED in $\rm erg \, s^{-1} \, \AA^{-1}$, $\lambda_0 = 912\AA$, c is the speed of light and h is the Planck constant. $Q_{\rm H^0}$ is a unit-less quantity calculated at each model snapshot just by solving the integral above. Assuming a Kroupa (2001) IMF, one can express the ionising photon rate as a function of the instantaneous star formation rate as (Falcón-Barroso & Knapen 2013):

$$Q_{\rm H^0} = \log_{10} 1.35 + \log_{10}({\rm SFR/M_\odot \, yr^{-1}}) + 53.0. \quad (7)$$

Combining Eq. 7 with the attenuation-corrected emission-line lists from Levesque et al. (2010), normalised to the Hα line flux, we compute the [O II] luminosity as:

$$L(\lambda_j) = 1.37 \times 10^{-12} Q_{\rm H^0} \frac{F(\lambda_j, q, Z_{\rm cold})}{F(H\alpha, q, Z_{\rm cold})}, \quad (8)$$

where $F(\lambda_j, q, Z_{\rm cold})$ is the MAPPINGS-III prediction of the desired emission line flux at wavelength λ_j for a given set of $(q, Z_{\rm cold})$ and $F(H\alpha, q, Z_{\rm cold})$ is the Hα normalisation flux.

The total luminosity of the [O II] doublet is the sum of the luminosities of the two lines at $\lambda_j = 3727, 3729 \, \AA$, both calculated using Eq. 8.

The [O II] luminosity in Eq. 8 does not include any dust contribution. In order to account for dust attenuation, we implement the correction detailed in next Section using Cardelli et al. (1989) extinction curve.

3.2 Dust extinction

In this study, the intrinsic [O II] luminosity given in Eq. 8, $L(\lambda_j)$, is attenuated by interstellar dust as follows:

$$L(\lambda_j)^{\rm att} = L(\lambda_j) 10^{-0.4 A_\lambda(\tau_\lambda^z, \theta)}, \quad (9)$$

where $A_\lambda(\tau_\lambda^z, \theta)$ represents the attenuation coefficient defined as a function of the galaxy optical depth τ_λ^z and the dust scattering angle θ. Explicitly we have (Spitzer 1978; Osterbrock 1989; Draine 2003; Izquierdo-Villalba et al. 2019):

$$A_\lambda(\tau_\lambda^z, \theta) = -2.5 \log_{10} \frac{1 - \exp(-a_\lambda \sec \theta)}{a_\lambda \sec \theta}, \quad (10)$$

where $a_\lambda = \sqrt{1 - \omega_\lambda} \tau_\lambda^z$ and ω_λ is the dust albedo, i.e. the fraction of the extinction that is scattering. We assume $\cos \theta = 0.60$ and $\omega_\lambda = 0.80$, meaning that the scattering is not isotropic but more forward-oriented, and that 80% of the extinction is scattering. These are the values that return the best agreement with DEEP2+VVDS observations in Fig. 9.

The galaxy optical depth τ_λ^z that enters Eq. 10 is defined as (Devriendt et al. 1999; Hatton et al. 2003; De Lucia & Blaizot 2007):

$$\tau_\lambda^z = \left(\frac{A_\lambda}{A_V} \right)_{Z_\odot} \left(\frac{Z_{\rm cold}}{Z_\odot} \right)^s \left(\frac{\langle N_H \rangle}{2.1 \times 10^{21} {\rm atoms \, cm^{-2}}} \right), \quad (11)$$

where the first two factors on the right-hand side represent the extinction curve. This depends on the cold gas metallicity $Z_{\rm cold}$ defined in Eq. 5 according to power-law interpolations based on the solar neighbourhood, the Small and the Large Magellanic Clouds. The exponent $s = 1.6$ (Guiderdoni & Rocca-Volmerange 1987) holds for the $\lambda > 2000\AA$ regime,

where the [O II] line is located. The $(A_\lambda/A_V)_{Z_\odot}$ term is the extinction curve for solar metallicity, which we take to be that of the Milky Way, and $\langle N_H \rangle$ the mean hydrogen column density. We adopt the values $Z_\odot = 0.0134$ (Asplund et al. 2009) for the solar metallicity.

Assuming the Cardelli et al. (1989) extinction law in $0.3\,\mu\text{m} \leqslant \lambda < 0.9\,\mu\text{m}$ (i.e., optical/NIR regime), one has:

$$\left(\frac{A_\lambda}{A_V}\right) = a(x) + b(x)/R_V, \qquad (12)$$

where $x \equiv \lambda^{-1}$, $R_V \equiv A_V/E(B-V) = 3.1$ is the ratio of total to selective extinction for the diffuse interstellar medium in the Milky Way, and

$$a(x) = 1 + 0.17699\,y - 0.50447\,y^2 - 0.02427\,y^3 +$$
$$0.72085\,y^4 + 0.01979\,y^5 - 0.77530\,y^6 + 0.32999\,y^7,$$
$$b(x) = 1.41338\,y + 2.28305\,y^2 + 1.07233\,y^3 - 5.38434\,y^4$$
$$- 0.62251\,y^5 + 5.30260\,y^6 - 2.09002\,y^7,$$
$$(13)$$

with $y = (x - 1.82)$.

The mean hydrogen column density is given by (Hatton et al. 2003; De Lucia & Blaizot 2007):

$$\langle N_H \rangle = \frac{M_{\text{cold}}^{\text{disc}}}{1.4\,m_p\,\pi\,(a\,R_{1/2}^{\text{disc}})^2} \text{ atoms cm}^{-2}, \qquad (14)$$

where $M_{\text{cold}}^{\text{disc}}$ is the cold gas mass of the disc, $m_p = 1.67 \times 10^{-27}$ kg is the proton mass, $a = 1.68$ is such that the column density represents the mass-weighted mean column density of the disc, and $R_{1/2}^{\text{disc}}$ is the disc half-mass radius.

Qualitatively for this dust attenuation model[12], galaxies with large amounts of cold gas, metal rich cold gas and/or small scale sizes, will be the most attenuated ones (see also Merson et al. 2016).

3.3 Instantaneous versus average SFR

The GET_EMLINES code described in Section 3.1 ideally requires as inputs the instantaneous SFR and cold gas metallicity of galaxies. The instantaneous SFR, which is defined on a smaller time-step compared to the average SFR (see Sec. 2.1.4), traces very recent or ongoing episodes of star-formation, that are the relevant ones for nebular emission.

Fig. 5 shows, as a function of SFR, the intrinsic (i.e. corrected from dust attenuation) $L[\text{O II}]$ that the coupling with GET_EMLINES gives for both the instantaneous (solid contours) and average (dashed) SFR from SAG at $z \sim 1$. The innermost (outermost) contours enclose 68% (95%) of our model galaxies. The diagonal lines show the correlations between SFR and $L[\text{O II}]$. These are tight correlations, whose best-fitting parameters are reported in Table 1. Under laid are the DEEP2-FF observational data at $0.9 < z < 1.1$. Overall, the model galaxy distributions presented in Fig. 5 are very similar for the $L[\text{O II}]$ derived from either the instantaneous or the average SFRs. These distributions show a bimodality that can also be seen in the observations.

The instantaneous and average SFR derived distributions differ the most at SFR$\lesssim 10^0\,\text{yr}^{-1}\text{M}_\odot$, with [O II] luminosities from average SFR being ~ 0.2 dex fainter than those from instantaneous SFR. At SFR$\sim 10^{1.5}\text{yr}^{-1}\text{M}_\odot$, there are slightly less bright [O II] emitters from instantaneous SFR.

[12] Our implementation of the dust attenuation model is available at https://github.com/gfavole/dust
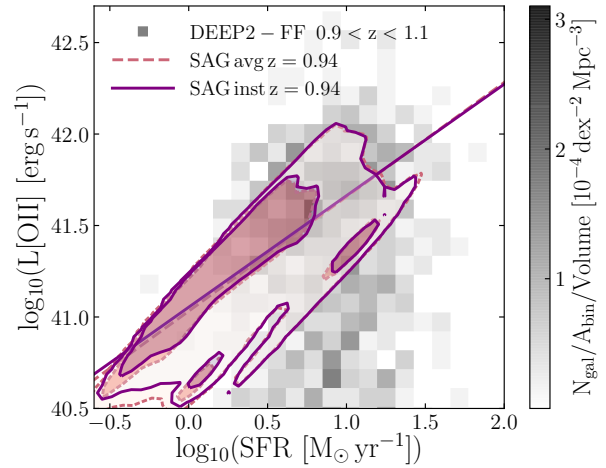
Figure 5. Intrinsic [O II] luminosity as a function of the SFR for the SAG model galaxies at $z \sim 1$ (contours) and the DEEP2-FF observations at $0.9 < z < 1.1$ (grey, shaded squares). The bar represents the number density of DEEP2-FF galaxies in each 2D bin normalised by the bin area in units of $[\text{dex}^{-2}\,\text{Mpc}^{-3}]$. We have imposed a minimum [O II] flux of $5\times10^{-18}\text{erg s}^{-1}\,\text{cm}^{-2}$ to both observations and models. The model $L[\text{O II}]$ values are calculated by assuming instantaneous (solid, purple contours) and average (dashed, salmon) SFR as input for the GET_EMLINES prescription. The innermost (outermost) model contours encompass 68% (95%) percent of the galaxy distributions. The diagonal lines represent the $L[\text{O II}]$-SFR correlations, whose coefficients are given in Table 1.

DEEP2-FF galaxies in the upper density peak of the observed bimodal distribution shown in Fig. 5 are older, more massive, more luminous and slightly more star-forming (mean values: $\langle\text{age}\rangle \sim 10^{9.28}\text{yr}$, $\langle M_\star \rangle \sim 10^{10.42}\text{M}_\odot$, $\langle L[\text{O II}] \rangle \sim 10^{41.48}\text{erg s}^{-1}$, $\langle\text{SFR}\rangle \sim 10^{1.13}\text{yr}^{-1}\text{M}_\odot$) compared to their counterparts in the lower density area ($\sim 10^{9.13}\text{yr}$, $\sim 10^{10.22}\text{M}_\odot$, $\sim 10^{39.07}\text{erg s}^{-1}$, $\sim 10^{1.08}\text{yr}^{-1}\text{M}_\odot$). Overall, we find an opposite trend for model galaxies. In fact, the upper peak of the bimodality is composed of younger, less massive, slightly more luminous, less star-forming galaxies with mean values: $\langle\text{age}\rangle \sim 10^{9.16}\text{yr}$, $\langle M_\star \rangle \sim 10^{9.58}\text{M}_\odot$, $\langle L[\text{O II}] \rangle \sim 10^{41.26}\text{erg s}^{-1}$, $\langle\text{SFR}\rangle \sim 10^{0.21}\text{yr}^{-1}\text{M}_\odot$); the lower peak has mean values: $\langle\text{age}\rangle \sim 10^{9.32}\text{yr}$, $\langle M_\star \rangle \sim 10^{10.06}\text{M}_\odot$, $\langle L[\text{O II}] \rangle \sim 10^{41.25}\text{erg s}^{-1}$, $\langle\text{SFR}\rangle \sim 10^{0.83}\text{yr}^{-1}\text{M}_\odot$.

At the end of this Section, we will discuss further the origin of the DEEP2-FF $L[\text{O II}]$-SFR bimodal trend in connection with other galactic properties shown in Fig. 8.

In the top panel of Fig. 6 we compare the average (dashed, salmon) and instantaneous (solid, purple) SAG SFR functions at $z \sim 1$, whose ratio is displayed in the bottom panel. The instantaneous and average SFR functions remain within 5% of each other at SFR$> 10^0\,\text{yr}^{-1}\text{M}_\odot$ (the 5% region is highlighted by the yellow shade). There is a slightly larger fraction, within 20%, of SAG galaxies having low average SFR, SFR$< 10^0\,\text{yr}^{-1}\text{M}_\odot$, than instantaneous values. The main difference between average and instantaneous SFRs is found for galaxies with the highest specific SFR (i.e., SFR/$M_\star$) and stellar masses below $10^{11}\text{M}_\odot$.
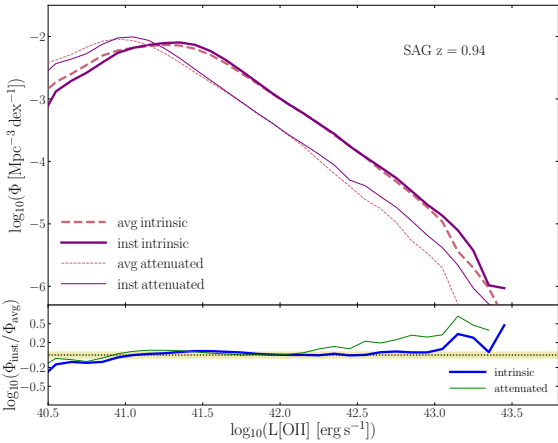
The top panel in Fig. 7 presents the intrinsic (thick lines) and attenuated (thin) [O II] luminosity functions derived from the average SFR (dashed, salmon line) and instantaneous SFR (solid, purple) from SAG. We impose on the SAG model galaxies the same [O II] flux limit of DEEP2-FF observations, $5 \times 10^{-18}\text{erg s}^{-1}\,\text{cm}^{-2}$ (see Sec. 2.2.2),
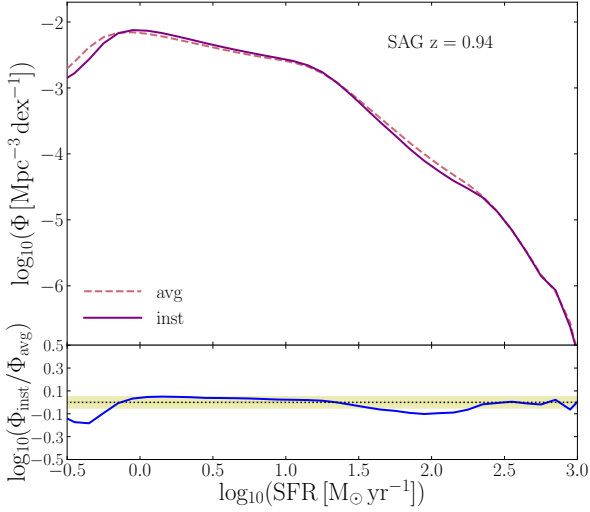
Figure 6. Average (dashed, salmon) versus instantaneous (solid, purple) SFR functions for SAG model galaxies. The bottom panel shows the ratio between the two, and the yellow, shaded region highlights the 5% region of agreement.

