Characterisation of the stellar content of SDSS EELGs through self-consistent spectral modelling

Iris Breda, José M. Vilchez, Polychronis Papaderos, Leandro Cardoso, Ricardo O. Amorin, Antonio Arroyo-Polonio, Jorge Iglesias-Páramo, Carolina Kehrig, and Enrique Pérez-Montero

1 Instituto de Astrofísica de Andalucía, Glorieta de la Astronomía, s/n, 18008 Granada, Spain
2 Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa, OAL, Tapada da Ajuda, 1349-018 Lisboa, Portugal
3 Departamento de Física, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, 1749-016 Lisboa, Portugal
4 Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
5 Departamento de Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile
6 Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán 1305, La Serena, Chile

Received 1 December 2021 / Accepted 4 May 2022

ABSTRACT

Extreme emission line galaxies (EELGs) are a notable galaxy genus, ultimately being regarded as local prototypes of early galaxies at the cosmic noon. Robust characterisation of their stellar content, however, is hindered by the exceptionally high nebular emission present in their optical spectroscopic data. This study is dedicated into recovering the stellar properties of a sample of 414 EELGs as observed by the SDSS Survey. Such is achieved by means of the spectral synthesis code F-starlight, which self-consistently considers the stellar and nebular emission in an optical spectrum. Additionally, a comparative analysis was carried on, by further processing the EELGs sample with the purely stellar spectral synthesis code STARLIGHT, and by extending the analysis to a sample of 697 normal star-forming galaxies, expected to be less affected by nebular contribution. We find that, for both galaxy samples, stellar mass and mean age estimates by STARLIGHT are systematically biased towards higher values, and that an adequate determination of the physical and evolutionary properties of EELGs via spectral synthesis is only possible when nebular continuum emission is taken into account. Moreover, the differences between the two population synthesis codes can be ascribed to the degree of star-formation activity through the specific star-formation rate and the sum of the flux of the most prominent emission lines. As expected, on the basis of the theoretical framework, our results emphasise the importance of considering the nebular emission while performing spectral synthesis, even for galaxies hosting typical levels of star-formation activity.

Key words. galaxies: star formation – galaxies: starburst – galaxies: evolution

1. Introduction

Throughout the many decades of research across the extra-galactic domain, several thousand galaxy specimens have been discovered that display substantially elevated equivalent widths (EWs; between 100 and up to ~2000 Å) of particular emission lines concomitant of violent star-formation (SF) events, such as Hα, [OIII]5007, [OII] and Ly-α, the latter being found mainly at the high-z regime. These galaxies, designated as extreme emission line galaxies (EELGs), may also be classified as green pea galaxies (PGPs; e.g., Cardamone et al. 2009; Amorín et al. 2015) or as blueberries (Yang et al. 2016), depending on their redshift and colours. EELGs can also be included in additional classification schemes such as blue compact galaxies or blue compact dwarfs (BCGs/BCDs; e.g., Kunth & Sargent 1986; Cairós et al. 2001; Papaderos et al. 1996; Thuan & Martin 1981; Loose & Thuan 1986; Reverte et al. 2007), or ELdots (Bekki 2015). The latter galaxy groups, however, are defined solely according to photometric properties, and include galaxies that are not EELGs. In addition, Wolf-Rayet features (e.g., Schaerer et al. 1999; Amorín et al. 2012) and nebular HeII emitters (e.g., Kehrig et al. 2018; Fernández et al. 2022) are found among EELGs. These are considered well-preserved relics alluding to the high-z Universe, where, in contrast with what is observed in the nearby Universe, such extreme events were increasingly more prevalent. Chiefly in virtue of their faint surface brightness levels, the true nature of these galaxies remains a mystery, with numerous fundamental questions yet to be addressed.

Aside from conventionally being regarded as windows to cosmic history, emulating the conditions observed in high-z galaxies at the peak of SF, EELGs harbour a substantial number of young, massive star forming complexes which produce a considerable amount of photoionising radiation, successively being absorbed by the metal-poor gas existing in their surroundings (e.g., Ravindranath et al. 2020). Considering the abnormally intensive SF events and the ensuing high number of super-novae explosions, the gas content is expected to present an irregular, punctured distribution from where hard ionising radiation may escape (e.g., Fletcher et al. 2019; Bergvall et al. 2006; Izotov et al. 2016; Pérez-Montero et al. 2020), making these galaxies perfect laboratories for the exploration of Ly-continuum escape. An additionally important remark with respect to these atypical galaxies is that these are consensually accepted as key contributors to the reionisation of the Universe (e.g., Wise et al. 2014; Madau & Haardt 2015; Stanway et al. 2016; Naidu et al. 2022; Matthee et al. 2022), and as the prime building blocks of more massive, local galaxies (e.g., Press & Schechter 1974). EELGs are amongst the most chemically poor extra-galactic objects in the local Universe, in respect to both their stellar
and gas content (e.g., Rosa-González et al. 2007; Jaskot & Oey 2013; Amorín et al. 2010; Isobe et al. 2022; Kojima et al. 2021). Hereupon, a comprehensive study of this galaxy genus, specially at lower redshifts, offers a unique opportunity to examine in detail the specifics of pristine environments, and how in these conditions processes such as SF, chemical enrichment or mass growth, unfold.

