

# Extreme Emission-Line Galaxies in SDSS. I. Empirical and model-based calibrations of chemical abundances

E. Pérez-Montero<sup>1</sup>★, R. Amorín<sup>2,3</sup>, J. Sánchez Almeida<sup>4,5</sup>, J.M. Vílchez<sup>1</sup>,  
R. García-Benito<sup>1</sup>, C. Kehrig<sup>1</sup>

<sup>1</sup>*Instituto de Astrofísica de Andalucía. CSIC. Apartado de correos 3004. 18080, Granada, Spain.*

<sup>2</sup>*Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raul Bitrán 1305, La Serena, Chile.*

<sup>3</sup>*Departamento de Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile.*

<sup>4</sup>*Instituto de Astrofísica de Canarias. C/ Vía Lactea s/n. La Laguna, Tenerife, Spain.*

<sup>5</sup>*Departamento de Astrofísica, Universidad de La Laguna, Tenerife, Spain.*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

Local star-forming galaxies show properties that are thought to differ from galaxies in the early Universe. Among them, the ionizing stellar populations and the gas geometry make the recipes designed to derive chemical abundances from nebular emission lines to differ from those calibrated in the Local Universe. A sample of 1969 Extreme Emission Line Galaxies (EELGs) at a redshift  $0 \lesssim z \lesssim 0.49$ , selected from the *Sloan Digital Sky Survey* (SDSS) to be local analogues of high-redshift galaxies, was used to analyze their most prominent emission lines and to derive total oxygen abundances and nitrogen-to-oxygen ratios following the direct method in the ranges  $7.7 < 12+\log(\text{O}/\text{H}) < 8.6$  and  $-1.8 < \log(\text{N}/\text{O}) < -0.8$ . They allow us to obtain new empirically calibrated strong-line methods and to evaluate other recipes based on photoionization models that can be later used for a chemical analysis of actively star-forming galaxies in very early stages of galaxy evolution. Our new relations are in agreement with others found for smaller samples of objects at higher redshifts. When compared with other relations calibrated in the local Universe, they differ when the employed strong-line ratio depends on the hardness of the ionizing radiation, such as O32 or Ne3O2, but they do not when the main dependence is on the ionization parameter, such as S23. In the case of strong-line ratios depending on [N II] lines, the derivation of O/H becomes very uncertain due to the very high N/O values derived in this sample, above all in the low-metallicity regime. Finally, we adapt the bayesian-like code HII-CHI-MISTRY for the conditions found in this kind of galaxies and we prove that it can be used to derive within errors both O/H and N/O, in consistency with the direct method.

**Key words:** ISM: abundances, galaxies: ISM, abundances, evolution, star formation

## 1 INTRODUCTION

The metal content of the gas in galaxies at different cosmological epochs is one of the main indicators of their evolution. There are important scaling relations in galaxies between metallicity ( $Z$ ) and other integrated properties such as stellar mass ( $M_*$ ), known as the mass-metallicity relation (MZR, Lequeux et al. 1979; Tremonti et al. 2004), or between them in combination with the star formation rate (SFR), in the known as fundamental metallicity relation (FMR, Lara-López et al. 2010; Mannucci et al. 2010). The evolution of these relations with cosmological time (e.g. Lamareille et al. 2009; Cresci et al. 2010; Pérez-Montero et al. 2013; Zahid et al. 2013;

Maier et al. 2014; Kashino et al. 2017) give important observational constraints to the models of galaxy formation and evolution.

However, one of the main obstacles to study the evolution of these scaling relations is that star-forming galaxies at intermediate and high redshift ( $z$ ) present very different properties compared with the well-studied sample of local star-forming regions in disk galaxies. This implies that the methods and the calibrations traditionally used to derive chemical abundances in low- $z$  objects must be adapted to the conditions of the high- $z$  samples.

Among the observations pointing out these differences, a direct correlation between  $Z$  and the size of galaxies at fixed stellar mass has been found, in the sense that smaller radii implies larger  $Z$  (e.g. Hoopes et al. 2007; Ellison et al. 2008; Brisbin & Harwit 2012). This has implications on the expected metal content of high-

★ E-mail: epm@iaa.es (EPM)

$z$  objects as it has been observed that these have in average smaller radii as compared to their low- $z$  analogs for the same stellar mass (e.g. Daddi et al. 2005). According to simulations, this relation can be pinned down to the relative time when each galaxy underwent the main episode of gas accretion (Sánchez Almeida & Dalla Vecchia 2018).