Figure 7. Intrinsic (thick lines) and attenuated (thin) [O II] luminosity functions based on SAG average (dashed, salmon) and instantaneous SFR (solid, purple). The bottom panel shows the ratios between the two and the yellow stripe highlights the 5% region of agreement. We apply the mocks the same [O II] flux limit of DEEP2-FF observations, 5×10^{-18} erg s^{-1} cm^{-2} (see Sec. 2.2.2).

which corresponds to L[O II] $\sim 10^{40.4}$ erg s^{-1} at $z = 1$ in Planck cosmology (Planck Collaboration et al. 2015). The instantaneous-to-average amplitude ratios are displayed in the bottom panel of Fig. 7. The intrinsic (attenuated) L[O II] functions have differences below 5% for luminosities in the range $10^{41} - 10^{43}$ erg s^{-1} ($10^{41} - 10^{42.2}$ erg s^{-1}), which are highlighted by the yellow shade. At lower (higher) luminosities, the discrepancies grow up to 20% (30%). For the brightest galaxies, the discrepancy remains within 50%. The difference produced in L[O II] by assuming average instead of instantaneous SFR does not change significantly with redshift over the range $0.6 < z < 1.2$ (see Appendix B for further details). Thus, the average and instantaneous SFR can be assumed interchangeably for average galaxies.

In Fig. 8, from top to bottom, we display the SAG broad-
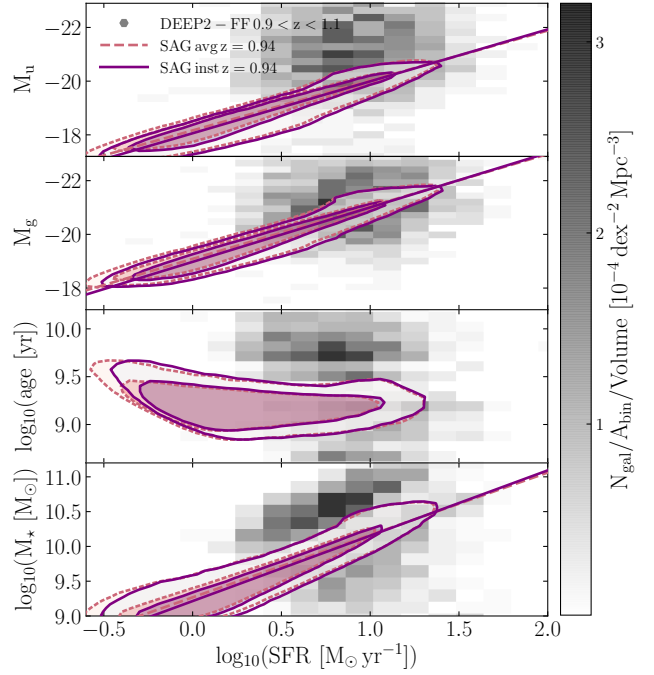
Figure 8. From top to bottom: intrinsic magnitudes, ages and stellar masses as a function of star formation rate for SAG (contours) at $z \sim 1$ and DEEP2-FF observations at $0.9 < z < 1.1$ (grey, shaded squares). The colour bar shows the number density of DEEP2 galaxies per bin area in units of [dex^{-2} Mpc^{-3}]. The dashed, salmon (solid, purple) contours represent the average (instantaneous) SFRs. The innermost (outermost) contours encompass 68% (95%) of the distributions. The diagonal lines are the linear fits showing the significant correlations (i.e. $r \geqslant 0.6$), whose coefficients are reported in Table 2, together with the best-fit parameters.

y=$Ax+B$	A	B	σ_y	r
y=$\log_{10}(L$[O II]$)$				
x=$\log_{10}(\mathrm{SFR_{avg}})$	0.625±0.001	41.03±0.01	0.40	0.83
x=$\log_{10}(\mathrm{SFR_{inst}})$	0.609±0.001	41.05±0.01	0.38	0.80
y=M_u				
x=$\log_{10}(\mathrm{SFR_{avg}})$	-1.859±0.001	-18.17±0.01	1.07	0.92
x=$\log_{10}(\mathrm{SFR_{inst}})$	-1.934±0.001	-18.06±0.01	1.07	0.90
y=M_g				
x=$\log_{10}(\mathrm{SFR_{avg}})$	-1.951±0.001	-19.09±0.01	1.11	0.93
x=$\log_{10}(\mathrm{SFR_{inst}})$	-2.029±0.001	-18.98±0.01	1.11	0.91
y=$\log_{10}(\mathrm{M_\star})$				
x=$\log_{10}(\mathrm{SFR_{avg}})$	0.897±0.001	9.27±0.01	0.54	0.89
x=$\log_{10}(\mathrm{SFR_{inst}})$	0.939±0.001	9.21±0.01	0.54	0.87

Table 1. Best-fit parameters of the linear scaling relations found for SAG model galaxies at $z = 1$ and shown in Fig. 8. The parameter r is the correlation coefficient and σ_y is the scatter in the y-axis. All the L[O II] values are intrinsic.

band u and g absolute magnitudes, ages and stellar masses as a function of the average SFR (dashed, salmon contour) and instantaneous SFR (solid, purple). We compare them with the DEEP2-FF observations at $0.9 < z < 11$ (grey, shaded squares). Except for the age, all these properties are tightly correlated with both SFRs. The lack of correlation between age and SFRs is clear for both the model and DEEP2-FF galaxies. We fit straight lines to the instan-

taneous and average contours and report the best-fit parameters and correlation coefficients in Table 1. For the broadband magnitudes, the slopes of the average SFR correlations are only ~ 0.07 shallower than the instantaneous ones; for the stellar mass they are even closer. Overall, the width of the distributions as a function of both SFRs does not vary significantly. The average SFR contours extend down to slightly smaller values compared to the instantaneous contours.

The DEEP2-FF age and stellar mass distributions as a function of SFR in Fig. 8 show a bimodal trend, with an upper population of older, more massive, luminous, quiescent galaxies ($\langle age \rangle \sim 10^{9.8}$ yr, $\langle M_\star \rangle \sim 10^{10.82}$ M$_\odot$, $\langle L[\mathrm{O\,II}] \rangle \sim 10^{40.57}$ erg s^{-1}, $\langle \mathrm{SFR} \rangle \sim 10^{1.02}$ yr^{-1}M$_\odot$) and a lower tail of younger, less massive, luminous, more star-forming emitters ($\langle age \rangle \sim 10^{8.48}$ yr, $\langle M_\star \rangle \sim 10^{9.72}$ M$_\odot$, $\langle L[\mathrm{O\,II}] \rangle \sim 10^{40.45}$ erg s^{-1}, $\langle \mathrm{SFR} \rangle \sim 10^{1.24}$ yr^{-1}M$_\odot$). We obtain the same mean galaxy properties splitting the DEEP2-FF sample with a cut in either the age-SFR or Mstar-SFR planes.

The mean DEEP2-FF values derived from splitting in age or stellar mass as a function of SFR are similar to those obtained by splitting in $L[\mathrm{O\,II}]$ versus SFR (see Fig. 5). The age/mass-SFR bimodal trend observed in DEEP2-FF galaxies is not reproduced by the SAG model galaxies, which instead look bimodal in the $L[\mathrm{O\,II}]$-SFR plane (Fig. 5) because of the non-trivial dependence of $L[\mathrm{O\,II}]$ on metallicity through the parameters q_0 and γ (see Sec. 3.1).

In this section, we have shown that using the SAG average SFRs as input for the GET_EMLINES code gives results within 5% from using the instantaneous value for galaxies with attenuated $L[\mathrm{O\,II}]$ in the range $10^{40.9} - 10^{42.2}$ erg s^{-1}, and with intrinsic $L[\mathrm{O\,II}]$ between $10^{40.9} - 10^{43}$ erg s^{-1}. These are the ELGs with SFR within $10^{-0.2} - 10^{1.6}$ yr^{-1} M$_\odot$. At higher and lower SFRs, there is a larger discrepancy between the average and instantaneous values, which translates into a larger difference ($< 60\%$) in the number of bright [O II] emitters. Thus, this effect is not significant for the average galaxy population.

3.4   Model [OII] luminosity functions

In the top panel of Fig. 9, we present the MULTIDARK-GALAXIES dust attenuated [O II] luminosity functions at $z = 0.94$ compared to a compilation of DEEP2 and VVDS data from Comparat et al. (2016). Note that similar results have been found within the redshift range $0.6 < z < 1.2$, although they are not shown here. The SAM [O II] luminosities have been derived using the GET_EMLINES code described above coupled with instantaneous SFR for SAG model galaxies, and average SFR for SAGE and Galacticus, for which the instantaneous quantity is not available. The dust attenuation has been accounted for by correcting these luminosities applying Eq. 9 with Cardelli et al. (1989) extinction curve. There are varying degrees of agreement between the models and observational data across the $\sim$3 decades in [O II] luminosity and redshift range considered. Nevertheless, the trends from all the data sources are consistent. This plot highlights that the shape and normalisation of a predicted [O II] luminosity function from a SAM are robust to both the precise prescriptions that govern galaxy evolution in the model, and the calculation of [O II] from either instantaneous or average SFR.

In the top panel of Fig. 9, we see a drop in the number of Galacticus [O II] emitters at intermediate luminosities that
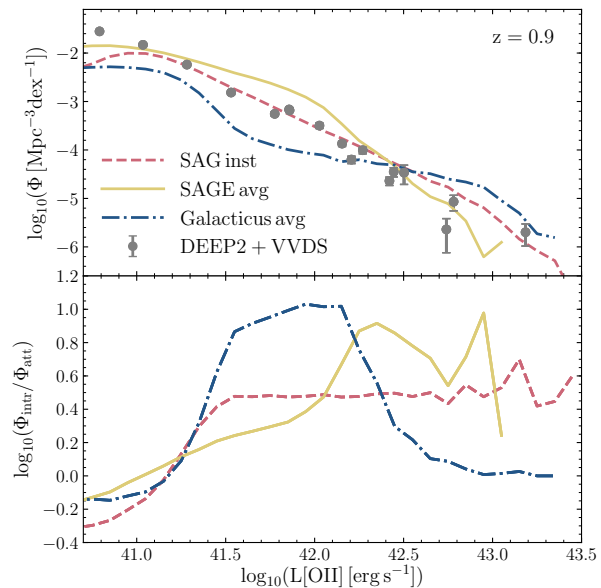
Figure 9. *Top:* Dust attenuated [O II] luminosity functions of the MULTIDARK-GALAXIES at $z \sim 1$ compared with DEEP2+VVDS observations (Comparat et al. 2016). We consider all SAM galaxies above 5×10^{-18} erg s^{-1} cm^{-2}. All the [O II] luminosities are computed using the GET_EMLINES code with SFR and cold gas metallicity as inputs (see Section 3.1). The SAG $L[\mathrm{O\,II}]$, which are estimated using the instantaneous SFR, are in good agreement with the SAGE and Galacticus results based on the average SFR. *Bottom:* Ratios between the model intrinsic [O II] luminosity functions (given by Eq. 8) and the dust attenuated ones (see Sec. 3.2) shown in the upper panel.

is independent of the stellar mass. This is mainly determined by the half mass radii of the disc, $\mathrm{R}_{1/2}^{disc}$, that enter the dust attenuation correction (see Eq. 14). These radii in Galacticus are about 50% smaller than in SAG and SAGE. At $L[\mathrm{O\,II}] \gtrsim 10^{42.5}$ erg s^{-1} and $z < 1.2$, Galacticus predicts about 0.5 dex more [O II] emitters than the other two models.

In the bottom panel of Fig. 9, we display the ratios of the attenuated-to-intrinsic $L[\mathrm{O\,II}]$ functions. As expected, the largest effect of attenuation occurs at $L[\mathrm{O\,II}] \gtrsim 10^{42}$ erg s^{-1}, where more massive galaxies are located, while in the low-luminosity, low-mass regime, the observed and intrinsic signals tend to overlap. The ratio between the intrinsic and attenuated luminosity functions is model dependent. In particular, the largest variations are due to the dust model, which depends on the metallicity, gas content and size of each galaxy, as described in §3.2. For SAGE model galaxies, this ratio increases for brighter $L[\mathrm{O\,II}]$ galaxies. For SAG, the ratio also increases up to $L[\mathrm{O\,II}] \gtrsim 10^{41.5}$ erg s^{-1}, although with a steeper slope, and beyond this value it reaches a plateau. The ratio for GALACTICUS has a prominent bump in the luminosity range $10^{41.5} - 10^{42.5}$ erg s^{-1}, where the effect of attenuation is more pronounced, and this feature corresponds to the drop seen at intermediate $L[\mathrm{O\,II}]$ in the upper panel. At higher luminosities, there is almost no difference between the intrinsic and dust attenuated GALACTICUS $L[\mathrm{O\,II}]$ functions.