Acknowledging how critical is the role that these galaxies play in the broader context of galaxy formation and evolution emphasises the importance of cultivating a holistic understanding of the subject. In this regard, there is an abundant number of studies that collectively aid to outline a coherent picture of the inherent properties of such galaxies across time, shedding some light upon the topic. A key detail for establishing a comprehensive and consistent framework relies on the assembly of a substantial number of exemplars across redshift. Predominantly through the exploration of deep narrow- or broadband photometry, several studies have contributed in this regard (e.g., Straughn et al. 2009; van der Wel et al. 2011; Amorín et al. 2010; Atek et al. 2011; Smit et al. 2014; Lumbrañas-Calle et al. 2022). However, the information that can be extracted from this data type is rather limited, with the ideal being to acquire spectroscopic information for an equivalently high number of galaxies. Pérez-Montero et al. (2021; hereafter EPM21) describes the assembly of such a sample. By scanning approximately one million galaxies in the SDSS-DR7 (Sánchez et al. 2010) with apparent magnitude brighter than 17.8 in the r band through the Automated Spectroscopic K-means-based (ASK) classification scheme, the authors identified approximately 2000 EELGs at 0 ≤ z ≤ 0.49 with associated SDSS spectroscopic data. The work here presented exploits the aforementioned galaxy sample, emerging as a complementary study.

The assembly of such samples has been instrumental for the assessment of the physical properties of these galaxies. For instance, several studies have significantly contributed on the characterisation of the gas content of EELGs (in addition to the review by Kunth & Östlin 2000, see Amorín et al. 2010, EPM21, and Amorín et al. 2015) where the analysis of high-quality, spectroscopic data through a variety of methods (e.g., direct t, method, strong line methods and the code HI-CHI-MISTRY, Pérez-Montero 2014) revealed low total oxygen abundances (7.3–7.7 ≤ log(O/H) ≤ 8.5–8.6) and equally low nitrogen-to-oxygen ratios (−1.8 ≤ log(N/O) ≤ −0.8). The prior study additionally concluded that these display significantly compact morphologies, sharing many characteristics with high-redshift analogues, and that these objects are likely to be in a transient phase of their early evolution, currently building a significant amount of their stellar mass.

As far as the gas kinematics in EELGs is concerned, previous research has demonstrated high turbulence and rapidly evolving scenarios. For instance, by investigating a sample of 22 EELGs at 1.3 < z < 2.3 through near-infrared spectroscopy, Maseda et al. (2014) concluded that, at these redshifts, these low-mass, low-metallicity systems display velocity dispersions of the order of 50 km s^{-1} which, to attain stability against SF, implies a gas fraction above two-thirds. The authors add that the ongoing star-formation activity is expected to endure solely for ~50 Myr, rapidly assembling stellar mass and decreasing the specific star-formation rate (sSRF) on a time-scale that is shorter than that estimated for the gas depletion (~100 Myr). Further evidence is provided by Bosch et al. (2019) where, upon inspection of GMOS-IFU data of one EELG, three individual kinematic components were identified; a rotating component that the authors correlate with a disk-like feature, a violent, extensive outflow and an intermediate component that is considered to be a turbulent mixing layer previously identified in local starburst galaxies. Further evidence is provided by Hogarth et al. (2020), who analysed high-dispersion long-slit data. These results once more highlight the underlying complex nature of these galaxy systems.

In respect to their stellar content, it is documented that a vast number of BCDs possess an old underlying stellar component, in addition to the young stellar complexes that recently emerged as a result of the ongoing, intensive SF activity (e.g., Papaderos et al. 1996; Schulte-Ladbeck et al. 1999; Cairós et al. 2002; Amorín et al. 2012). Despite the fact that there is not a substantial body of knowledge describing the stellar populations of EELGs through redshift, complementary studies have been conducted, being mostly based on spectro-photometric evolutionary synthesis such as spectral energy distribution (SED) fitting (Janowiecki et al. 2017; Cohn et al. 2018). It is critical however, to stand out the many considerable biases and degeneracies inherent to SED fitting techniques (see e.g., Walcher et al. 2011). In this regard, and with all their shortcomings (see Gomes & Papaderos 2017, for a review), population spectral synthesis (PSS) techniques are by far the most reliable method for the challenging task of accurately characterising stellar populations in galaxies. Nevertheless, as most of the available PSS codes are purely stellar, that is, they do not consider the nebular continuum contribution, these are not suited for the study of star-forming galaxies. For EELGs in particular, which register especially high nebular contributions, the use of conventional PSS tools would introduce significant biases, impeding the accurate estimation of star-formation and chemical enrichment histories (SFH and CEH, respectively). In this respect, FADO (Gomes & Papaderos 2017; Cardoso et al. 2019; Pappalardo et al. 2021), with its unique capability of self-consistently considering the nebular component whilst fitting the stellar templates to the observed optical spectrum, is the only PSS tool that is qualified to produce sensitive results (it was recently employed to characterise the stellar content of three GPGs in the work by Fernández et al. 2022).

The primary goal of the present study is to accurately characterise the stellar content of a representative sample of EELGs, by processing with FADO a subsample of approximately 400 galaxies from EPM21. The secondary goal is to evaluate how the estimated stellar properties, such as stellar mass and mean stellar age, vary when considering full self-consistent or purely stellar spectral synthesis tools. To reach this goal, extra runs with the purely stellar PSS code STARLIGHT (Cid et al. 2005) were performed. Finally, we aim to compare EELGs with normal SF galaxies. This supplementary analysis allows for the identification of the physical properties compelling the observed differences, and to quantify the extent of these differences in the two galaxy families. The article is organised as follows: in Sect. 2 we introduce the samples to be analysed and the adopted methodology, and in Sect. 3 we discuss the obtained results, simultaneously exploring the causes for the disagreement between the two PSS codes and the differences between EELGs and normal SF galaxies. Finally, Sect. 4 summarises the main results obtained in the course of this analysis.

2. Sample description and data analysis

The present work is subdivided into three stages, the first pertaining to isolating the best candidates from the original sample of EPM21, along with assembling a control sample of normal SF galaxies, the second referring to the spectral modelling of
the SDSS spectra, and finally correction of the obtained physical properties for aperture effects.