The physical parameters characterizing the nebular emission vary with redshift. This is the case of the ionization parameter ( $U$ ), that is known from different local samples of H II regions and galaxies to be lower for higher metal content (e.g. Dopita et al. 2000; Pérez-Montero 2014). This dependence can lead to an enhancement of systematic uncertainties in the derivation of  $Z$  and, hence, to the dispersion in the  $MZR$  found at higher  $z$ , when the mean  $U$  of galaxies is higher (Nakajima & Ouchi 2014; Shapley et al. 2005; Shapley et al. 2015). High- $z$  objects also present higher average electron densities (e.g. Brinchmann et al. 2008; Shirazi et al. 2014; Kaasinen et al. 2017). Moreover, the radiation field shows a well defined sequence with redshift in which emission lines in higher- $z$  objects are produced in harder ionizing conditions (Kewley et al. 2013).

In addition, high- $z$  galaxies can have different chemical evolution histories leading to abundance ratios different from the local sample. This is the case for the nitrogen-to-oxygen ratio (N/O), for which several authors have found to be higher at a given metallicity (e.g. Masters et al. 2014; Hayashi et al. 2015; Kojima et al. 2017), although there is no general consensus on its evolution (e.g. Shapley et al. 2015; Steidel et al. 2016; Strom et al. 2018; Sanders et al. 2020). The difference in behavior could be due to a different star formation efficiency (Khochfar & Silk 2011) or the existence of metal-poor gas inflows thought to be responsible for the low O/H and relatively high N/O values observed in Green Pea galaxies (Amorín et al. 2010, 2012) and Lyman-break analogues (Loaiza-Agudelo et al. 2020). This observed difference can have a non-negligible impact on the determination of  $Z$  using [N II] emission-lines (Pérez-Montero & Contini 2009).

Gas-phase  $Z$  in star-forming galaxies are mainly derived by interpreting optical emission-lines from ionized gas, which readily provide the oxygen abundance O/H. Thus,  $Z$  and O/H are used interchangeably in literature, as we do here. The scale and the uncertainty of the resulting O/H, depend basically on the detected collisionally excited lines and the method used to calibrate the relation between the emission-line fluxes and the metal content of the gas. In the absence of any metal recombination line, usually around  $10^{-4}$  times fainter than H $\beta$  and hence very difficult to be measured in high- $z$  galaxies, the most reliable method comes from the determination of the electron temperature and the use of the so-called direct method (e.g. Pérez-Montero 2017). This method depends, however, on the detection of weak auroral lines, such as for instance [O III]  $\lambda 4363$  Å, difficult to detect in weak or metal-rich objects. In case these lines are too faint, other methods based on more easily observable stronger lines are calibrated to estimate O/H (e.g. Pagel et al. 1979; Pérez-Montero & Díaz 2005; Maiolino & Mannucci 2019).

In the last years, several attempts have been made to enlarge the sample of high- $z$  galaxies with *bona-fide* determination of oxygen abundance using the direct method to establish calibrations consistent with the high- $z$  regime, either using stacking of spectra of local analogs to high- $z$  star-forming galaxies (e.g. Bian et al. 2018), or with the direct acquisition of auroral lines in high- $z$  spectra (e.g. Jones et al. 2015; Sanders et al. 2020). Instead, we use in this work a sample of local objects selected from the Sloan Digital Sky Survey (SDSS) to carry out an extensive study of the chemical properties of

Extreme Emission Line Galaxies (EELG) considered to be analogs to high- $z$  objects.

In the context of providing a calibration for O/H based on strong lines, local EELGs present clear advantages with respect to high- $z$  galaxy sets. The number of available galaxies is much larger, and the signal-to-noise of the individual spectra often allows us to determine, simultaneously, their O/H and N/O using the direct method. In high- $z$  targets this is possible only using stacked spectra which may bias the results. Part of the O/H derivations in our sample of local EELGs are made with the code HII-CHI-MISTRY (hereafter HCM, Pérez-Montero 2014) to complement the empirical calibrations of strong-lines. This code is based on the use of photoionization models but it results consistent with the direct method and presents two important advantages for the study of high- $z$  galaxies: i) it allows us to select several different combinations of emission-lines, which is of special relevance for objects observed with the same instrumental setup at different  $z$  (e.g. Sanders et al. 2018), and ii) it allows us to use [N II] emission lines to determine N/O independently of O/H, reducing the uncertainty owing to the unknown O/H-N/O relation, as may occur in dense environments (Edmunds 1990; Köppen & Hensler 2005).

The paper is organized as follows: Section 2 presents our sample of selected EELGs including their properties. In Section 3, we describe the derivation of physical properties and chemical abundances following the direct method in those galaxies with a reliable measurement of [O III]  $\lambda 4363$  Å. In Section 4, we provide different empirical calibrations based on strong collisionally excited emission-lines using the abundances derived from the direct method or from the direct calibration of the oxygen abundance with the electron temperature. In Section 5, we extend our analysis of the chemical abundances to the use of the code HCM, using appropriate photoionization models to find a solution for O/H, N/O and the ionization parameter consistent with the direct method. Finally, in Section 6 we summarize our results and present our conclusions.