4   [OII] LUMINOSITY PROXIES

Observational studies have shown tight correlations between the [O II] luminosity, SFR (Kennicutt 1998; Sobral et al.

2012; Kewley et al. 2004; Moustakas et al. 2006; Comparat et al. 2015) and the galaxy UV-emission (Comparat et al. 2015), *without* the need to introduce any dependence on metallicity (Moustakas et al. 2006). This has prompted authors of theoretical papers to treat star-forming galaxies as ELGs when making predictions for upcoming surveys (e.g. Orsi & Angulo 2018; Jiménez et al. 2019).

Here we explore the possibility of using simple, linear relations to infer the [O II] luminosity from global galaxy properties that are commonly output in SAMs. For this purpose, we investigate both observationally motivated prescriptions (Section 4.1), and we derive model relations from the GET_EMLINES code coupled with the SAMs considered (Sections 4.2 and 4.3). For this last study, we quantify the correlation between the model L[O II] from GET_EMLINES with the average SFR, broad-band magnitudes, stellar masses, ages and cold gas metallicities. Directly using the measured L[O II]-SFR linear relation is useful to understand when is adequate to consider ELGs equivalent to star-forming galaxies and when it is not.

We find that the stellar mass of the MULTIDARK-GALAXIES are unaffected by the change in proxies for estimating their [O II] luminosities. As a consequence, the stellar-to-halo mass relation (SHMR) is also unchanged using different L[O II] proxies.

We remind the reader that, unless otherwise specified, we exclusively select emission line galaxies with fluxes above $5 \times 10^{-18}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ in both the DEEP2-FF observations and MULTIDARK-GALAXIES. This flux limit corresponds to a L[O II] $> 10^{40.4}\,\mathrm{erg\,s^{-1}}$ at $z = 1$ in the Planck cosmology (Planck Collaboration et al. 2015). All the results in what follows have these minimum cuts applied.

4.1 The SFR–L[OII] relation

In this Section, we derive **intrinsic** L[O II] from the average SFR of the MULTIDARK-GALAXIES using three different, published relations assuming a Kennicutt (1998) IMF. These are: the Moustakas et al. (2006) conversion (see also Comparat et al. 2015) calibrated at $z = 0.1$,

$$L_{[OII]}^{Moust}(\mathrm{erg\,s^{-1}}) = \frac{\mathrm{SFR(M_\odot\,yr^{-1})}}{2.18 \times 10^{-41}}, \quad (15)$$

the Sobral et al. (2012) formulation optimised at $z = 1.47$,

$$L_{[OII]}^{Sob}(\mathrm{erg\,s^{-1}}) = \frac{\mathrm{SFR(M_\odot\,yr^{-1})}}{1.4 \times 10^{-41}}, \quad (16)$$

the Kewley et al. (2004) conversion calibrated at $z = 1$,

$$L_{[OII]}^{Kew}(\mathrm{erg\,s^{-1}}) = \frac{\mathrm{SFR(M_\odot\,yr^{-1})}}{7.9 \times 10^{-42}} \\ \times (a[12 + \log_{10}(O/H)_{cold}] + b). \quad (17)$$

The coefficients (a, b) in the equation above are the values from Kewley et al. (2004) derived for the R_{23} metallicity diagnostic (Pagel et al. 1979). The $[12+\log_{10}(O/H)_{cold}]$ term is the [O II] ELG gas-phase oxygen abundance, which we proxy with the cold gas-phase metallicity Z_{cold} given in Eq. 5 through the solar abundance and metallicity. Explicitly we have:

$$12 + \log_{10}(O/H)_{cold} = [12 + \log_{10}(O/H)_\odot]\frac{Z_{cold}}{Z_\odot}, \quad (18)$$

where we assume $Z_\odot = 0.0134$ (Asplund et al. 2009), and $[12 + \log_{10}(O/H)_\odot] = 8.69$ (Allende Prieto et al. 2001). As
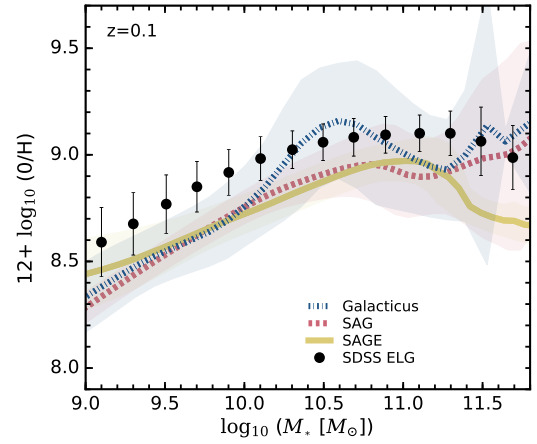
Figure 10. Mean gas-phase oxygen abundance in bins of stellar mass of the SDSS emission line galaxies at $z \sim 0.1$ (Favole et al. 2017) compared to the MULTIDARK-GALAXIES models. The abundance is computed for the SAMs using Eq. 18. The error bars on the SDSS measurements are the 1σ scatter around the mean.

the above relations are for intrinsic luminosities, dust attenuated quantities are obtained following the description in § 3.2.

For SAG and GALACTICUS, galaxies' cold gas is broken into bulge and disc components (see their respective papers for their definitions of a 'gas bulge'); we therefore take a mass-weighted average of these components' metallicities to obtain Z_{cold}. SAGE instead always treats cold gas as being in a disc. In addition, the SAG catalogues also output the $(O/H)_{cold}$ values, which are mass density ratios, that we use in the calculation of Eq. 17 for SAG model galaxies. In order to derive the correct abundances in terms of number densities, we need to rescale them by the Oxygen-to-Hydrogen atomic weight ratio, $A_O/A_H \sim 15.87$.

Fig. 10 displays the comparison between the gas-phase oxygen abundances of our SAM galaxies computed using Eq. 18 and the observed abundance of the SDSS [O II] ELGs at $z \sim 0.1$ from Favole et al. (2017). The SDSS metallicity values have been derived from the MPA-JHU DR7[13] catalogue of spectrum measurements and are built according to the works of Tremonti et al. (2004) and Brinchmann et al. (2004). Overall, we find that the gas-phase oxygen abundance in MULTIDARK-GALAXIES increases with stellar mass up to $M_\star \sim 10^{11}\,\mathrm{M_\odot}$. Beyond this value it drops and reaches a plateau.

The SAG and SAGE model galaxies under-predict the gas-phase oxygen abundance by an average factor of $\sim 0.02\,\mathrm{dex}$. This systematic offset for SAGE is not predictive, but purely due to the fact that this model was calibrated by assuming a different value of $(O/H)_\odot/Z_\odot$, specifically $[12 + \log_{10}(O/H)] = [9 + \log_{10}(Z_{cold}/0.02)]$; for further details, see Knebe et al. (2018).

At $M_\star < 10^{10.2}\,\mathrm{M_\odot}$, GALACTICUS also under-predicts the gas-phase abundance by the same factor. However, this model exhibits a bump at $M_\star \sim 10^{10.5}\,\mathrm{M_\odot}$. This feature is related to the excess of galaxies around this stellar mass, which is seen in the galaxy stellar mass function (see Fig. 3). This excess was found to be produced by the depletion of gas due to the extreme AGN feedback mechanism imple-

---

[13] http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/

mented in GALACTICUS, where the galaxies have almost no inflow of pristine gas, and rapidly consume their gas supply (for further details, see Knebe et al. 2018).

We have investigated further this feature finding that, if we exclude galaxies with progressively higher cold gas fraction (CGF), which is defined as CGF=$M_{cold}/M_\star$, the bump shrinks continuously. Fig. 10 is produced by combining two cuts: CGF> 0.1 and sSFR> 10^{-11} yr^{-1}. The first one eliminates about half of the GALACTICUS model galaxies, most of them with unrealistically small CGFs, possibly meaning that their metallicities are not reliable due to the precision used in evolving the relevant ordinary differential equations (Benson 2012). The second cut selects only very star-forming galaxies. The bump completely disappears for CGF> 0.5, but in that case ∼ 70% of the galaxies are excluded from the sample.

Fig. 11 compares the intrinsic [O II] luminosity as a function of SFR for the three SAMs (coloured, filled contours) with the DEEP2-FF data at $z \sim 1$ (grey, shaded squares). We also show the results of the conversions given in Eqs. 15-17 (diagonal, black and green lines). The model L[O II] is computed using the GET_EMLINES code coupled with instantaneous SFR for SAG, and average SFR for the other semi-analytic models. The distributions of SAG, SAGE and GALACTICUS behave in a similar way, reproducing the bimodality observed in the data. The coloured lines (dashed, salmon; solid, yellow; dot-dashed, blue) are the linear fits to the model L[O II]-SFR correlations. The best-fit parameters, correlation coefficients (r-values) and dispersions in both directions are reported in Table 2.

Fig. 11 shows that all the model galaxies considered overlap with the DEEP2-FF observations and extend further towards lower SFR values. All three SAMs cover the L[O II] observational range with their 2σ regions. SAGE and GALACTICUS get to the very bright domain of the parameter space, while SAG is limited to fainter L[O II] values.

All the SAMs are tightly correlated in the SFR–luminosity plane and such a trend is in reasonable agreement with the observationally derived relations from Eqs. 15-17 (diagonal, black and green lines).

In Fig. 11, the Kewley et al. (2004) parametrisation (green line and contours in Fig. 11) appears above all the GET_EMLINES derivations. These contours are obtained from Eq. 17, by inputting instantaneous (average) SFR and cold gas metallicity for SAG (SAGE, GALACTICUS) model galaxies. The green, straight lines are calculated by feeding the median metallicity values in bins of SFR into Eq. 17. Although both the Kewley et al. (2004) relation and the GET_EMLINES code assume the same cold gas metallicity values as inputs, the obtained distributions are very different. The width of the distributions is model-dependent and the L[O II] obtained for galaxies in SAG and Galacticus present bimodal distributions. This bimodality comes from the MAPPINGS-III term $F(\lambda_j, q, Z_{cold})$ in Eq. 8, that is a non-linear function of Z_{cold}.

4.2   L[OII] versus broad-band magnitudes

At a given redshift range, the broad-band magnitudes tracing the rest-frame UV emission of a galaxy are expected to be tightly correlated with the SFR and the production of emission line galaxies. The rest-frame UV slope $(1000 - 3000\,\text{Å})$ at $z \sim 1$ is measured between the u and the $g-$bands ($\sim 2000\text{Å}$). As expected, these are the bands

that correlate the most with both SFR and [O II] luminosity for the sample under study.

The correlations between the broad-band u and g absolute magnitudes and the intrinsic [O II] luminosity in MULTIDARK-GALAXIES at $z \sim 1$ are displayed in the first two columns of panels in Fig. 12 together with DEEP2-FF observations. Data and all model galaxies show a good overlap in this parameter space. The observations populate a smaller region of the parameter space, while the SAMs extend down to lower SFR and L[O II] values. We over plot all the strong correlations (i.e. those with correlation coefficient $r \geqslant 0.6$) as linear scaling laws with an associated scatter σ_y. Their best-fit parameters (A, B) and correlation coefficients (r) can be found in Table 2, where relations with $r < 0.6$ have been omitted. We find both the u and g magnitudes to be tightly correlated with L[O II], and thus they have the potential to be used as proxies for the [O II] luminosity, using the relations presented in Table 2.

4.3   L[OII] versus age, metallicity and stellar mass

We also study the dependence of the [O II] luminosity on galaxy properties that are relevant to the L[O II] and $(k+e)$ calculations: the age, metallicity, and stellar mass.

The right column of panels in Fig. 12 shows the relationship between the intrinsic [O II] luminosity and the stellar mass in both DEEP2-FF and our model galaxies. In SAGE we identify a correlation, but none is found for SAG and GALACTICUS model galaxies. The DEEP2-FF data do not exhibit any particular trend, maybe due to the narrow luminosity range that the sample covers.

In the third column of Fig. 12, we display the relationship between the intrinsic L[O II] and age, which is mostly flat both in MULTIDARK-GALAXIES and DEEP2-FF observations, with the latter showing a bimodal distribution. Only GALACTICUS model galaxies exhibit an anti-correlation in the age-L[O II] plane.

No correlation is found between the metallicity and L[O II] for any of the models (this is not shown in Fig. 12). We conclude that none of the galaxy properties explored in this Section are good candidates as proxies for L[O II].

## 4.4   From galaxy properties to L[OII]

The L[O II] derived from the GET_EMLINES code is tightly related to the SFR by construction, but we found it to be also tightly related with the broad-band u and g magnitudes ($r \geqslant 0.64$, see Table 2). Here, we quantify the usability of the linear relations found as proxies to derive L[O II] from average SFR and broad-band magnitudes. For this purpose, we compare the luminosity functions and galaxy clustering signal for [O II] emitters selected using the aforementioned linear relations and the relations from Section 4.1, with those obtained by coupling the SAMs with the GET_EMLINES code (see Section 3.1).