2.1. EELG sample

As briefly aforementioned, the galaxy sample here analysed was identified in EPM21 by employing the ASK classification scheme on the SDSS-DR7 data-base, composed by roughly \(10^6\) galaxies with apparent magnitude <17.8 in the \(r\) band (the reader is referred to EPM21 for additional details on sample acquisition and specifics on ASK classification). The sample was identified by selecting galaxies that display an \(EW(H/\beta) > 30\) Å and an \(EW([OIII]) > 100\) Å, yielding 1969 objects with a redshift distribution between 0 < \(z\) < 0.49. However, the execution of this study requires the fulfilment of certain additional conditions:

i. To have available SDSS photometric data. The photometric analysis provides the means of visually inspecting the data thus confirming (a) the galaxy morphology and (b) the galaxy fraction covered by the fibre. In addition, it permits to estimate the radial extent of the galaxy, essential to correct from aperture effects. From the original 1969 objects of the starting sample, 1740 contain photometric information.

ii. After visual inspection, cometary and tadpole galaxies to which the fibre did not incorporate a significant portion of the galaxy’s radial extent and highly star-forming nodes in spiral galaxies, galaxy mergers or low surface brightness (LSB) galaxies were excluded from the main sample, thereafter listing 1119 objects.

iii. By applying SEXTRACTOR (Bertin & Arnouts 1996) to the photometric \(r\) frames, aperture photometry was performed, permitting to recover total radial extent \(R_T\) (i.e., the galactocentric radius at which the surface brightness level equalizes the background), effective radius \(R_e\) and the galaxy’s flux fraction integrated within the SDSS aperture \(f_{ap}\). Comparison with the photometric information provided by the SDSS-DR12 Photometric catalogue (Alam et al. 2015) has revealed an overall good agreement except for 10 galaxies (<1%) where the latter provides significantly over-estimated sizes. To assure that the fibre does not encompass only a residual amount of the total galaxy light, galaxies with \(f_{ap} < 30\)\% were excluded, thus obtaining the main sample of 997 sources.

iv. Considering the study by Pappalardo et al. (2021), where it was established a consistent dependence of the quality of the fits on the signal-to-noise ratio \(S/N\), and the dependence of the inclusion/exclusion of the Balmer jump at 3646 Å (and Paschen at 8207 Å for galaxies at \(z \leq 0.12\)) on the resulting best-fitting model, it was selected a sub-sample of galaxies with \(S/N > 8\) (measured at the restframe emission-line-free pseudo-continuum between 6390 Å and 6490 Å) and a rest-frame wavelength starting at <3640 Å, listing 464 galaxies.

Figure 1 illustrates the distribution of the redshift and of some typically high EWs for the main sample, encompassing all the EELGs from EPM21 that satisfy the previously mentioned criteria (i.e., have available SDSS photometric data, are individual, compact galaxies and have \(f_{ap} > 30\)\%), and the subsample consisting of the EELGs with \(S/N_{[OIII]6390:6490} > 8\), and to which the Balmer jump is present in the observed spectrum. The first panel shows that the vast majority of the sampled galaxies have \(z \leq 0.2\), reaching the peak of the distribution at 0.05 < \(z\) < 0.1. With the exception of galaxies with \(z < 0.05\) and \(z > 0.2\), which are not well represented by the reduced, higher-\(S/N\) sample, both distributions are comparable. However, the redshift range is not sufficiently wide to conduct a sound exploration across time. The additional panels of Fig. 1 display the distribution of the EWs of several emission lines, namely, from left to right, the flux-weighted average of the \([OII]\) doublet at 3727 Å and 3729 Å, the \([OIII]\) at 5007 Å and \(H/\alpha\) at 6563 Å. It is apparent that the entire range of EWs is properly represented by the reduced, higher-\(S/N\) sample. For illustrative purposes, 24 EELGs were uniformly selected across redshift, portraying some of the typical morphologies of the sampled EELGs across time, as exemplified by Fig. 2. Inspection of this figure reveals that although many EELGs display compact, circular morphologies, a significant fraction thereof are elongated, as apparent from the overlay contours.

2.2. Control sample of normal SF galaxies

For completeness, and aiming for a comparison study, the analysis was extended to a sample of normal SF galaxies. After identifying the plate, MJD and fibre of the ASK galaxies that conform to the defined selection criteria, spectroscopic SDSS data were retrieved from the SDSS DAS User Fiber List Submission Form\(^1\). The selection was based on their BPT \([NII]/H\alpha\) versus \([OIII]/H\beta\) values (Baldwin et al. 1981), specifically by selecting the galaxies falling below the Kauffmann et al. (2003) demarcation line (corresponding to pure SF systems), with a \(S/N > 3\) in the most relevant emission lines (i.e., \([OIII]\), \(H\alpha\), \(H\beta\)) and within the redshift range of 0.04 \(\leq z \leq 0.1\).

Additionally, to secure the extraction of normal SF galaxies (i.e., excluding galaxies experiencing residual or violent SF events), the range intervals 10 \(\leq EW([OIII]_{[5007]}) \leq 100\) Å, 20 \(\leq EW(H/\alpha) < 100\) Å and 3 \(\leq EW(H/\beta) < 30\) Å were selected. Subsequently, the obtained sample was cross-matched with the

\(^1\) http://das.sdss.org/www/html/post_fibers.html
Fig. 2. True-colour SDSS images of 24 randomly selected EELGs at different redshifts. It is shown the RA and Dec, redshift and a scale-bar depicting 2 kpc. The black contours show the galaxy morphology by highlighting the $r$ band isophotes, the white circle portrays the estimated galactic radial extent and the red circle represents the SDSS aperture ($\Theta = 3''$). The galaxies are ordered from top to bottom and from left to right, in ascending redshift order.
ASK catalogue (Ascasibar & Sánchez 2011) and all galaxies with primary ASK class corresponding to EELGs were withdrawn. This approach resulted in the selection of 11,832 objects, from which a sample of 1,350 objects was extracted, by randomly selecting 150 galaxies from each of the 9 redshift bins. The EW distribution of the relevant emission lines in the final randomly selected sample was contrasted with the same distribution of the original sample of ~12k objects, thus confirming that the reduced sample accurately represents the extended sample of normal SF galaxies. Comparatively with the EELGs sample, objects with S/N > 8 within the wavelength range between 6390 Å and 6490 Å were selected, resulting in a final control sample of 777 galaxies.