## 2 DESCRIPTION OF THE EELG SAMPLE

Our EELG sample selection is based on the Automated Spectroscopic  $K$ -means-based (ASK) classification of galaxy spectra (about one-million galaxies with apparent magnitude brighter than 17.8) in the SDSS-DR7 (Sánchez Almeida et al. 2010). The  $k$ -means clustering algorithm used for the ASK classification is a well-known robust tool, which is employed in data mining, machine learning and artificial intelligence (e.g. Everitt 1995; Bishop 2006) and guarantees that similar rest-frame spectra belong to the same class. A thorough description and technical details regarding the ASK classification is provided in Sánchez Almeida et al. (2010) and additional properties of the ASK classes inferred from them can be found in a series of papers (Sánchez Almeida et al. 2011; Ascasibar & Sánchez Almeida 2011; Huertas-Company et al. 2011; Aguerri et al. 2012; Sánchez Almeida et al. 2012). Here, we only summarize their main properties in the specific context of our EELG selection. A more extended discussion on the ASK-based EELG selection will be provided in a forthcoming paper (Amorín et al, in prep.).

In brief, the ASK classification scheme allows to separate and characterize all the SDSS galaxy spectra into 28 ASK classes, which are driven only by the shape of the visible spectrum. As described in Sánchez Almeida et al. (2010), most ( $\sim 99\%$ ) galaxies in the SDSS-DR7 were classified into only 17 ASK major classes, with 11 additional minor classes including the remaining  $\sim 1\%$ . All the galaxies in a class have very similar spectra and their average spectrum is the

template spectrum of the class. Both the ASK classification of all galaxies in the SDSS-DR7 and their template spectra are publicly available<sup>12</sup> through the Spanish Virtual Observatory.

For the present study, we selected galaxies undergoing intense star formation, which are characterised by their high emission line equivalent widths, i.e.  $EW(H\beta) > 30\text{\AA}$  and  $EW([O\ III]) > 100\text{\AA}$  (e.g. Amorín et al. 2014, 2015; Calabrò et al. 2017). According to Sánchez Almeida et al. (2012), these are typically HII galaxies (e.g. Terlevich et al. 1991) which belong to a limited number of minor ASK classes, specifically ASK 15, ASK 17, ASK 20, ASK 21, ASK 25, ASK 26 and ASK 27, and a small number of objects (43) which are outliers of any of the ASK classes defined in Sánchez Almeida et al. (2010), i.e. they have nearly zero probability to belong to any class. Their ASK template spectra has two main components (i) a very faint and blue stellar continuum with extremely weak Balmer stellar absorption lines, and (ii) significant nebular continuum and very strong Balmer recombination and collisionally-excited emission lines. The qualitative physical properties of these galaxies based on their ASK template spectra were presented in Sánchez Almeida et al. (2012, Table 1). All of them are metal-poor starbursts dominated by young ( $< 10\text{ Myr}$ ) stellar populations.

The final sample collects all the galaxies in the ASK classes mentioned above, which all together yield 1969 EELGs with redshifts from  $z \sim 0$  to  $z \sim 0.49$ . Nonetheless, the redshift distribution of this sample is not homogeneous and the average  $z$  value is 0.08, with a median value of 0.05. A histogram of this distribution can be seen in upper left panel of Figure 1. The sample includes well-known families of local analogs of high- $z$  galaxies, such as Green Pea galaxies (Cardamone et al. 2009; Amorín et al. 2010, 2012) and their cohort (Izotov et al. 2011). After visual inspection, we discarded all duplicated objects within the selected sample. The EELG sample includes nearly integrated spectra from extremely compact galaxies such as GPs, and spectra of bright star-forming clumps within larger galaxies. For this work, we do not apply any aperture correction on the spectroscopic measurements.

We used the package `SHIFU`<sup>3</sup> to obtain the flux of the emission lines from the spectra taken from the Data Release 16 of SDSS (Ahumada et al. 2019). The package contains a suite of routines to easily analyze emission or absorption lines. The core of the code uses CIAO's Sherpa package (Freeman et al. 2001). Several custom algorithms are implemented in order to cope with general and ill-defined cases. A first order polynomial was chosen for the continuum, while single gaussians were selected for the lines.