4.4.1   [OII] luminosity functions

In the left column of Fig. 13, from top to bottom, we show the attenuated [O II] luminosity functions of the SAG, SAGE and GALACTICUS model galaxies at $z \sim 1$. We compare the L[O II] predictions from coupling the models with the GET_EMLINES code (thick, coloured lines without error bars) with those from using the SFR (solid, black),
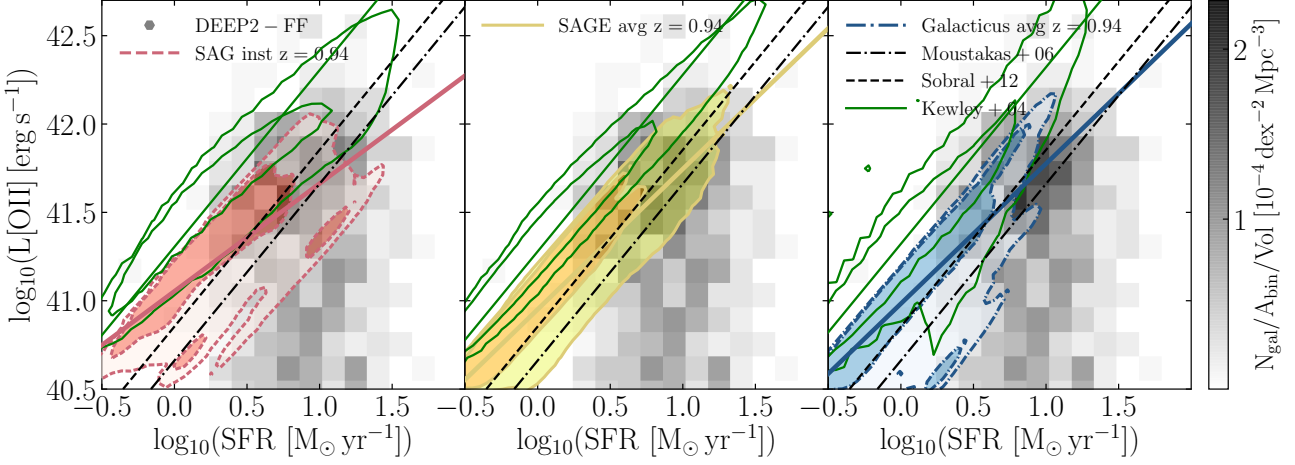
Figure 11. Intrinsic [O II] luminosity as a function of the SFR from the MULTIDARK-GALAXIES at $z \sim 1$ (salmon, yellow and blue, filled contours), compared with the DEEP2-FF observations at $0.9 < z < 1.1$ (grey, shaded squares, colour-coded with the density of emitters per 2D bin area). The innermost (outermost) contour represents 68% (95%) of the galaxy distributions. For SAG model galaxies, the [O II] luminosities have been computed from instantaneous SFRs, while for the other SAMs they are based on average SFRs. Both data and model ELGs are selected imposing a minimum [O II] flux of 5×10^{-18} erg s^{-1} cm^{-2}. The thick, coloured, diagonal lines are the linear fits to each SAM distribution, and their best-fit parameters are reported in Table 2. The black dot-dashed and dashed, diagonal lines are the L[O II] predictions obtained from the SFR range of interest using Eqs. 15 and 16, respectively. The green, empty contours are the Kewley et al. (2004) predictions obtained using Eq. 17 with SFR and cold gas metallicity as inputs. The green, solid lines are the same predictions assuming median metallicity values in bins of SFR.
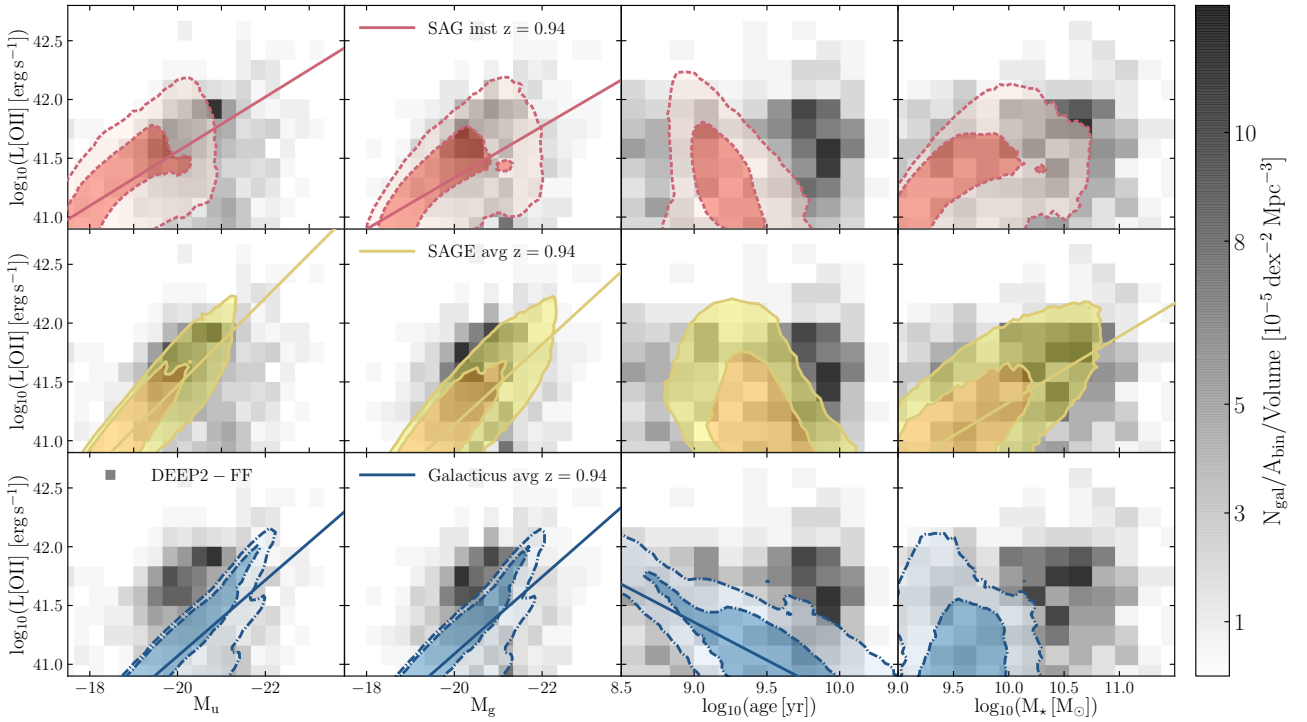
Figure 12. From top to bottom and from left to right: SAG, SAGE and GALACTICUS $z \sim 1$ intrinsic [O II] luminosities versus broad-band magnitudes, ages and stellar masses (contours) compared with the DEEP2-FF observations at $0.9 < z < 1.1$ (grey, shaded squares). The L[O II] values are computed using the GET_EMLINES code with instantaneous SFR for SAG and average SFR for SAGE and GALACTICUS. The innermost and outermost model contours represent 68% (1σ) and 95% (2σ) of the distribution. A minimum [O II] flux cut of 5×10^{-18} erg s^{-1} cm^{-2} has been applied to both data and model galaxies. The diagonal lines are the linear fits for strong correlations with $r > 0.6$, as reported in Table 2.

z=1		SAG	SAGE	GALACTICUS
$\log_{10}(L[\text{O\,\textsc{ii}}]/\text{erg\,s}^{-1}) = A\log_{10}(\text{SFR}/\text{M}_\odot\,\text{yr}^{-1})+B$	A	0.609±0.001	0.792±0.001	0.795±0.001
	B	41.05±0.01	40.98±0.01	40.95±0.01
	$\sigma_{\log(\text{SFR})}$	0.50	0.53	0.48
	$\sigma_{\log(\text{L[OII]})}$	0.38	0.45	0.46
	r	0.80	0.92	0.83
$\log_{10}(L[\text{O\,\textsc{ii}}]/\text{erg\,s}^{-1}) = A\,M_u+B$	A	-0.231±0.001	-0.373±0.001	-0.323±0.001
	B	36.93±0.01	34.01±0.01	34.61±0.01
	σ_{M_u}	1.07	1.05	1.18
	$\sigma_{\log(\text{L[OII]})}$	0.38	0.45	0.46
	r	0.65	0.86	0.83
$\log_{10}(L[\text{O\,\textsc{ii}}]/\text{erg\,s}^{-1}) = A\,M_g+B$	A	-0.218±0.001	-0.342±0.001	-0.328±0.001
	B	36.97±0.01	34.29±0.01	34.53±0.01
	σ_{M_g}	1.11	1.08	1.15
	$\sigma_{\log(\text{L[OII]})}$	0.38	0.45	0.46
	r	0.64	0.81	0.82
$\log_{10}(L[\text{O\,\textsc{ii}}]/\text{erg\,s}^{-1}) = A\log(\text{age}/\text{yr})+B$	A	—	—	-0.646±0.001
	B	—	—	47.17±0.01
	$\sigma_{\log(\text{age})}$	—	—	0.54
	$\sigma_{\log(\text{L[OII]})}$	—	—	0.46
	r	-0.44	-0.47	-0.76
$\log_{10}(L[\text{O\,\textsc{ii}}]/\text{erg\,s}^{-1}) = A\log(M_\star/\text{M}_\odot)+B$	A	—	0.563±0.001	—
	B	—	35.70±0.01	—
	$\sigma_{\log(\text{M}\star)}$	—	0.52	—
	$\sigma_{\log(\text{L[OII]})}$	—	0.45	—
	r	0.54	0.64	0.03

Table 2. Best-fit parameters of the linear scaling relations shown in Figs. 11 and 12. All the [O\,\textsc{ii}] luminosities here are intrinsic and computed using the GET_EMLINES code with input the instantaneous SFR for SAGE and average SFR for SAGE and GALACTICUS.

M_u (dashed, green) and M_g (dot-dashed, orange) proxies established above and summarised in Table 2. The shaded regions represent the effect of the scatter σ_y on the proxy-$L[\text{O\,\textsc{ii}}]$ relation and are derived from LFs estimated from 100 Gaussian realisations $G(\sigma_y,\mu)$ with mean $\mu = (\text{SFR}, M_u, M_g)$ and fixed scatter $\sigma_y = (\sigma_{\text{SFR}}, \sigma_{\text{M}_u}, \sigma_{\text{M}_g})$ from Table 2.

The [O\,\textsc{ii}] luminosity functions derived from the proxies are strongly model dependent, with varying levels of success for each model and proxy, as can be seen in Fig. 13. In SAG, the M_u proxy produces a luminosity function which, in the $L[\text{O\,\textsc{ii}}]$ range $10^{41.7} - 10^{42.5}$ erg s^{-1}, is consistent with that derived from coupling the model with the GET_EMLINES code, while the other two proxies are lower. In SAGE, the M_g proxy returns a LF in very good agreement with the GET_EMLINES estimate on all luminosity scales. M_u gives good agreement at $L[\text{O\,\textsc{ii}}] \lesssim 10^{42}$ erg s^{-1}, while beyond this value it slightly overestimates the number of [O\,\textsc{ii}] emitters. The SFR proxy is consistent with the GET_EMLINES result at $L[\text{O\,\textsc{ii}}] \lesssim 10^{41.7}$ erg s^{-1}, while at higher $L[\text{O\,\textsc{ii}}]$ values it overpredicts the luminosity function by ~ 1.5 dex.

The $L[\text{O\,\textsc{ii}}]$ function based on the SFR proxy from GALACTICUS is in reasonable agreement with that from coupling the model with GET_EMLINES, while the magnitude proxies produce a lack of emitters on all luminosity scales (~ 1.4 dex at $\sim 10^{40.5}$ erg s^{-1}, ~ 0.4 dex at $\sim 10^{41.5}$ erg s^{-1} and ~ 1.8 dex at $\sim 10^{42}$ erg s^{-1}). Fig. 12 shows that GALACTICUS magnitudes are below those from DEEP2-FF. This discrepancy is likely to be the cause of the lack of [O\,\textsc{ii}] emitters.