2.3. Spectral modelling

After de-redshifting the SDSS spectra (resolution λ/Δλ ~ 2.5 at ~3800 Å and λ/Δλ ~ 3.6 at ~9000 Å), rebinning to a step of 1 Å (for STARLIGHT runs, it is recommended to provide the input spectrum de-redshifted and rebinned to 1 Å), and correcting for Galactic extinction (following the extinction curve by Cardelli et al. 1989), with R_V = 3.1, spectral modelling of both samples was performed by means of the PSS codes FADO and STARLIGHT (as an integrity check, supplementary runs for the EELG sample were carried out with FADO purely stellar mode, FADO_ST, that is, without considering the nebular contribution). As Balmer and Paschen jumps are prominent nebular features in SF galaxies (the latter being observed in SDSS galaxies at z ≤ 0.12), and considering that the purely stellar code STARLIGHT cannot account for such (inclusion of these features would result in inaccurate fits, translating into stellar mass overestimation for younger ages, see Cardoso et al. 2019), STARLIGHT runs were performed by not considering these spectral ranges (i.e., from 3670 Å to 8100 Å). As for FADO, the whole wavelength range was considered while adopting full-consistency mode, with the auxiliary FADO_NST runs being carried out within the fitting range adopted for STARLIGHT. The main stellar library was designed, comprising 152 simple stellar population (SSP) spectra from Bruzual & Charlot (2003), for 38 ages between 1 Myr and 13 Gyr for four stellar metallicities (0.05, 0.2, 0.4 and 1.0 Z⊙), referring to a Salpeter initial mass function and Padova 2000 tracks. The individual SSP libraries for each galaxy were derived from the main library, by removing a varying number of SSPs with corresponding ages that are older than the age of the Universe at the z of the galaxy under study.

In addition to the stellar properties pertaining to spectral modelling, FADO provides fluxes and EWs of a comprehensive array of emission lines. The obtained fluxes were corrected for the effects of intrinsic extinction using the Balmer decrement as measured by the Hα/Hβ ratio, following the prescription proposed by Vogt et al. (2013), Eqs. (A11) and (A12), with R_V = 4.5 and R_a,β = 2.86. Subsequently, star-formation rates (SFRs) and specific star-formation rates (sSFRs) were derived by following the prescription by Kennicutt et al. (2009):

\[
\text{SFR} \left(\frac{\text{M}_\odot}{\text{yr}}\right) = 5.5 \times 10^{-42} \text{L(Hα)} \text{ (ergs s}^{-2}),
\]

where L(Hα) represents the dust-corrected Hα luminosity and the sSFR is the SFR divided by the estimated stellar mass. Considering that this prescription has the underlying assumption of a constant dust-heating stellar population, which might substantially underestimate the true SFR in young stellar systems (Kennicutt et al. 2009; Weilbacher & Fritze-v. Alvensleben 2001), the additional prescription by Duarte et al. (2017, Eq. (6)) was used. Both are in excellent agreement, not introducing any substantial alteration in the forthcoming correlations. However, it is important to note that SFR calibrations predicated on the Hα luminosity enclose the assumptions of (a) a continuous SF at a constant SFR for a minimum period of 70–80 Myr, and (b) stellar populations of solar metallicity, which might not hold true. For instance, in the case that the stellar metallicity is strongly sub-solar, the Lyman continuum photon production rate per unit mass is by a factor 3–4 times higher than for solar-metallicity stellar populations. This would imply a strong overestimation (under-estimation) of SFRs (specific SFRs) based on Hα luminosities. In addition, throughout this work it is considered a null escape fraction of ionising photons. Accordingly, the estimated SFRs might be regarded as first-order estimates.

2.4. Aperture corrections

The main goal of the present study is to characterise a sample of a considerable number of EELGs through systematic examination of single-fibre spectroscopic data. To this end, it is imperative to extrapolate the derived quantities (aperture based) to quantities that reflect the whole radial extent of the sampled galaxies.

Aperture- and extinction-corrected Hα luminosities and SFRs were obtained following the prescription by Duarte et al. (2017), where it were derived aperture corrections for approximately 210,000 SDSS star-forming galaxies based on empirical Hα growth curves extracted from the Calar Alto Legacy Integral Field Area (CALIFA) survey (Sánchez et al. 2012).

Concerning total stellar mass estimates, bearing in mind that spectroscopic information outside of the aperture is not available, a flawless correction for aperture effects is an unattainable concept. However, it is possible to obtain approximate estimates by assuming that these objects exhibit a constant mass-to-light (M/L) ratio, which is a reasonably valid assumption, considering that most of these objects are compact and uniform. In any case, it should be noted that, if the SF component (low-M/L ratio) is embedded within a more extended, older (higher-M/L ratio) stellar host, the total stellar mass would be underestimated, while the sSFR would be strongly overestimated. In addition, postulating a constant M/L ratio denotes that the previously estimated mean stellar age and metallicities within the fibre are similar throughout the whole galaxy extent.