Prior to the emission-line measurement, the spectra were shifted to the rest frame and corrected for Galactic extinction according to Schlegel et al. (1998). Then, we subtracted the underlying stellar population using the spectral synthesis code STARLIGHT (Cid Fernandes et al. 2004, 2005). STARLIGHT fits an observed continuum SED using a nonparametric linear combination of synthesis spectra of different single stellar populations (SSPs), simultaneously solving the ages, metallicities, and the average reddening. In this case, we considered the SSP spectra from Bruzual & Charlot (2003) based on the STELIB library from Le Borgne et al. (2003), Padova 1994 evolutionary tracks, and a Chabrier (2003) initial mass function (IMF) between 0.1 and  $100 M_{\odot}$ . We selected 41 ages spanning from 1 Myr up to 14 Gyr with four metallicities, from Z

= 0.0001 up to  $Z = 0.008$ . We used the reddening law from Cardelli et al. (1989) with  $RV = 3.1$ .

The emission-line fits are then performed in the stellar-continuum-subtracted spectra, allowing for the modeling of the continuum to take into account small deviations in the stellar continuum residuals. A sigma clipping was independently applied to the residual spectra, and then this was parsed to the composite line plus (residual) continuum model. Uncertainties in the measured values are evaluated by perturbing the residual spectra according to the error vector 100 times. Only line fluxes with a signal-to-noise ratio (SNR) larger or equal to 3 were used for our analysis. The continuum was evaluated in the original spectra to determine equivalent widths. In Figure 1 (a) we show the resulting histogram of the measured EW for  $[O\ III]$  at  $5007\text{\AA}$ . The mean value of the distribution is  $354\text{\AA}$ , with a median value of  $249\text{\AA}$ . These values are slightly higher when we analyze only the objects at a redshift larger than the 90th percentile of the redshift distribution ( $z = 0.194$ ) that are, respectively, of  $483\text{\AA}$  and  $377\text{\AA}$ .

In order to carry out our abundance diagnostic analysis, we corrected emission line fluxes from absorption of interstellar dust using the reddening constant  $c(H\beta)$ . Its value was obtained from the direct comparison between the observed  $H\alpha/H\beta$  ratio and the theoretical value for standard physical conditions of the gas ( $T_e = 10^4\text{ K}$  and  $n_e = 100\text{ cm}^{-3}$ ), and assuming the extinction curve of Cardelli et al. (1989).

We compared the emission-line properties and chemical abundances of our selected EELGs with a local sample of 59 547 star-forming galaxies, which is used as control sample. This sample was selected from the parent SDSS-DR7 sample limiting the redshift range to  $0.04 < z < 0.10$  to minimize aperture effects and ensure a S/N of at least 3 in all the relevant strong emission lines used in this work.

In Figure 2, we represent the  $[N\ II]/H\alpha$  vs.  $[O\ III]/H\beta$  diagram, (Baldwin et al. 1981), for both the control sample and the selected EELGs. As it can be seen, both samples are in regions of this diagram well consistent with emission-lines dominated by stellar photoionization (Kewley et al. 2001; Kauffmann et al. 2003). EELGs present a specific position in this diagram owing to their particular chemical and ionization properties as compared with the rest of star-forming galaxies. On the one hand, EELGs tend to have larger  $[O\ III]/H\beta$  ratios than the rest (i.e. the mean value of  $O3$  [ $= \log([O\ III]/H\beta)$ ] for EELGs is 0.62 while for the rest of star-forming galaxies is -0.28). On the other hand, the mean  $N2$  [ $= \log([N\ II]/H\alpha)$ ] for EELGs (-1.32) is lower than the control star-forming sample (-0.53). This difference is significant even if we restrict the star-forming control sample to the observed  $[O\ III]/H\beta$  range in EELGs (i.e.  $\pm 2\sigma$  the mean value in EELGs). This makes, as noted in the same figure, that the EELGs appear close to the separation curve between star-forming dominated galaxies and AGNs defined by Kauffmann et al. (2003) and present a very similar position in this diagram to other star-forming galaxies at high-redshift (e.g. Bian et al. 2020).