In the right column of Fig. 13, we display the intrinsic $L[\text{O\,\textsc{ii}}]$ functions colour-coded as the left panels. In SAG and SAGE model galaxies, the effect of dust attenuation is stronger at higher luminosities, while in GALACTICUS it is more significant at $L[\text{O\,\textsc{ii}}] \lesssim 10^{42}$ erg s^{-1}. We overplot, as
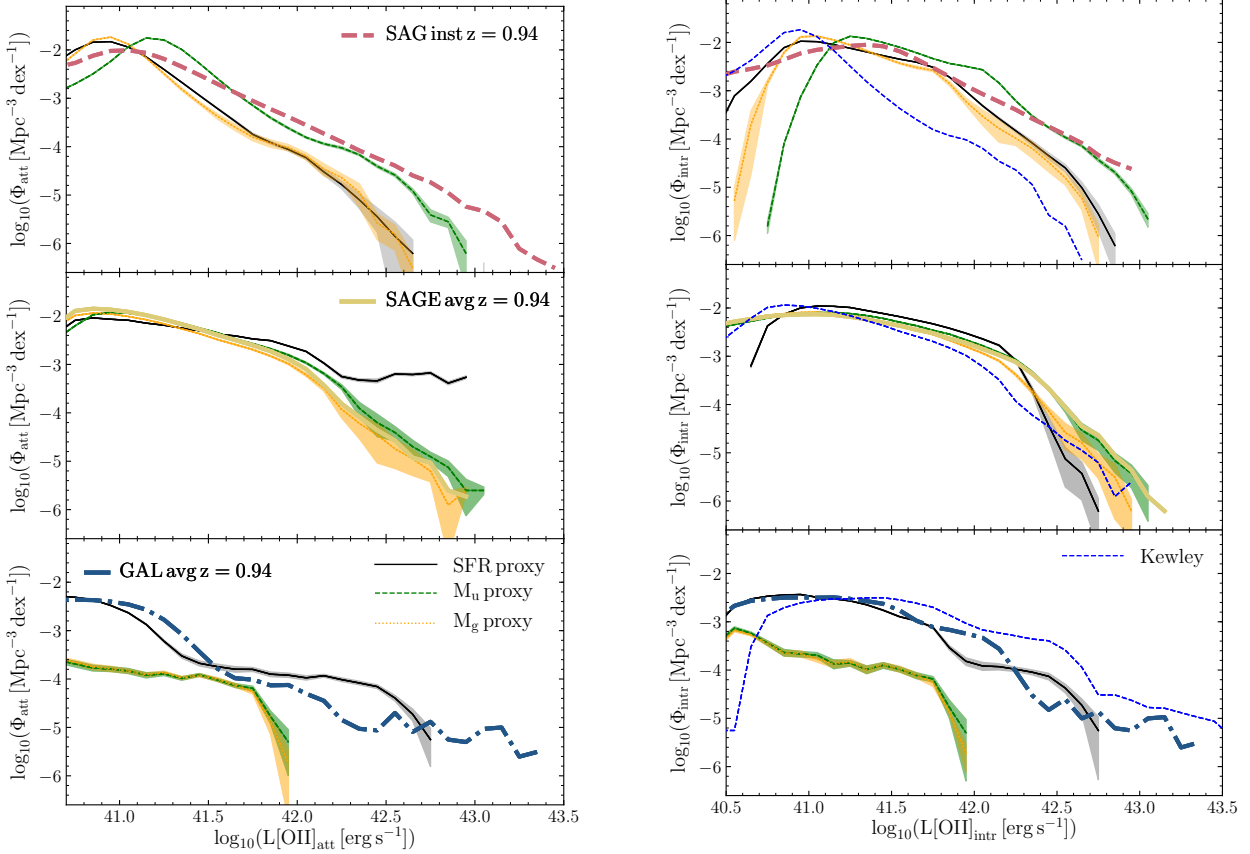
dashed, blue lines, the [O\,\textsc{ii}] luminosity functions obtained by applying the Kewley et al. (2004) conversion (Eq. 17) to each one of the model catalogues. This lies below (above) the other results in the bright end for SAG and SAGE (GALACTICUS) model galaxies. The relation from Kewley et al. (2004) produces very different $L[\text{O\,\textsc{ii}}]$ functions compared to the ones obtained from the SAM model galaxies coupled with the GET_EMLINES prescription. This result highlights that the dispersion in the model gas metallicities is not the only source of the variation seen in the luminosity function in Fig. 13.

In this Section, we have investigated the impact in the [O\,\textsc{ii}] luminosity function of using the $L[\text{O\,\textsc{ii}}]$ proxies established above. We find the $L[\text{O\,\textsc{ii}}]$ proxies to be model-dependent and to overall result in either a lack or an excess of bright [O\,\textsc{ii}] emitters. These outcomes emphasise the inappropriateness of using simple relations to derive the [O\,\textsc{ii}] emission from global galaxy properties. In fact, besides introducing systematic uncertainties, they can also result in [O\,\textsc{ii}] luminosity functions with very different shapes depending which properties are used.

4.4.2 Galaxy clustering

We further check how the clustering of our model ELGs is sensitive to an [O\,\textsc{ii}] luminosity selection, where $L[\text{O\,\textsc{ii}}]$ is computed either from the GET_EMLINES code, or the proxies established above. We consider SAG, SAGE and GALACTICUS model galaxies at $z \sim 1$ and impose on them a minimum luminosity threshold of $L[\text{O\,\textsc{ii}}] > 10^{40.4}$ erg s^{-1}.

Fig. 14 shows the ratios between the projected two-point correlation functions obtained from the proxy-to-$L[\text{O\,\textsc{ii}}]$ relations and those derived from $L[\text{O\,\textsc{ii}}]$ computed using the

Figure 13. *Left column:* From top to bottom, attenuated [O II] luminosity functions of the SAG, SAGE and GALACTICUS model galaxies at $z \sim 1$. We show as thick lines the results with L[O II] computed using the GET_EMLINES code described in Section 3.1 with either instantaneous or average SFR and metallicity as inputs. We compare these results with the L[O II] functions derived from the three L[O II] proxies established above: SFR (solid, black line), M_u (dashed, green) and M_g (dotted, orange). The shaded regions represent the $\pm\sigma_y$ scatter in the proxy-L[O II] linear scaling laws, which is given in Table 2. *Right column:* Same as left column, but here the L[O II] are intrinsic. The lines are colour-coded as the left panels. We show as blue dashed lines the results from the Kewley et al. (2004) conversion.

GET_EMLINES code with instantaneous SFR (SAG) or average SFR (SAGE and GALACTICUS). In Fig. 14, we also show the clustering of the data obtained using the conversion from Kewley et al. (2004) given in Eq. 17. For all the models, this clustering is in excellent agreement with the data derived from the GET_EMLINES L[O II] estimation.

For the clustering we adopt the Landy & Szalay (1993) estimator and the two-point function code from Favole et al. (2016b). The shaded regions present the effect of the σ_y scatter given in Table 2 in the proxy-L[O II] linear relations. The dispersion is computed from the covariance of 100 Gaussian realisations with mean the desired proxy and scatter σ_y (see Section 4.4.1 for further details).

The clustering amplitude remains similar (within 12%) for the different L[O II] calculations in all the SAM considered. In particular, in SAG and SAGE galaxies, all the proxies agree within 5% with the GET_EMLINES and Kewley et al. (2004) results on all scales. On small scales, the SFR proxy in SAG declines by 5% and in SAGE it shows some small fluctuations. In Galacticus, the clustering amplitude diminishes by up to 12% (4%) on small (intermediate) scales when assuming any proxy.

The two point correlation functions at $r_p > 1h^{-1}$Mpc are consistent with each other, agreeing within the $1\sigma_y$ dispersion.

We have investigated further the redshift evolution at $0.6 < z < 1.2$ and the dependence of different L[O II] thresholds of the MULTIDARK-GALAXIES clustering amplitude, both based on estimates from coupling the models with GET_EMLINES and on the proxies above. In general, we find that increasing both the redshift and the L[O II] thresholds, the galaxy number density decreases, resulting in a noisier clustering. Despite this increased noise, we find that model galaxies with L[O II] $> 10^{41}$erg s^{-1} can be more clustered when L[O II] is derived from proxies. We find variation among the different proxies used together with one of the three SAMs explored here. This possible dependency with L[O II] should be taken into account when using proxies to create fast galaxy catalogues for a particular survey.

Overall, we find that the MULTIDARK-GALAXIES clustering signal is model-dependent. The linear bias is mostly unchanged, however differences are seen at small scales, below $1h^{-1}$Mpc. The dispersion changes between the different proxies, with the SFR presenting the largest scatter, overall.

Our ELG clustering results show that simple L[O II] estimates based on a linear relation with SFR are sufficient for modelling the large scale clustering of [O II] emitters, even if they are not accurate enough to predict the [O II] luminosity function.
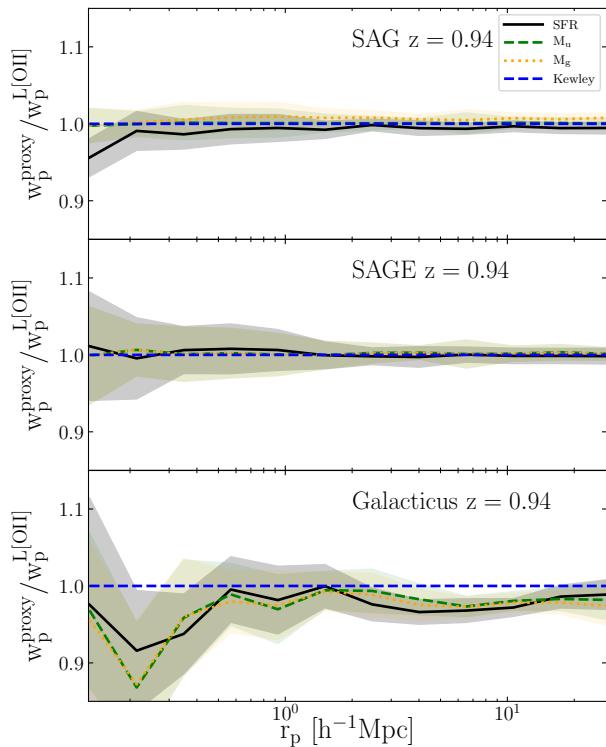
Figure 14. Proxy-to-L[O II] ratios of the projected two-point correlation functions of, from top to bottom, SAG, SAGE and GALACTICUS model galaxies at $z \sim 1$. The SAG L[O II] is estimated using the GET_EMLINES code with instantaneous SFR, while SAGE and GALACTICUS using the average quantity. Galaxies have been selected to have L[O II] $> 10^{40.4}$ erg s^{-1}. The shaded regions represent the effect of the σ_y scatter in the proxy-L[O II] linear relations reported in Table 2. These regions are the 1σ uncertainties derived from the co-variance of 100 Gaussian realisations with the L[O II] proxy considered as mean and σ_y as scatter. We over plot the Kewley et al. (2004) result as a dashed, blue line.

4.4.3   [O II] ELG Halo Occupation Distribution

In Fig. 15, we show the MULTIDARK-GALAXIES mean halo occupation distribution (HOD) for model galaxies selected with L[O II] $> 10^{40.4}$ erg s^{-1}. Here, the model [O II] luminosities have been calculated using the GET_EMLINES code. We highlight contributions from central and satellite model galaxies. The shapes of the HODs are qualitatively consistent among the different models, with an asymmetric Gaussian for central galaxies, plus maybe a plateau, and a very shallow power law for satellite galaxies. A similar shape has been found using different models for either young or star-forming galaxies, selected in different ways (Zheng et al. 2005; Contreras et al. 2013; Cochrane & Best 2018; Gonzalez-Perez et al. 2018) and also in measurements derived from observations (Geach et al. 2012; Cochrane et al. 2017; Guo et al. 2018).

The shape of the HOD for central star-forming galaxies is very different from those selected with a cut in either rest-frame optical broad-band magnitudes or stellar mass, which is close to a smooth step function that reaches unity (e.g. Berlind & Weinberg 2002; Kravtsov et al. 2004). As it can be seen in Fig. 15, the HOD of MULTIDARK-GALAXIES central [O II] emitters does not necessarily reach unity, i.e. it is not
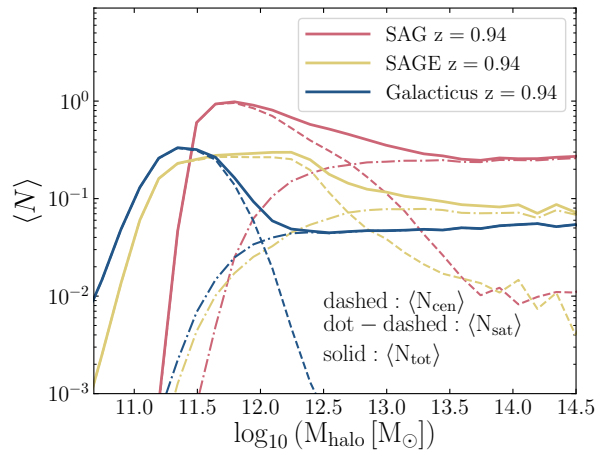
Figure 15. Mean halo occupation distribution of the SAG (salmon solid line), SAGE (yellow solid line) and GALACTICUS (blue solid line) model galaxies with L[O II] $> 10^{40.4}$ erg s^{-1} at $z \sim 1$. The model L[O II] has been computed using GET_EMLINES with instantaneous SFR for SAG galaxies and average SFR for SAGE and GALACTICUS. The contribution from central galaxies is shown by dashed lines and that for satellites by dot-dashed lines.

guaranteed to find an [O II] emitter in every dark matter halo above a given mass.

We find that the SAG HODs peak at higher halo masses compared to the other two SAMs. The mean halo masses predicted by the SAG, SAGE and GALACTICUS model galaxies are in agreement with the results of Favole et al. (2016a) for BOSS [O II] ELGs at $z \sim 0.8$ and Favole et al. (2017) for SDSS [O II] ELGs at $z \sim 0.1$.

The HOD of MULTIDARK-GALAXIES satellite [O II] emitters is a very shallow power law, closer to a smooth step function. This is similar to what has been inferred for eBOSS [O II] emitters (Guo et al. 2018), but very different to the findings using the GALFORM semi-analytical model (Gonzalez-Perez et al. 2018). This difference is most likely related to a different treatment of gas in this model, as the distribution of satellites in dark matter haloes of different masses is very sensitive to both the modelling of feedback and environmental processes.