The fair assumption that the surface brightness profile (µ; mag/2^"^) obtained in the spectral region where emission lines are absent or display a minimal contribution is a suitable proxy of the stellar surface density (Σ; i.e., the stellar mass M_⋆, over the projected area, M_⋆/πR_e^2) provided the means to extrapolate outwards the aperture the stellar mass previously derived through spectral modelling of the SDSS spectra. Summarising the complete procedure:

i. For each galaxy, µ for the three SDSS bands g, r and i were extracted from the photometric frames (after sky modelling and subtraction) by fitting ellipses to the galactic isophotes.

ii. As expected, the SDSS filters cover different intervals of the rest-frame spectrum according to the redshift of the source. Such can amplify or diminish colour gradients between the SF component and the surrounding LSB host, depending on redshift and the colour considered (Papaderos & Ostlin 2012). For instance, for galaxies at z = 0, the i band covers a spectral region which encompasses minimal nebular contribution, as this band-pass in the most sensitive to the stellar surface density profile of the galaxy. At z = 0.12 however, the g band provides a better tracer of the stellar mass, given that the strong [OIII]λ5007 falls in
the lower-transmission edge of that filter. With these considerations, the appointed $\mu \equiv \Sigma_*$ was extracted from the $i$ band for galaxies with $z \leq 0.067$, from the $r$ band for $0.067 < z \leq 0.105$ and from the $g$ band for $z < 0.105$.

iii. The $\mu$ were converted to linear units of counts/"$''$ ($\mu_{\text{int}}$), spline interpolated to a constant finer step and integrated within the radius of the fibre, 1.5"' such as,

$$I_{1.5} = 2\pi \Sigma_* \int_0^{1.5} \mu_{\text{int}} R \, dR,$$

where $R$ is the galactocentric radius.

iv. The previously obtained $I_{1.5}$ over the stellar mass within the SDSS aperture $M_{*,1.5}$, provides the factor $f$ necessary to convert the integrated $\mu_{\text{int}}$ (i.e., the total number of counts in the specific filter) into $M_*$. As a result, the total stellar mass $M_{*,T}$ is obtained by integrating the entire $\mu_{\text{int}}$ (i.e., until $R_T$), and multiplying by $f$ in the following manner:

$$M_{*,T} = 2\pi f \Sigma_* \int_0^{R_T} \mu_{\text{int}} R \, dR.$$  

Application of this procedure to the stellar masses obtained with both FADO and STARLIGHT resulted in the estimated total stellar mass for the two PSS codes.

Figure 3 exhibits the relation between absolute magnitude in the different SDSS optical bands and the derived stellar mass. It illustrates that the aperture-corrected stellar mass versus the total absolute magnitude within an aperture with radius $R_T$ and the stellar mass versus the absolute magnitude within the SDSS aperture manifests an equivalent behaviour, supporting the soundness of the aforementioned approach. In addition, Fig. 4 displays the relations between aperture-corrected and non-corrected quantities, namely for the sSFR (left-hand-side panel), present-day stellar mass as obtained by FADO (central panel) and stellar surface density (right-hand-side panel), for the EELG sample. While the sSFR does not exhibit significant variation, the mass and $\Sigma_*$ display a clear shift towards higher masses and lower $\Sigma_*$, respectively. The observed behaviour is expected, considering that for these galaxies the stellar mass fraction outside the fibre is typically low, whereas their radial dimension extends quite beyond the fibre, with an average galactic radius of $\sim 3.1'' \pm 0.1$, in the SDSS bands $r, g$ and $i$.

Regarding the normal SF galaxies, in virtue of their considerable size and heterogeneous morphology and nature, no attempt was made to perform aperture corrections. Accordingly, the comparison study can only be carried out between non-corrected stellar properties, such as mass- and luminosity-weighted age and metallicity, and physical properties as estimated before aperture correction.

3. Spectroscopic analysis

3.1. Recovering fundamental correlations between galactic properties

The EELGs and normal SF galaxy samples were processed as described in the previous section, providing a great amount of relevant results for both the stellar and gas components. However, considering that the gas-phase of these EELGs was extensively studied in EPM21, the present study focuses on their stellar properties (mainly mean stellar age and stellar mass) and on how both PSS codes here probed recover these properties, while contrasting with the same obtained for normal SF galaxies.

Both samples were once more curtailed by rejecting poor fits, that is, the galaxies for which the reduced $\chi^2$ of the fit by both PSS codes is higher than 3. The above resulted in 414 EELGs and 697 normal SF galaxies. Inspection of Fig. 5 reveals that, when comparing both samples, greater differences are found for the properties as obtained by FADO. This can be seen mostly for the logarithm of the present-day stellar mass, where the great majority of EELGs display values between 8.5 and 9.5 whereas the bin 9.5 to 10.5 is mainly populated by normal SF galaxies. Mass- and luminosity-weighted mean stellar age also reflects similar behaviour, considering that a substantial fraction...
Fig. 4. Differences between aperture-corrected (x axis) and non-corrected quantities (y axis), such as sSFR (left-hand-side panel), present-day stellar mass as obtained by FADO (central panel) and stellar surface density (right-hand-side panel), for the EELG sample. Identity lines are represented in grey. As in Fig. 3, data points are coloured as specified by their redshift distribution.