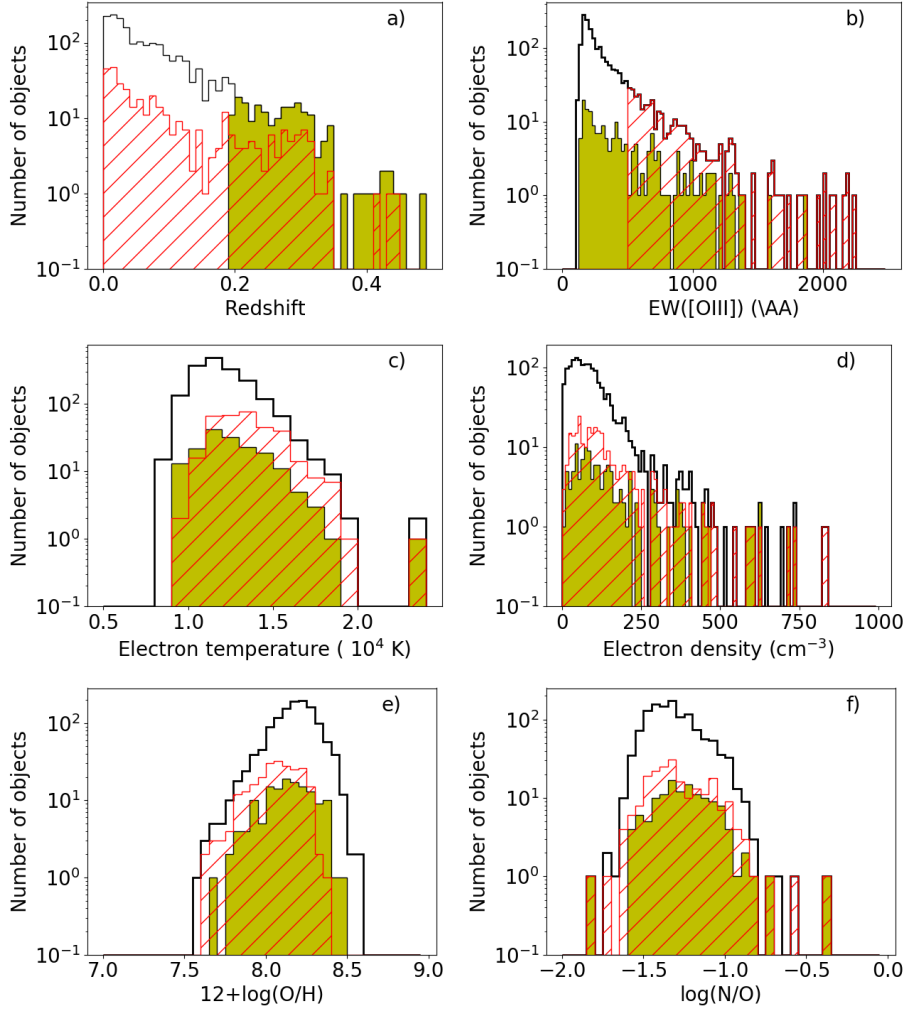
### 3 PHYSICAL PROPERTIES AND CHEMICAL ABUNDANCES FROM THE DIRECT METHOD

We derived ionic and total chemical abundances of the gas-phase in the selected sample of EELGs using the so-called direct method when the needed emission lines have a S/N larger than 3. To calculate all the physical properties and chemical abundances described below using these emission lines, we employed the analytical expres-

<sup>1</sup> <ftp://ask.galaxy@ftp.iac.es/>

<sup>2</sup> <https://sdc.cab.inta-csic.es/ask/index.jsp>

<sup>3</sup> Sherpa IFU line fitting package, (García-Benito in prep.).



**Figure 1.** Histogram of some of the measured and derived properties of the selected sample of EELGs, including (a) redshift, (b) equivalent width of [O III] 5007 Å, (c) electron temperature of [O III], (d) electron density of [S II], (e) total oxygen abundance as derived from the direct method, and (f) nitrogen-to-oxygen ratio, also derived from the direct method. The yellow histograms show the objects with a redshift larger than 0.194, i.e., larger than 90% of the galaxies in the sample. The red hatched histograms represent objects with  $\text{EW}([\text{O III}]) > 500 \text{ \AA}$ .

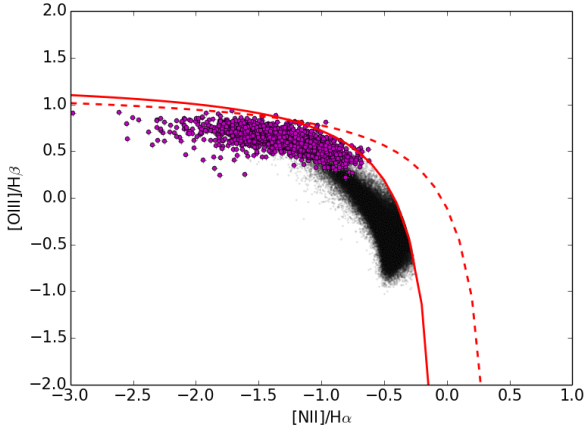
sions listed in Pérez-Montero (2014) and Pérez-Montero (2017), but taking care of using the same sets of atomic data used in the photoionization models described in the sections below. This implies using collisional atomic data from Kisielius et al. (2009) for  $\text{O}^+$ , and from Storey et al. (2014) for  $\text{O}^{2+}$ .