In Fig. 16, we display the ratios between the MULTIDARK-GALAXIES HODs selected in L[O II], where the luminosity is calculated from either using the GET_EMLINES code or the proxies indicated in the legend. We find that the differences in the HODs from proxies and GET_EMLINES are negligible for GALACTICUS and less than 20% for SAG at $M_{\rm halo} \gtrsim 10^{12} M_\odot$, while SAGE shows differences above a factor 1.5 in most cases. The L[O II] proxies behave very similarly, with negligible differences between them, except for the GALACTICUS SFR proxy, which is slightly lower than the magnitude ones.

In summary, we find different levels of agreement with the GET_EMLINES results depending on the model considered. However, the HOD remains almost unchanged when different L[O II] proxies are assumed.

5   SUMMARY AND CONCLUSIONS

In this work, we have explored how the [O II] luminosity can be estimated for semi-analytic models of galaxy formation
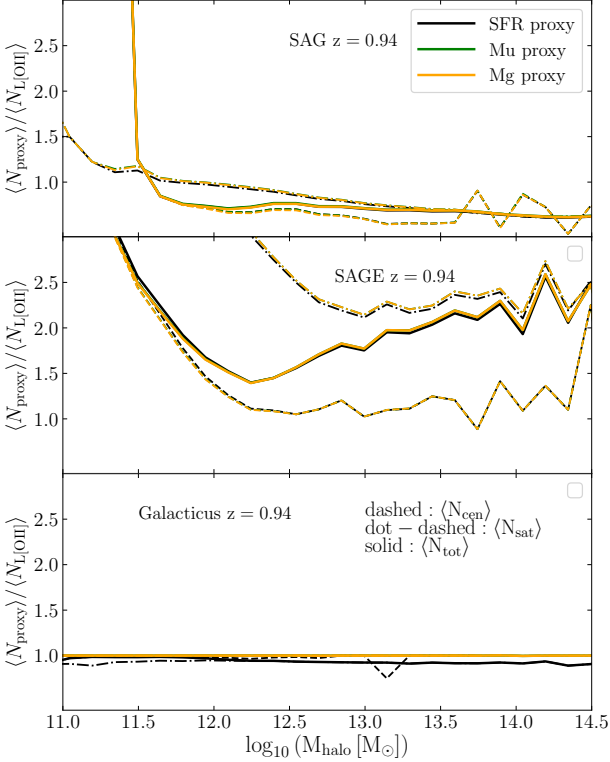
Figure 16. Ratio between the HOD obtained from the L[O II] calculated from the proxies indicated in the legend, and L[O II] obtained using GET_EMLINES. From top to bottom, results are shown for the SAG, SAGE and GALACTICUS models.

and evolution using different methods: (i) by coupling the SAMs with the GET_EMLINES code (Section 3.1) and (ii) using simple relations between L[O II] and global properties such as SFR, broad-band magnitudes and metallicity (Section 4.1).

We have studied the following models from the MULTIDARK-GALAXIES products (Knebe et al. 2018): SAG (Cora et al. 2018), SAGE (Croton et al. 2016) and GALACTICUS (Benson 2012). All these models are run on the MDPL2 cosmological simulation (Klypin et al. 2016). They were calibrated to a number of observations within $0 < z < 2$, and they produce SFR and stellar mass functions that evolve similarly to what is observed in this redshift range.

Throughout this study, we have compared our model results with different observational data sets, including DEEP2-FF galaxies with absolute magnitudes (see Section 2.2).

The GET_EMLINES code to calculate nebular emission lines is publicly available and ideally uses instantaneous SFR as input. However, usually SAMs only output SFRs that are averaged over long time intervals, corresponding to the outputs of the underlying dark matter simulation. From the SAMs under study, only SAG provides instantaneous SFRs. We have coupled the GET_EMLINES code with the SAG model using both instantaneous and average SFRs to study the impact that this choice has on the L[O II] calculation in post-processing. Assuming as input for the GET_EMLINES code either the instantaneous or the average SFR, we see a variation below 5% for the dust attenuated [O II] luminosity functions in the range $10^{41} - 10^{42.2}$ erg s^{-1}, and in the range $10^{41} - 10^{3}$ erg s^{-1} for the intrinsic [O II] lumi-

nosity functions. These ranges correspond to model ELGs with $1 <$SFR (yr^{-1}M$_\odot$) $< 10^{1.5}$. At higher and lower SFRs, there is a larger discrepancy, $< 50\%$, in the luminosity functions, when using either the average or the instantaneous SFR. Thus, we find that using average SFRs as inputs for GET_EMLINES is a good approach when studying average galaxy populations.

The luminosity functions of the MULTIDARK-GALAXIES with L[O II] computed using the GET_EMLINES algorithm are in good agreement with the DEEP2 and VVDS observations over the redshift range $0.6 < z < 1.2$. The [O II] luminosity, SFR and stellar mass functions of the SAMs all consistently predict a smaller number of massive, star-forming emitters as the redshift increases. The match we find in the [O II] luminosity functions of model and DEEP2 galaxies, where the model L[O II] values are computed using the GET_EMLINES code, cannot guarantee that they are the same identical population of galaxies. In other words, we select the SAG, SAGE and GALACTICUS model galaxies to best reproduce the characteristics of the observed DEEP2 [O II] emitters. These selections return different levels of agreement in the explored parameter spaces, as shown in Figs. 8, 12, A2, C1. A remarkable result from this study is that our model galaxies span the same regions as the observed ones, in all the parameter spaces under study, with overall consistent trends. This suggests that our modelling approach captures the most important physical processes that shape the DEEP2 galaxy sample.

We have also investigated the viability of obtaining L[O II] from simple relations with global galactic properties that are usually outputted by galaxy formation models. For this purpose, we use observationally derived relations (Kewley et al. 2004) and linear relations derived for each model. In particular, we explore the L[O II] derived using the GET_EMLINES code as a function of SFR, broad-band magnitudes, age and stellar mass. The SFR, both instantaneous and average, is the physical quantity that, by construction, is most correlated with the [O II] luminosity (with correlation coefficients $r \geqslant 0.80$ for all the SAMs). Such a tight correlation is well described by a linear scaling law with an associated scatter $\sigma_{\log(\text{SFR})}$ that varies with L[O II] (see Table 2). Other valuable proxies to derive L[O II] are the observed-frame u and g broad-band magnitudes, M_u and M_g, which trace the rest-frame UV emission in our redshift range of interest.

We test how feasible it is to use these correlations as proxies for L[O II] by studying the evolution of the derived [O II] luminosity functions, mean halo occupation distribution (HOD) and the galaxy clustering signal in L[O II] thresholds.

The different methods explored to calculate L[O II] result in a range of [O II] luminosity functions. Taking into account the effect of the scatter in the SAG L[O II]–proxy relations, the luminosity functions from the proxies (including the Kewley et al. (2004) relation from Eq. 17) are in reasonable agreement with the direct GET_EMLINES estimates. The differences are larger for the relations derived from M_g and SFR in SAG at all luminosities, for SFR in SAGE at L[O II] $> 10^{42}$ erg s^{-1}, and for the magnitude proxies in GALACTICUS. At high luminosities, L[O II] derived with most linear proxies result in a lack of bright emitters that increases with luminosity, but remains approximately constant with redshift. The Kewley et al. (2004) relation (Eq. 17) results in a lower number of bright [O II] emitters compared to all

the other methods to obtain L[O II] in SAG and SAGE, and in a higher number in GALACTICUS.

We find a large variation between the derived [O II] luminosity functions among both the SAMs and the methods used to obtain L[O II]. Thus, it is important to highlight that, despite the model SFR density evolution being in reasonable agreement with observations, simple relations based on global galaxy properties are not robust estimators for L[O II].

We further test the use of simple relations to obtain L[O II] for SAMs by measuring the galaxy two-point autocorrelation function for [O II] emitters selected above a given L[O II] threshold. We compare the clustering measured from the [O II] proxies with direct predictions from the SAMs coupled with the GET_EMLINES code and with the Kewley et al. (2004) relationship. The results vary from model to model and the largest fluctuations are seen below $1h^{-1}$Mpc. However, if we account for the effect of the scatter in the proxy-L[O II] relation, the discrepancies reconcile with direct luminosity predictions. The large scale bias remains similar for all the models.

By increasing the L[O II] threshold, the galaxy number density drops considerably resulting in a noisier clustering signal, which makes the comparison difficult. Despite this increased noise, we find that model galaxies with L[O II] $> 10^{41}$erg s^{-1} can be more clustered when L[O II] is derived from proxies (this depends both on the model and the proxy used). This possible dependency with L[O II] should be taken into account when using proxies to create fast galaxy catalogues for a particular survey.

There is no direct correspondence between a proxy resulting in a good luminosity function and providing a similar outcome for the clustering.

We also test how the mean HOD of [O II] emitters changes when assuming different proxies compared to the GET_EMLINES code in the L[O II] calculation of our SAMs. We find that the shape of the HOD is consistent with that expected for a star-forming population of galaxies. Quantitatively, the HOD is strongly model-dependent, and we find different levels of agreement between the proxies and the GET_EMLINES results, in particular at $M_{\rm halo} \gtrsim 10^{12}M_\odot$. However, the distributions remain substantially unchanged from one proxy to another for all the models under study.

Our results show that ELGs are different from SFR-selected samples and that the L[O II] estimation needs more complex modelling than assuming a linear relation with SFR. Simple L[O II] estimates are not accurate enough to predict direct statistics of L[O II], as the luminosity function, but they are sufficient for modelling the large scale clustering of [O II] emitters.

New-generation optical and infra-red surveys will provide enormous data sets with unprecedented spectroscopic precision and imaging quality. These observations, together with models of galaxy formation and evolution, will enable us to reach a complete and consistent understanding of both the Universe large scale structure, and the galaxy formation and evolution processes within dark matter haloes. In this context, simple derivations of L[O II] might be adequate for the clustering above $1h^{-1}$Mpc, although at least two simple approximations might be needed to determine the uncertainties.

6   DATA AVAILABILITY

The data produced for this article are available at http://popia.ft.uam.es/MultiDarkEmissionLines/. Here we provide the DEEP2–FIREFLY observations, both with and without dust attenuation, and the galaxy properties from the SAG, SAGE and GALACTICUS models. For the latter, besides the [O II] emission line, we also include the Hα, Hβ and [O III] luminosities.

PLOTLIB[14] 2012-2016 (Hunter 2007); PYTHON SOFTWARE FOUNDATION[15] 1990-2017, version 2.7., PYTHONBREW[16]; we use whenever possible in this work a colour-blind friendly colour palette[17] for our plots.

REFERENCES

Alam S. et al., 2015, ApJS, 219, 12
Allen M. G., Groves B. A., Dopita M. A., Sutherland R. S., Kewley L. J., 2008, The Astrophysical Journal Supplement Series, 178, 20
Allende Prieto C., Lambert D. L., Asplund M., 2001, ApJ, 556, L63
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Baugh C. M., 2006, Reports of Progress in Physics, 69, 3101
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Benson A. J., 2010, Phys. Rep., 495, 33
Benson A. J., 2012, New Astron., 17, 175
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014, A&A, 564, A66
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Brinchmann J., Charlot S., Heckman T. M., Kauffmann G., Tremonti C., White S. D. M., 2004, ArXiv, astro-ph/0406220
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Chabrier G., 2003, PASP, 115, 763
Chabrier G., Hennebelle P., Charlot S., 2014, ApJ, 796, 75
Cochrane R. K., Best P. N., 2018, MNRAS, 480, 864
Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913
Collacchioni F., Cora S. A., Lagos C. D. P., Vega-Martínez C. A., 2018, MNRAS, 481, 954
Comparat J. et al., 2017, ArXiv e-prints:1711.06575
Comparat J. et al., 2015, A&A, 575, A40
Comparat J. et al., 2016, MNRAS, 461, 1076
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, MNRAS, 432, 2717
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Cora S. A., 2006, MNRAS, 368, 1540
Cora S. A. et al., 2018, MNRAS, 479, 2
Croton D. J. et al., 2006, MNRAS, 365, 11
Croton D. J. et al., 2016, ApJS, 222, 22
Dawson K. S. et al., 2016, AJ, 151, 44
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
De Lucia G., Boylan-Kolchin M., Benson A. J., Fontanot F., Monaco P., 2010, MNRAS, 406, 1533
Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, A&A, 350, 381
Dickey C. M. et al., 2016, ApJ, 828, L11
Draine B. T., 2003, Annual Review of Astronomy and Astrophysics, 41, 241
Driver S. P. et al., 2018, MNRAS, 475, 2891
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Etherington J. et al., 2017, MNRAS, 466, 228

[14] http://matplotlib.org/
[15] http://www.python.org
[16] https://github.com/utahta/pythonbrew
[17] https://personal.sron.nl/~pault/