Fig. 5. Series of histograms summarising the main results obtained for both samples (EELGs in blue and normal SF galaxies in green), as obtained by FADO (left column) and STARLIGHT (right column). From top to bottom, it displays the distribution of the present-day stellar mass, the mass-weighted mean stellar age, the luminosity-weighted mean stellar age, the mass-weighted mean stellar metallicity, and the luminosity-weighted mean stellar metallicity.

of EELGs are significantly younger, as expected. STARLIGHT, on the other hand, does not recover any fundamental difference between these two galaxy families. Bearing in mind that STARLIGHT is a purely stellar code, it is incapable of disentangling between the stellar and nebular contributions. As a result, the nebular continuum generated by photoionisation of the gas by young stellar populations, which has a higher contribution in the blue wavelength range, is regarded as old, high-M/L ratio stellar populations. This effect, first pointed out by Izotov et al. (2011), subsequently translates into over-estimating masses and mean stellar ages (see Cardoso et al. 2019, for a detailed analysis), as it is properly demonstrated by Fig. 5, where both galaxy families are, on average, mainly regarded by STARLIGHT as more massive, significantly older stellar systems.

Figure 6 clearly demonstrates the previously reported phenomenon: whereas FADO recovers a positive trend between mass-weighted mean stellar age and total stellar mass (although highly scattered) and surface stellar density, STARLIGHT’s mass-weighted age determinations appear to be saturated, being primarily concentrated at the high-mass locus (although to a lesser degree, the same behaviour is seen for luminosity-weighted mean age determinations). The fact that STARLIGHT is unable to recover these fundamental relations suggests that correct determination of the stellar properties in SF galaxies requires proper treatment of the nebular contribution, as supported by the theoretical framework. Equally important is the fact that the same behaviour occurs for normal SF galaxies (although to a lesser extent, considering that these galaxies display lower levels of nebular contamination), as can be appreciated upon inspection of Fig. 7. Once more, this piece of evidence suggests that purely stellar codes such as STARLIGHT can be severely affected by typical SF activity.

Additional evidence demonstrating the limited capability of a purely stellar code to soundly extract the physical properties of the stellar component of SF galaxies is provided by Figs. 8 and 9 where it is displayed the relation between the sSFR and the total stellar mass, stellar surface density and mean stellar age for the main EELG sample. Although STARLIGHT is capable of recovering the negative trends between sSFR and the total stellar mass and stellar surface density, the low-mass regime is significantly underpopulated, reflecting the tendency for the purely stellar code to provide enhanced mass estimates. This behaviour is also seen for the normal SF galaxies in Fig. 9, which, alternatively, shows non-aperture-corrected quantities. Regarding the correlation between mean stellar age and sSFR, STARLIGHT fits yield significantly high values for the majority of the sample, with both mass- and luminosity-weighted mean stellar age estimates displaying a severe saturation at the high end. Once
more, the same behaviour is observed for the normal SF sample. In opposition, FADO recovers clear negative correlations between the sSFR and the aforementioned stellar properties, in agreement with previous studies (e.g., Brinchmann et al. 2004; Noeske et al. 2007).

3.2. Inspecting the factors impacting the determination of the stellar properties

The following stage is to assess the observed differences between spectral synthesis quantities (namely, stellar mass and mean stellar age) and identify the physical properties accountable for such contrast. The results outlined in the previous section strongly suggest that not accounting for the nebular contribution when fitting spectra of SF galaxies produces substantial biases towards higher ages and masses, ultimately concealing fundamental relations between galaxy properties. The results obtained by employing FADO_ST support the preceding conjecture.

Figure 10 displays a comparison between the present-day stellar mass and mass-weighted stellar age determined by STARLIGHT (blue points), FADO in its full consistency mode (orange points), and FADO_ST (purple points). This comparison clearly reveals an overestimation of the two physical properties when not considering the nebular continuum, which is
more pronounced for the log$_{10}$ $M_\odot$ estimates by STARLIGHT. The right-hand side panels exhibit a positive correlation between the differences in estimated stellar mass and mass-weighted age when considering and not considering the nebular contribution and the logarithm of the flux resulting from the sum of the most prominent emission lines (namely, [OII] (λλ 3727, 3729), [OIII] (λλ 4363, 4959, 5007), Hβ, HeI (λλ 5876), [NII] (λλ 6548, 6584), Hα, and [SII] (λλ 6717, 6731)), a direct proxy for the nebular continuum level. The over-plotted linear regressions clearly highlight the interdependence between the tested parameters, demonstrating that the higher the nebular contamination, the greater the stellar mass and mass-weighted age overestimation will be when using purely stellar codes (similar behaviour is also observed for luminosity-weighted stellar age estimates). These findings are in overall agreement with the theoretical framework, as evidenced by Figs. 4 and 5 of Cardoso et al. (2019) where the authors perform a series of controlled tests on synthetic spectra with FADO and STARLIGHT. Herein they demonstrate that, during evolutionary phases with a high sSFR, translating to an $EW(H\beta) \geq 200$ Å (such as the sampled EELGs which display an average $EW(H\alpha)$ of $\sim$400 Å), STARLIGHT tends to severely overestimate both the stellar mass and the mass-weighted stellar age estimates, whilst simultaneously strongly underestimating the light-weighted age and metallicity. FADO, on the other hand, is able to recover the mean stellar properties (i.e., total mass, mean age, and mean metallicity) with high accuracy (~0.2 dex). In addition, the theoretical work by Pappalardo et al. (2021) demonstrates that the overestimation of mass-weighted quantities observed in purely stellar codes such as STARLIGHT and STECKMAP (Ocvirk 2011) relates to the inability of purely stellar models to fit the observed SED in the vicinity of the Balmer jump.