Briefly, the direct method is based on the determination of the electron temperature of the ionized gas using the [O III] emission-line ratio  $I(5007)/I(4363)$ . This ratio could be used for a subsample of 1813 EELGs in our sample, leading to a mean value of 12 100 K value, with a standard deviation of 1 800 K. Other electron temperatures corresponding to other ions can be measured as well, but the corresponding auroral line fluxes are usually weaker than that for of  $\text{O}^{2+}$ . We checked that none of these auroral lines were neither measurable in the 156 galaxies for which the [O III] auroral line was not measurable. In those galaxies with just one estimate of electron temperature, ionic abundances are usually calculated assuming a model-based relation between temperatures. In this work we use for all the galaxies with a direct measurement of  $t([\text{O III}])$ , the model-based relation between  $t(\text{O}^{2+})$  and  $t(\text{O}^+)$  given by Pérez-Montero (2014) to derive the electron temperatures associated to the low-

excitation regions necessary to calculate their corresponding ionic abundances.

When the temperatures associated to different ionization layers in the gas have been derived, it is possible to calculate the two main ions abundances ( $\text{O}^+$  using [O II]  $\lambda 3727 \text{ \AA}$  and  $\text{O}^{2+}$  using [O III]  $\lambda 5007 \text{ \AA}$ ). Then the total oxygen abundance is calculated as the addition of these two abundances relative to  $\text{H}^+$ . For very high ionization, it is expected that a fraction of oxygen can be ionized to  $\text{O}^{3+}$ , and the use of an ionization correction factor (ICF) is needed (e.g. Pérez-Montero et al 2020), but this is not expected since the high-excitation He II emission line at  $\lambda 4686 \text{ \AA}$  is not observed in most of the objects. Finally, the nitrogen-to-oxygen ratio is calculated by means of  $\text{N}^+/\text{H}^+$  (via the [N II]  $\lambda 6583 \text{ \AA}$  line) and assuming that  $\text{N}^+/\text{O}^+ = \text{N}/\text{O}$ . Both  $\text{O}^+$  and  $\text{N}^+$  have non-negligible dependence on electron density, which was calculated using the values derived from the [S II] emission-line ratio  $I(6731)/I(6717)$ . Our sample of EELGs shows gas density values with a mean around  $100 \text{ cm}^{-3}$  and a standard deviation of  $70 \text{ cm}^{-3}$ , and only in 3 objects we measure electron densities larger than  $500 \text{ cm}^{-3}$ .

The number of objects in our EELGs sample for which it is



**Figure 2.** Relation between the emission-line ratios  $[N\ II]/H\alpha$  and  $[O\ III]/H\beta$  for our sample of EELGs and for star-forming SDSS galaxies. The sample of selected EELGs are symbolized as magenta circles. The solid line represents the separation curve defined by Kauffmann et al. (2003) between star-forming and AGN-dominated galaxies, while the dashed line represents the theoretical separation curve defined by Kewley et al (2001).

possible to derive total oxygen abundances and nitrogen-to-oxygen ratios following the direct method using the available SDSS spectra are 1268 and 1240 respectively. These numbers are substantially smaller than the number of galaxies with estimated electron temperature, a drop caused by the fact that the  $[O\ II]\ \lambda 3727\ \text{\AA}$  emission-line cannot be detected in the most local objects (i.e.  $z < 0.02$ ) in the observed spectral range of SDSS. The results that we obtain are in general consistent with those of a family of metal-poor objects, with a mean total oxygen abundance from the direct method of  $12+\log(O/H) = 8.17$  and  $\log(N/O) = -1.31$ , i.e.  $\sim 30\%$  and  $\sim 38\%$  the solar ratios, respectively (Asplund et al. 2009). In Figure 1 we show the distribution of electron temperature as derived from  $[O\ III]$ , electron density from  $[S\ II]$  lines, and chemical abundances from the direct method as discussed above.

To assess the sample homogeneity, we also show in Figure 1 the distributions of the same properties for the subsample of EELG with a redshift larger than the 90th percentile of the  $z$  distribution in the whole sample (i.e.  $z = 0.194$ , with a mean value  $z = 0.271$ ). These objects at higher  $z$  tend to have slightly higher electron temperatures (12600 K on average) and densities ( $200\ \text{cm}^{-3}$ ), what translates into slightly lower oxygen abundances (average of  $12+\log(O/H) = 8.13$ ), but a slightly higher value for  $\log(N/O)$  ( $-1.24$ ) as compared to the whole sample. These differences can be explained in terms of a larger mean SFR and stellar mass of the objects in the high redshift subsample. This effect is even more clearly seen in the case of  $O/H$  when we restrict our sample to those objects with an  $EW([O\ III])$  of  $\lambda 5007\ \text{\AA} > 500\ \text{\AA}$ , that are also shown in the panels of Figure 1. In this case the mean electron temperature is even larger (13 500 K), the mean  $12+\log(O/H)$  smaller (8.05), and  $N/O$  has slightly larger mean value ( $-1.25$ ). The chemical properties of these galaxies are discussed and compared with other integrated properties, such as  $M_*$  or SFR, in Paper II of this series (Amorín et al, in prep.).