Faber S. M. et al., 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, Iye M., Moorwood A. F. M., eds., pp. 1657–1669
Falcón-Barroso J., Knapen J. H., 2013, Secular Evolution of Galaxies
Favole G. et al., 2016a, MNRAS, 461, 3421
Favole G., McBride C. K., Eisenstein D. J., Prada F., Swanson M. E., Chuang C.-H., Schneider D. P., 2016b, MNRAS, 462, 2218
Favole G., Rodríguez-Torres S. A., Comparat J., Prada F., Guo H., Klypin A., Montero-Dorta A. D., 2017, MNRAS, 472, 550
Font A. S., Johnston K. V., Ferguson A. M. N., Bullock J. S., Robertson B. E., Tumlinson J., Guhathakurta P., 2008, ApJ, 673, 215
Gargiulo I. D. et al., 2015, MNRAS, 446, 3820
Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, MNRAS, 426, 679
Gonzalez-Perez V. et al., 2018, MNRAS, 474, 4024
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264
Groves B. A., Dopita M. A., Sutherland R. S., 2004, The Astrophysical Journal Supplement Series, 153, 9
Gruppioni C. et al., 2015, MNRAS, 451, 3419
Guiderdoni B., Rocca-Volmerange B., 1987, A&A, 186, 1
Guo H. et al., 2018, arXiv:1810.05318 [astro-ph], arXiv: 1810.05318
Guo Q. et al., 2011, MNRAS, 413, 101
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
Henriques B., Maraston C., Monaco P., Fontanot F., Menci N., De Lucia G., Tonini C., 2011, MNRAS, 415, 3571
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
Hirschmann M., Charlot S., Feltre A., Naab T., Choi E., Ostriker J. P., Somerville R. S., 2017, MNRAS, 472, 2468
Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760
Hirschmann M., Naab T., Somerville R. S., Burkert A., Oser L., 2012, MNRAS, 419, 3200
Hou J., Lacey C. G., Frenk C. S., 2017, ArXiv e-prints
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Izquierdo-Villalba D., Bonoli S., Spinoso D., Rosas-Guevara Y., Henriques B. M. B., Hernández-Monteagudo C., 2019, MNRAS, 488, 609
Jiménez E., Contreras S., Padilla N., Zehavi I., Baugh C. M., Gonzalez-Perez V., 2019, arXiv e-prints, arXiv:1906.04298
Jouvel S. et al., 2009, A&A, 504, 359
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Kennicutt, Jr. R. C., 1998, ARA&A, 36, 189
Kewley L. J., Geller M. J., Jansen R. A., 2004, AJ, 127, 2002
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Knebe A. et al., 2018, MNRAS, 474, 5206
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Kroupa P., 2001, MNRAS, 322, 231
Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649
Lagos C. D. P., Cora S. A., Padilla N. D., 2008, MNRAS, 388, 587
Lagos C. d. P., Lacey C. G., Baugh C. M., 2013, MNRAS, 436, 1787
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Laureijs R. et al., 2011, ArXiv e-prints: 1110.3193

Levesque E. M., Kewley L. J., Graham J. F., Fruchter A. S., 2010, ApJ, 712, L26

Li C., White S. D. M., 2009, MNRAS, 398, 2177

Madau P., Dickinson M., 2014, Annual Review of Astronomy and Astrophysics, 52, 415

Maraston C. et al., 2013, MNRAS, 435, 2764

Maraston C., Strömbäck G., 2011, MNRAS, 418, 2785

Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359

Matthews D. J., Newman J. A., Coil A. L., Cooper M. C., Gwyn S. D. J., 2013, ApJS, 204, 21

McConnell N. J., Ma C.-P., 2013, ApJ, 764, 184

Merson A., Wang Y., Benson A., Faisst A., Masters D., Kiessling A., Rhodes J., 2018, MNRAS, 474, 177

Merson A. I., Baugh C. M., Gonzalez-Perez V., Abdalla F. B., Lagos C. d. P., Mei S., 2016, MNRAS, 456, 1681

Mohammad F. G. et al., 2018, A&A, 610, A59

Monaco P., Benson A. J., De Lucia G., Fontanot F., Borgani S., Boylan-Kolchin M., 2014, MNRAS, 441, 2058

Moustakas J. et al., 2013, ApJ, 767, 50

Moustakas J., Kennicutt, Jr. R. C., Tremonti C. A., 2006, ApJ, 642, 775

Muñoz Arancibia A. M., Navarrete F. P., Padilla N. D., Cora S. A., Gawiser E., Kurczynski P., Ruiz A. N., 2015, MNRAS, 446, 2291

Newman J. A. et al., 2013, ApJS, 208, 5

Nuza S. E. et al., 2013, MNRAS, 432, 743

Orsi Á., Padilla N., Groves B., Cora S., Tecce T., Gargiulo I., Ruiz A., 2014, MNRAS, 443, 799

Orsi Á. A., Angulo R. E., 2018, MNRAS, 475, 2530

Osterbrock D. E., 1989, Astrophysics of gaseous nebulae and active galactic nuclei

Pagel B. E. J., Edmunds M. G., Blackwell D. E., Chun M. S., Smith G., 1979, MNRAS, 189, 95

Planck Collaboration et al., 2015, ArXiv e-prints

Prugniel P., Soubiran C., Koleva M., Le Borgne D., 2007, ArXiv, astro-ph/0703658

Ruiz A. N. et al., 2015, ApJ, 801, 139

Sartoris B. et al., 2015, ArXiv e-prints: 1505.02165

Schlegel D. J. et al., 2015, in American Astronomical Society Meeting Abstracts, Vol. 225, American Astronomical Society Meeting Abstracts

Sobral D., Best P. N., Matsuda Y., Smail I., Geach J. E., Cirasuolo M., 2012, MNRAS, 420, 1926

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

Spitzer L., 1978, Physical processes in the interstellar medium

Stevens A. R. H., Brown T., 2017, MNRAS, 471, 447

Talia M. et al., 2015, A&A, 582, A80

Tomczak A. R. et al., 2014, ApJ, 783, 85

Tonini C., Maraston C., Devriendt J., Thomas D., Silk J., 2009, MNRAS, 396, L36

Tremonti C. A. et al., 2004, ApJ, 613, 898

Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16

Tully R. B., Fisher J. R., 1977, A&A, 54, 661

Valentino F. et al., 2017, MNRAS, 472, 4878

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, MNRAS, 372, 1161

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

Wilkinson D. M., Maraston C., Goddard D., Thomas D., Parikh T., 2017, MNRAS, 472, 4297

Yates R. M., 2014, The chemical evolution of galaxies in semi-analytic models and observations

Zheng Z. et al., 2005, ApJ, 633, 791

Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760
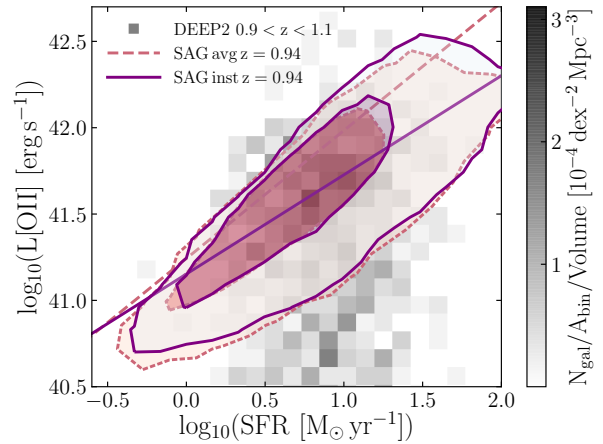
Figure A1. Non-attenuated [O II] luminosity as a function of SFR for the SAG model galaxies at $z \sim 1$ (contours) and the DEEP2-FF observations at $0.9 < z < 1.1$ (grey, shaded squares). The colour bar represents the observed galaxy number density in each 2D bin normalised by the bin area in units of $[\mathrm{dex}^{-2}\,\mathrm{Mpc}^{-3}]$. Here the SAG model galaxies are selected following the spline fit to the observed DEEP2-FF L[O II] distribution shown in Fig. 4. The model L[O II] values are calculated by assuming instantaneous (solid, purple contours) and average (dashed, salmon) SFR as input for the GET_EMLINES prescription. The innermost (outermost) model contours encompass 68% (95%) percent of the galaxy distributions. The diagonal lines represent the L[O II]-SFR correlations, whose coefficients are reported in Tab. A1.

| y=A x+B	A	B	σ_y	r
$y=\log_{10}(L[\mathrm{O\,II}])$				
$x=\log_{10}(\mathrm{SFR}_{avg})$	0.741±0.002	41.24±0.01	0.41	0.92
$x=\log_{10}(\mathrm{SFR}_{inst})$	0.574±0.003	41.15±0.01	0.38	0.77
$y=M_u$				
$x=\log_{10}(\mathrm{SFR}_{avg})$	-2.021±0.006	-18.04±0.01	1.16	0.88
$x=\log_{10}(\mathrm{SFR}_{inst})$	-2.084±0.005	-17.88±0.01	1.17	0.90
$y=M_g$				
$x=\log_{10}(\mathrm{SFR}_{avg})$	-2.006±0.005	-18.96±0.01	1.11	0.92
$x=\log_{10}(\mathrm{SFR}_{inst})$	-2.032±0.005	-18.84±0.01	1.11	0.93
$y=\log_{10}(M_\star)$				
$x=\log_{10}(\mathrm{SFR}_{avg})$	0.859±0.002	9.15±0.01	0.47	0.92
$x=\log_{10}(\mathrm{SFR}_{inst})$	0.846±0.002	9.11±0.01	0.47	0.90

Table A1. Best-fit parameters of the linear scaling relations found for SAG model galaxies at $z = 1$ and shown in Fig. A2. The parameter r is the correlation coefficient and σ_y is the scatter in the y-axis. SAG galaxies have been selected in L[O II] randomly drawn from the DEEP2-FF spline fit shown in Fig.4.

APPENDIX A: SAG MODEL GALAXIES SELECTED FROM DEEP2-FF SPLINE

We study the properties of a subset of SAG model galaxies selected to reproduce the L[O II] distribution of the DEEP2-FF data approximated by a spline fit in Fig. 4. We compare these model properties with the observational ones from DEEP2-FF.

Fig. A1 displays the SAG non-attenuated [O II] luminosities computed from average and instantaneous SFRs as a function of SFR. The bimodality observed in Fig. 5, where the model galaxies are selected by cutting at SFR > 0 and $\log(M_\star/M_\odot) > 8.87$, has now disappeared, but the discrep-
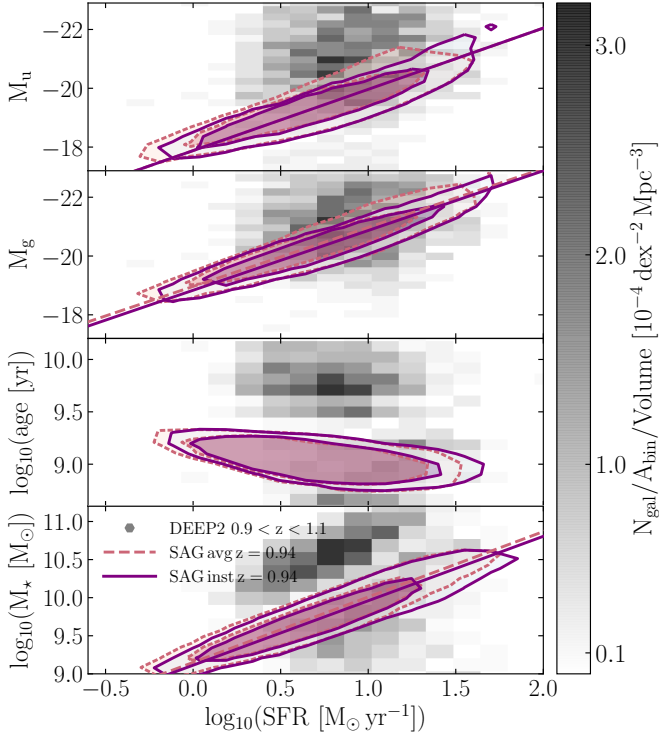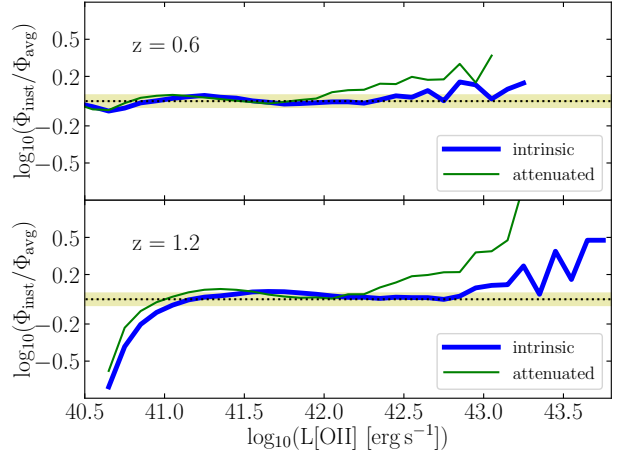
Figure A2. From top to bottom: intrinsic magnitudes, ages and stellar masses as a function of star formation rate for SAG model galaxies at $z \sim 1$ (contours) and DEEP2-FF observations at $0.9 < z < 1.1$ (grey, shaded squares). The model galaxies have been selected following the observed DEEP2-FF distribution approximated by a spline fit, as explained in Sec. 2.2.2. The colour bar shows the number density of DEEP2 galaxies per bin area in units of $[\mathrm{dex}^{-2}\,\mathrm{Mpc}^{-3}]$. The dashed, salmon (solid, purple) contours represent the average (instantaneous) SFRs. The innermost (outermost) contours encompass 68% (95%) of the distributions. The diagonal lines are the linear fits showing the significant correlations, whose coefficients are reported in Table 2, together with the best-fit parameters.