There is, however, a non-negligible disagreement between the difference in the mass-weighted stellar age as obtained by FADO$_{ST}$ and FADO, versus STARLIGHT and FADO. Most likely, such reflects the significantly different optimisation algorithms employed by the two PSS codes (genetic differential evolution optimisation in the case of FADO and Markov chain Monte Carlo for STARLIGHT), which translates into a complex behaviour enrolling several other factors, such as how the code engages the age–metallicity–extinction degeneracy (e.g., the PSS code might try to compensate for the nebular contribution not by selecting older SSPs but by increasing the stellar metallicity or intrinsic extinction). It is likewise impracticable to accurately predict in what manner a purely stellar code such as STARLIGHT will fit the absorption stellar spectrum after being diluted by the nebular continuum. In any case, accurate evaluation of the role of the inclusion of the nebular contribution relative to the technical specifics of the fitting procedure (e.g., construction of the SSP library, the different mathematical recipes of the codes, in what manner the codes manage the age–metallicity–extinction degeneracy) requires an extensive series of tests on synthetic spectra, which on one hand is beyond the scope of this study and on the other was previously assayed, as aforementioned.

To note that there are additional problems which propagate in a non-linear manner, such as the excess at 1 Gyr which can be identified for both FADO modes, being more prominent for FADO in its full consistency mode. This effect, a known issue in spectral synthesis for which no solution currently exists (see Asari et al. 2007, and references therein) also prevents a direct and clear comparison.

Furthermore, in view of the fact that both codes handle the emission lines in an entirely different fashion, an additional factor that could impact stellar-property determinations, and consequently affect the registered differences as obtained by the two spectral synthesis codes, is the treatment of the emission lines (i.e., in case of STARLIGHT the emission lines must be masked, whereas FADO exploits some of them, in particular Hα and Hβ
in addition to the Balmer and Paschen jumps, to constrain of the nebular continuum contribution). However, both codes use the same set of emission lines, rendering it highly improbable that the origin of the discrepancies is the differences in the masking. Additionally, both codes auto-reject (faint) emission lines that might not be listed in the spectral mask. In the case where an emission line contains kinematically different components, or when an emission line comprises both a narrow and a broad (low-intensity) component, a single Gaussian fitting can be an issue, resulting in an inaccurate estimate of the stellar emission. In such circumstances, considering that the non-detection of the broad Hα and Hβ components would lead to a slight underestimation of the line flux, fitting a single Gaussian could underestimate the Hα flux. Therefore, possible imperfections in the single-Gaussian fitting of lines cannot significantly account for the differences between STARLIGHT, FADO and FADO²ST.

Complementary, Fig. 11 demonstrates that the lower the mean stellar age estimated by FADO, the higher the difference between both age and mass estimates, as obtained by both PSS codes. Not surprisingly, in light of the fact that younger (older) galaxies are expected to be less (more) massive, and as demonstrated by Figs. 8 and 9, the mean stellar age derived by FADO is anti-correlated with the sSFR, and furthermore, the mean stellar age estimated by FADO is also related to the emission flux, that is, the sum of the most prominent emission lines. Figure 12 documents these interrelationships: the left panel demonstrates how both galaxy families can be identified in the parameter space defined by the logarithm of the luminosity-weighted mean stellar age versus the logarithm of the sSFR, as given by FADO. Here two distinct linear relations are evident, displaced by a factor of ~1 dex, however conserving approximately the same slope (~0.62 for the EELGs and ~0.53 for the normal SF galaxies).

By employing this equation to predict the ages, given the values for the two additional quantities, and computing the differences between the predicted and observed ages, it is obtained a reasonable agreement, with a mean error of ~0.02 and a RMS of 0.29 dex. The outcome from this simple exercise corroborates our finding of the critical role of the nebular emission in the sound determination of stellar properties derived from spectral synthesis, simultaneously acting as a prognostic instrument: on the basis of their sSFR and the level of their emission flux, the mean stellar age of SF galaxies can be approximately inferred. Additionally, by considering the relations displayed in Fig. 11, with these two physical properties it is possible to provide a rough estimate of the extent to which mass and age estimates from pure stellar codes such as STARLIGHT should be adjusted to account for the nebular contribution.

Categorically, the series of tests herein, in addition to the theoretical modelling by Cardoso et al. (2019) and Pappalardo et al. (2021) soundly point to the nebular continuum contribution as the main reason for the observed discrepancies between
Fig. 12. Luminosity-weighted mean stellar age versus the logarithm of the sSFR for both galaxy samples (normal SF and EELGs), colour coded according to the different galaxy family (left panel) and according to the logarithm of sum of all the most prominent emission line fluxes (middle panel, units of $10^{-16}$ erg cm$^{-2}$ s$^{-1}$). The right panel illustrates the inter-connectivity of these three physical properties in a 3D parameter space, additionally displaying the resulting 3D surface that was fitted to the data.

Table 1. Table summarising the obtained mean values for both galaxy families (E for EELGs and nSF for normal SF galaxies) in four log of stellar mass bins, comprising information on, from top to bottom, number of galaxies, luminosity- and mass-weighted mean stellar age, and stellar surface density.

| Log $M_\ast$ ($M_\odot$) (6.5, 7.5] (7.5, 8.5] (8.5, 9.5] (9.5, 10.5] | n$^\circ$ of E | 0 | 36 | 322 | 56 |
|---------------------------------------------------------------|---------------|----------------|---------------|----------------|---------------|----------------|---------------|----------------|---------------|----------------|
| n$^\circ$ of nSF                                             | 1 | 28 | 491 | 176 |
| log $\bar{t}_L$ E                                           | – | 7.71 | 8.56 | 8.88 |
| log $\bar{t}_M$ E                                           | – | 8.32 | 9.23 | 9.72 |
| log $\bar{t}_L$ nSF                                         | 7 | 8.31 | 9.05 | 9.28 |
| log $\bar{t}_M$ nSF                                         | 7 | 8.66 | 9.53 | 9.85 |
| log $\bar{\Sigma}_\ast$ E                                  | – | 7.11 | 7.76 | 8.23 |
| log $\bar{\Sigma}_\ast$ nSF                                | 6.74 | 7.44 | 8.10 | 8.55 |

STARLIGHT and FADO, in addition to the non-negligible effect of the individual mathematical recipes of the two PSS codes.