## 4 ABUNDANCE DERIVATION FROM EMPIRICAL CALIBRATIONS

### 4.1 Relation between O/H and electron temperature

One of the difficulties to apply the direct method for the derivation of total chemical abundances is the fact that not all needed lines corresponding to all ionization stages are available in the optical spectrum. If the missing ionization stages are not dominant in the total abundance, it is possible to resort to the use of ICFs based on models or observations. However, this leads to very high uncertainties when a large fraction of an element appears in unobserved stages. This can happen for oxygen when the  $[O\ II]\ \lambda 3727\ \text{\AA}$  cannot be measured (e.g. when the spectral coverage does not reach the wavelength region of this line), which is the case for a subsample of galaxies studied here.

One alternative can be to measure the weak auroral  $[O\ II]$  lines at 7319, 7330  $\text{\AA}$ , but these do not present an acceptable minimal SNR to be used in these data. Amorín et al. (2015) suggested the use of an empirical relation between the  $[O\ III]$  electron temperature, which only requires the measurement of  $[O\ III]\ \lambda 4363\ \text{\AA}$  and  $\lambda 4959, 5007\ \text{\AA}$ , with the total oxygen abundance. Since this relation was only proposed for local star-forming regions not selected on the basis of their  $EW([O\ III])$ , we study a new relation fitting the data of our sample of EELGs for which both total oxygen abundances and  $[O\ III]$  electron temperatures were derived. Taking only those objects with an  $O/H$  error lower than 0.2 dex, the sample consists of 1 268 galaxies. The mean  $[O\ III]$  electron temperature for this sample is  $t_e([O\ III]) = 11\ 800\ \text{K}$  and it approximately ranges from 8 500 to 19 700 K. They are plotted in Fig. 3. The error-weighted quadratical polynomial

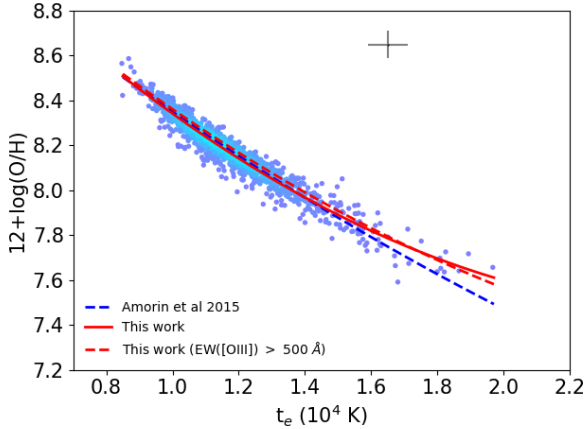
$$12 + \log(O/H) = (9.72 \pm 0.04) - (1.70 \pm 0.07) \times t_e + (0.32 \pm 0.03) \times t_e^2 \quad (1)$$

with  $t_e = t_e([O\ III])$  in units of  $10^4\ \text{K}$ . The standard deviation of the residuals is 0.04 dex. In Figure 3, this fit is close to the linear fit derived for a sample of green-pea galaxies in the Local Universe from Amorín et al. (2015). A fit considering only objects with  $EW([O\ III]) > 500\ \text{\AA}$ , also shown in the same figure, yields a polynomial very similar to that corresponding to the whole sample.

### 4.2 Strong-line methods to derive oxygen abundances

Strong-line methods to derive  $O/H$  appear as a common resource when the direct method or any other prescription based on a previous determination of the electron temperature cannot be applied. The collisionally excited lines are usually prominent in the optical spectrum and their relative intensities depend on the abundances of the ions that produce them.

Among the different strategies that could be followed to calibrate these strong-line methods, we select only the EELGs whose chemical abundances were derived following the direct method, or using a previous estimation of the electron temperature when the  $[O\ II]$  line is not measurable (i.e. from eq. 1). From this subset of EELGs, we only use for the calculation of empirical calibrations those objects that ender  $O/H$  with an uncertainty lower than 0.2 dex.



**Figure 3.** Relation between  $12+\log(\text{O}/\text{H})$  derived from the direct method and the  $[\text{O III}]$  electron temperature for the sample of EELGs where all required lines were available in the SDSS spectra. The solid red line represents a quadratic polynomial fit to the data, while the dashed blue line represents the linear fit published in Amorín et al. (2015). The red dashed line represents the polynomial fit to the objects with  $\text{EW}([\text{O III}]) > 500 \text{ \AA}$ . The cross in the upper right of the diagram represents the mean errors for both quantities.