Figure B1. Ratios between the SAG [O II] luminosity functions (thick, blue lines: intrinsic LFs; thin, green: attenuated LFs) at different redshifts computed from average and instantaneous SFR using the method presented in 3.1. The yellow, shaded areas represent the 5% confidence region.

ancy between the two sets of contours is larger, with the instantaneous correlation much shallower than the average one. At $\mathrm{SFR} \gtrsim 10^1\,\mathrm{yr}^{-1}\mathrm{M}_\odot$, the instantaneous SFR returns higher $L[\mathrm{O\,II}]$ values compared to the average SFR. Instead, at $\mathrm{SFR} \lesssim 10^0\,\mathrm{yr}^{-1}\mathrm{M}_\odot$, the average contours reach fainter luminosities. Compared to the SAG results based on simple SFR and stellar mass cuts (see Fig. 5), here both sets of contours span a higher range of $L[\mathrm{O\,II}]$ and SFR values. The $L[\mathrm{O\,II}]$-SFR correlation based on average (instantaneous) SFR is stronger (less strong) and with a steeper (shallower) slope compared to that for galaxies selected with simple cuts (compare Tabs. 1 and A1), while the scatter is the same.

The DEEP2-FF observations in Fig. A1 seem to span a narrower range in SFR **and to go fainter in** $L[\mathrm{O\,II}]$ compared to the model galaxy contours. However, we highlight that the low-luminosity observational tail has a very low-density of emitters ($\sim 10^{-4}$ in Fig. A1).

In Fig. A2, from top to bottom, we display the intrinsic u- and g-band absolute magnitudes, the age and stellar mass of the SAG model galaxies selected from the DEEP2-FF spline fit as a function of the average and instantaneous SFRs. Compared to the results based on simple cuts at $\mathrm{SFR} > 0\,\mathrm{yr}^{-1}\,\mathrm{M}_\odot$ and $M_\star > 10^{8.87}\,\mathrm{M}_\odot$ (see Fig. 8), here

the correlations between SFR and magnitudes are steeper and M_u shows a wider scatter in the y-axis. On the contrary, the correlation between SFR and stellar mass is shallower for SAG galaxies drawn from the DEEP2-FF spline fit and with less scatter in the y-axis. The specific values of the correlation parameters and coefficients are reported in Table A1. Overall, SAG galaxies selected from the spline fit reach brighter values of u- and g-band magnitudes compared to their counterparts based on simple SFR and stellar mass cuts (see Fig. 8), which also extend down to fainter magnitudes and smaller stellar masses. As already noticed in Fig. 8, model galaxies have lower ages and stellar masses and they extend into larger SFR values, compared to the DEEP2-FF sample.

APPENDIX B: EVOLUTION OF L[O II] FROM INSTANTANEOUS AND AVERAGE SFR

We investigate further the redshift evolution of the small discrepancy generated in $L[\mathrm{O\,II}]$ by assuming average instead of instantaneous SFR as input for the GET_EMLINES code. In Section 3.3, we have studied what happens at $z \sim 1$, now we look over the redshift range $0.6 < z < 1.2$ to see if there is some evolution.

Fig. B1 compares the ratios of the intrinsic (thick, blue lines) and attenuated (thin, green) [O II] luminosity functions obtained from average and instantaneous SFR at different redshifts. We have explored the entire range $0.6 < z < 1.2$ finding that, as the redshift increases, the instantaneous and average $L[\mathrm{O\,II}]$ results tend to agree on a larger luminosity domain. Specifically, the 5% agreement threshold (yellow, shaded region in the plot) is reached for the first time at $z = 0.6, 1.2$ by galaxies with attenuated $L[\mathrm{O\,II}] = 10^{41.9}, 10^{42.3}\,\mathrm{erg\,s}^{-1}$ and with intrinsic $L[\mathrm{O\,II}] = 10^{42.6}, 10^{42.9}\,\mathrm{erg\,s}^{-1}$, respectively. This result is independent on the presence of attenuation in the [O II] luminosity. At $z \sim 1.2$, we observe a larger discrepancy in both ratios in the faint region due to the larger effect of incompleteness.

APPENDIX C: GLOBAL PROPERTIES OF MULTIDARK-GALAXIES

We compare pair properties in MULTIDARK-GALAXIES and DEEP2-FF observations to better understand their mutual correlations. We then fit these dependences using linear scaling relations. Fig. C1 displays, from top to bottom, the correlations between broad-band magnitudes, age and stellar mass as a function of SFR and stellar mass of the DEEP2-FF galaxies (grey, shaded squares, colour-coded according to their galaxy number density normalized by the 2D bin area) compared to the MULTIDARK-GALAXIES (contours indicating the 68% and 95% of each distribution). Data and models overlap covering the brighter, more massive and more star-forming region of the parameter space. In particular, the MULTIDARK-GALAXIES only cover the SFR range above the knee shown in Fig. 2.

For such a small observational sample, it is difficult to establish and fit clear correlations among these quantities and between these quantities and L[O II] (see also Fig. 12). In order to do this properly, one should account for all the DEEP2-FF incompleteness effects, which goes beyond the scope of our work. Here we show the comparison between the DEEP2-FF emitters and the MULTIDARK-GALAXIES only to verify that our models cover the parameter space of the observational data set.

From the model point of view, we do find clear correlation among most of the physical quantities presented in Fig. C1. Each set of panels shows the results for one model: from top to bottom we display SAG, SAGE and GALACTICUS model galaxies. The relevant correlations ($r \geqslant 0.6$) are represented as linear fits and the optimal parameters are reported in Table C1, together with their correlation coefficient (r) and the associated scatter in the y-axis (σ_y).

As expected, tight correlation is observed between the star formation rate and the broad-band u and g magnitudes that trace the rest-frame UV emission of a galaxy (see also Section 4.2). Tight correlation is observed also between the magnitudes and the stellar mass in all our model galaxies, except for GALACTICUS. Overall, the DEEP2-FF observations and the MULTIDARK-GALAXIES show a good overlap in the brighter, more star-forming and massive portion of any parameter space. All the model galaxies then extend further down in SFR, stellar mass and magnitudes.

Age does not correlate with SFR neither in the observations, nor in SAG and SAGE mocks. In GALACTICUS, we observe an anti-correlation between age and SFR, meaning that older galaxies are more star-forming, as expected. Age does seem to correlate with stellar mass in DEEP2-FF, however this feature is not reproduced by any of our model galaxies. DEEP2-FF galaxies show a bimodal distribution in age and stellar mass, with an older, less star-forming, very massive population ($age \gtrsim 10^{9.3}$ yr; $M_\star \gtrsim 10^{10.3}$ M$_\odot$) and a younger, more star-forming distribution with less massive galaxies. None of the model galaxies seem to reproduce this bimodality.

SAG and SAGE stellar masses are tightly correlated with their SFRs, but no dependence is observed in GALACTICUS. While the DEEP2-FF quenched population is too sparse to identify any dependence in the stellar mass–SFR plane, the star-forming selection might show some correlation in the higher-mass end of the distribution. However, as already mentioned above, in order to correctly quantify this correlation, we should take into account the incompleteness effects in the data set, but this calculation goes beyond the aim
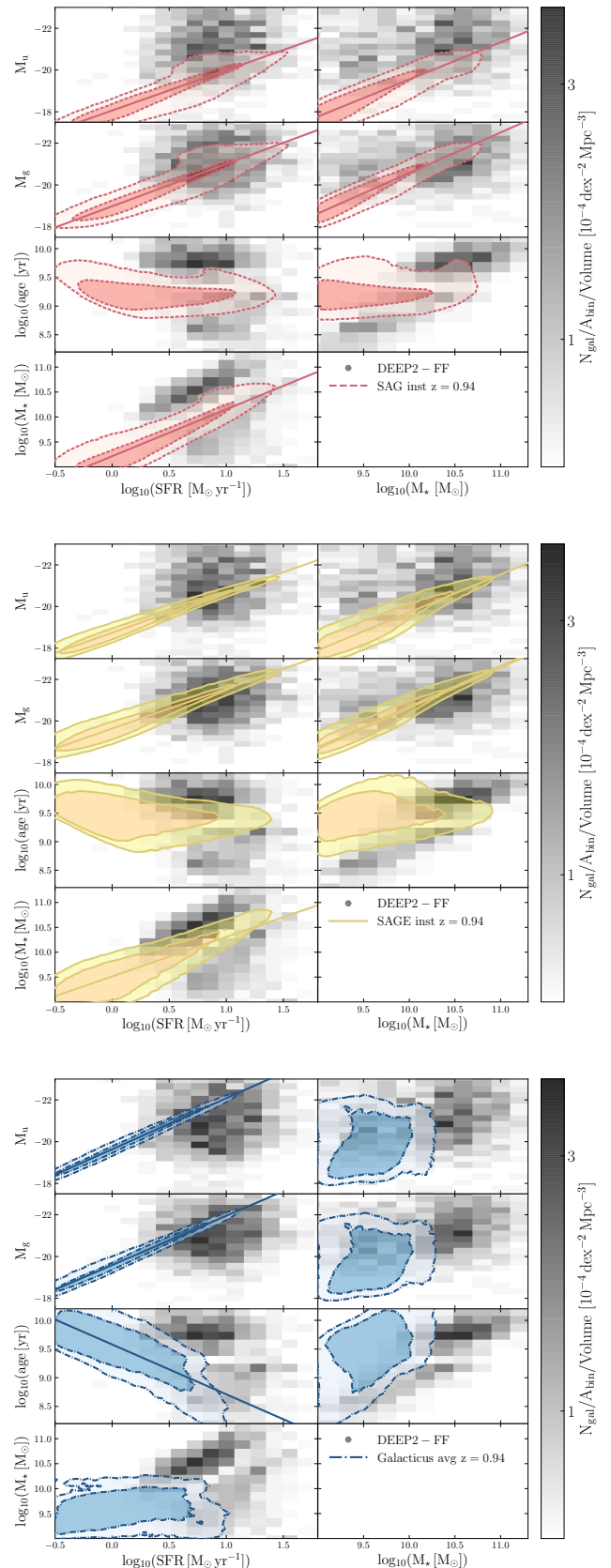
Figure C1. Comparison of pairs of properties for MULTIDARK-GALAXIES at $z = 1$ (contours) and DEEP2–FF observations at $0.9 < z < 1.1$ (grey, shaded squares). The colour bars show the number density of DEEP2 galaxies in each square. From top to bottom, we display SAG, SAGE and GALACTICUS results. A minimum [O II] flux cut of 5×10^{-18} erg s^{-1} cm^{-2} has been applied to both data and model galaxies. In each set of panels, from top to bottom, we compare broad-band u and g absolute magnitudes, age and stellar mass as a function of, from left to right, SFR and stellar mass. The model contours, from inner to outer, represent 68% and 95% of the distributions. The diagonal lines are the linear fits showing the significant correlations, whose coefficients are given in Table 2.

z=1		SAG	SAGE	Galacticus
$M_u = A\log_{10}(\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1}))+B$	A	-1.934±0.001	-1.941±0.001	-2.363 ±0.001
	B	-18.06±0.01	-18.74±0.01	-19.59±0.01
	$\sigma_{\log(\mathrm{SFR})}$	0.50	0.53	0.48
	$\sigma_{\mathrm{M_u}}$	1.07	1.05	1.15
	r	0.90	0.97	0.99
$M_g = A\log_{10}(\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1}))+B$	A	-2.029±0.001	-1.916±0.001	-2.362±0.001
	B	-18.98±0.01	-19.63±0.01	-19.59±0.01
	$\sigma_{\log(\mathrm{SFR})}$	0.50	0.53	0.48
	$\sigma_{\mathrm{M_g}}$	1.11	1.08	1.15
	r	0.91	0.94	0.99
$\log_{10}(\mathrm{age/yr}) = A\log_{10}(\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1}))+B$	A	—	—	-0.869±0.002
	B	—	—	9.58±0.1
	$\sigma_{\log(\mathrm{SFR})}$	—	—	0.48
	σ_{age}	—	—	0.54
	r	-0.21	-0.34	-0.77
$\log_{10}(\mathrm{M}_\star/\mathrm{M}_\odot) = A\log_{10}(\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1}))+B$	A	0.939±0.001	0.794±0.001	—
	B	9.21±0.01	9.51±0.01	—
	$\sigma_{\log(\mathrm{SFR})}$	0.54	0.52	—
	$\sigma_{\log(\mathrm{M}_\star)}$	0.50	0.52	—
	r	0.87	0.81	0.18
$M_u = A\log_{10}(\mathrm{M}_\star/\mathrm{M}_\odot)+B$	A	-1.779±0.001	-1.820±0.002	—
	B	-1.75±0.02	-1.52±0.01	—
	$\sigma_{\log(\mathrm{M}_\star)}$	0.54	0.52	—
	$\sigma_{\mathrm{M_u}}$	1.07	1.05	—
	r	0.89	0.89	0.15
$M_g = A\log_{10}(\mathrm{M}_\star/\mathrm{M}_\odot)+B$	A	-1.941±0.001	-1.951±0.001	—
	B	-1.17±0.01	-1.15±0.02	—
	$\sigma_{\log(\mathrm{M}_\star)}$	0.54	1.08	—
	$\sigma_{\mathrm{M_g}}$	1.11	0.52	—
	r	0.94	0.94	0.18

Table C1. Best-fit parameters of the linear scaling relations found in MULTIDARK-GALAXIES at $z \sim 1$ and shown in Fig. C1. The parameter r is the correlation coefficient and σ_y is the scatter in the y-axis. The SFR values are instantaneous for SAG and average for SAGE and GALACTICUS. We highlight that we do not quantify the correlation in the DEEP2-FF sample, since this calculation would require accounting for all the observational incompleteness effects, which goes beyond the aim of this work.

of our analysis. We do not to show the dependence of the above quantities on metallicity since they do not correlate significantly in any of the model galaxies considered.