3.3. Comparison study between EELGs and normal SF galaxies

The stellar properties of 414 EELGs and 697 normal SF galaxies were retrieved and compared. Generally we find that, for the same mass range, EELGs display, on average, systematically (but not substantially) lower mass- and luminosity-weighted ages. Unsurprisingly, however, the number of galaxies in each bin significantly varies depending on the galaxy family. Concerning the stellar surface density, in comparison to normal SF galaxies, EELGs are marginally less compact. These results are summarised in Table 1.

The top row of Fig. 13 evidences the contrast between EELGs and normal SF galaxies as obtained by processing the SDSS spectra with FADO, with respect to their compactness and mean stellar age (luminosity-weighted at the left-hand side and mass-weighted at the right-hand side). Although there is a significant overlap, a linear fit to these relations reveals that EELGs display higher slopes. Nevertheless, such behaviour may not reflect a true disagreement between the nature of these two galaxy families, but instead the under-representativeness of normal SF galaxies with low mean ages and stellar surface densities. The left-hand side (right-hand side) of the middle row of the same figure reveals that, although there is a considerable displacement (i.e., for the same stellar mass or stellar surface density, in comparison to normal SF galaxies, EELGs are marginally less compact).
density, EELGs yield sSFRs ∼1 dex higher), both galaxy families display similar slopes in the relation between stellar mass (stellar surface density) and the logarithm of sSFR. The same behaviour is visible in the lower panels, which display the relations between mean stellar age (luminosity- and mass-weighted) and the logarithm of the sSFR.

In summary, comparison between EELGs and normal SF galaxies reveals that, despite the fact that EELGs consistently yield significant higher values of sSFR for the same stellar mass, no significant difference is seen in the mean ages of the galaxies of the two samples (the mean age of the EELGs is lower on average, but not substantially lower). This might occur due to the presence of a pronounced underlying old stellar component: despite the high values of sSFR, the contribution of the young stellar component is significantly lower than the same of the old. It might additionally reflect different SFHs: generally, normal SF galaxies may display more continuous SFHs, in contrast to EELGs which suddenly endure an extensive but brief star-formation episode. Detailed studies of high signal-to-noise-ratio integral field spectroscopy (IFS) data (which allows to spatially resolve the galaxy, thus revealing individual spectra of different morphological regions) with FADO might help to further disclose the nature of these galaxies and how they compare with normal SF galaxies.

4. Summary and conclusions

The nebular emission pertaining to young stellar populations that inhabit SF galaxies contaminates optical spectroscopic data, impeding the thorough characterisation of the stellar properties of the latter. The robust extraction of stellar properties in SF systems is only possible through an approach that self-consistently considers the nebular contribution, such as the one taken by FADO.

In this work, we attempt to quantify differences in the stellar properties obtained using a purely stellar PSS code such as STARLIGHT, versus FADO, which accounts for the nebular emission, for two set of galaxies: a sample of approximately 400 EELGs and a supplementary sample of approximately 700 normal SF galaxies. We confirm that familiar fundamental relations, such as stellar age versus stellar mass and stellar surface density, and sSFR versus stellar mass and stellar surface density and stellar age, can only be accurately inferred when accounting for the nebular contribution, yet again reflecting the importance of considering the nebular emission when characterising the stellar content of SF galaxies. This outcome is valid not only for EELGs but also for normal SF galaxies, demonstrating that even a mild nebular contribution is sufficient to substantially confuse purely stellar PSS codes, which systematically tend to retrieve higher stellar ages and masses. As anticipated, further exploration of the data links revealed that the two parameters that appear to regulate the observed differences between the stellar properties recovered through spectral synthesis are the sSFR and the emission flux, that is, the sum of the flux of the more prominent emission lines.

Regarding the differences observed between EELGs and normal SF galaxies, the low-mass, low-stellar surface density locus of normal SF is underpopulated as compared to the EELG sample. For this region of the parameter space in particular, EELGs are significantly (∼0.6 dex) younger than their normal SF counterparts. For higher mass bins, the stellar age differences are not as relevant as for the lower mass bins, despite the significantly higher sSFR. Such might be due to different SFHs: the recent star-forming activity in EELGs probably translates into a peaky SFH, while normal SF galaxies might endure SFHs of more continuous nature. High-quality IFS data, allowing to resolve stellar populations in time and space, might provide further insight in the near future.

Acknowledgements. We thank the anonymous referee for valuable comments and suggestions. I.B., J.V.M., J.I.P., C.K., E.P.M and A.A.P. acknowledge financial support from the Spanish MICU through the “Center of Excellence Severo Ochoa” award to the Instituto de Astrofísica de Andalucía (SEV-2017-0709). J.V.M., J.I.P., C.K., and E.P.M. acknowledge financial support from projects Estallidos6 AYA2016-79724-C4 (Spanish Ministerio de Economía y Competitividad), Estallidos? PID2019-107408GB-C44 (Spanish Ministerio de Ciencia e Innovación); and grant P18-FR-2666 (Junta de Andalucía). R.A. acknowledges support from ANID Fondecyt Regular 1202007. P.P. thanks Fundação para uma Ciência e a Tecnologia (FCT) for managing research funds graciously provided to Portugal by the EU and was supported through FCT grants UID/FIS/04434/2019, UIDB/04434/2020, UIDP/04434/2020 and the project “Identifying the Earliest Supermassive Black Holes with ALMA (ideeS with ALMA)” (PTDC/FIS-AST/29245/2017). This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

Kunth, D., & Ostlin, G. 2000, A&ARv, 10, 1