#### 4.2.1 Parameters based on $[\text{O II}]$ and $[\text{O III}]$ emission lines

One of the most used strong-line parameters to derive O/H is R23, defined by Pagel et al. (1979) as,

$$\text{R23} = \frac{I([\text{O III}]4959, 5007) + I([\text{O II}]3727)}{I(\text{H}\beta)} \quad (2)$$

Among the several limitations of this parameter to provide a precise determination of the oxygen abundance (see Pérez-Montero & Díaz 2005), it shows a non-negligible dependence on effective temperature, ionization parameter, and it is double-valued with O/H. Indeed R23 increases with increasing O/H at low  $Z$  ( $12+\log(\text{O}/\text{H}) < 8.0$ ) and decreases for increasing O/H at  $Z$  ( $12+\log(\text{O}/\text{H}) > 8.4$ ). In between, in the so-called turn-over region, it is not possible to provide an accurate relation between R23 and O/H.

In Figure 4, we show the relation between  $\log(\text{R23})$  and the total oxygen abundance for those galaxies in our sample for which an estimate of the electron temperature could be derived. In this case all O/H were determined by means of the direct method as the measurement of the  $[\text{O II}]$  emission line at  $\lambda 3727 \text{ \AA}$  was required to compute the behavior of R23. No reliable polynomial fit can be provided as most of the sample lie in the turnover region and no clear relation between R23 and O/H exists. Although some galaxies lie either in the upper or lower branches, these are not enough to provide a calibration.

Other alternative to derive abundances using  $[\text{O II}]$  and  $[\text{O III}]$  emission lines is using the ratio between these two lines,

$$\text{O32} = \log \frac{I([\text{O III}]4959, 5007)}{I([\text{O II}]3727)} \quad (3)$$

This ratio is mostly sensitive to ionization parameter and equivalent effective temperature (e.g. Pérez-Montero & Díaz 2005; Dors & Copetti 2003), but it can also give an estimate of the total oxygen abundance on the basis of the relation between  $Z$  and  $\log U$  (e.g. Dopita et al. 2000). The right panel of Figure 4 shows the relation between this ratio and  $12+\log(\text{O}/\text{H})$  for those local analogs for which it the direct method or Equation 1 could be used (i.e. 1268 objects).

The correlation is not very marked (i.e. the correlation coefficient,  $\rho = -0.52$ ) but, as expected, objects with a larger metal content have on average a lower O32 parameter. The red solid line represents a linear fit to the data that leads to,

$$12 + \log(\text{O}/\text{H}) = (8.308 \pm 0.007) - (0.352 \pm 0.016) \times \text{O32} \quad (4)$$

This expression is valid in the range  $-0.1 < \text{O32} < 1.2$ , which corresponds to the range  $7.9 \lesssim 12+\log(\text{O}/\text{H}) \lesssim 8.35$ . The standard deviation of the residuals in  $12+\log(\text{O}/\text{H})$  is 0.12 dex. This fit is similar to the one obtained in Jones et al. (2015) (i.e. the data present +0.01 dex mean offset with respect to this calibration) for a sample of galaxies at intermediate-redshift, showing that our local sample has on average excitation properties similar to those in other star-forming galaxies at that redshift range. However, it is sensibly flatter than the linear expression derived by Bian et al. (2018) (i.e. the mean offset is -0.14 dex) for a sample of SDSS selected local-analogs calibrated from the resulting stacked spectra in bins of the N2 parameter.

#### 4.2.2 Parameters based on $[\text{Ne III}]$ emission lines

In high- $z$  objects, the  $[\text{O III}]$  emission line is often outside the observed spectral range. One alternative to derive O/H is using  $[\text{Ne III}]\lambda 3869 \text{ \AA}$ , which is particularly well detected in high excitation objects as those we are studying in the context of this work.

In left panel of Figure 5, we represent a scatter plot between the reddening-corrected relative-to- $\text{H}\beta$  intensities of  $[\text{O III}]\lambda 5007 \text{ \AA}$  and  $[\text{Ne III}]\lambda 3869 \text{ \AA}$ . As can be seen, these emission lines follow a tight relation (i.e. Spearman rank correlation coefficient  $\rho = 0.89$ ) whose slope is  $\alpha = 11.0 \pm 1.4$ , that is slightly lower than the value derived in Pérez-Montero et al. (2007) for a sample of local star-forming objects ( $15.4 \pm 1.5$ ). This tight relation is due to the quite similar excitation structure of O and Ne in the ionized gas and it implies that all additional dependencies of the  $[\text{O III}]/[\text{O II}]$  emission-line ratio are also present in the case of  $[\text{Ne III}]/[\text{O II}]$ .

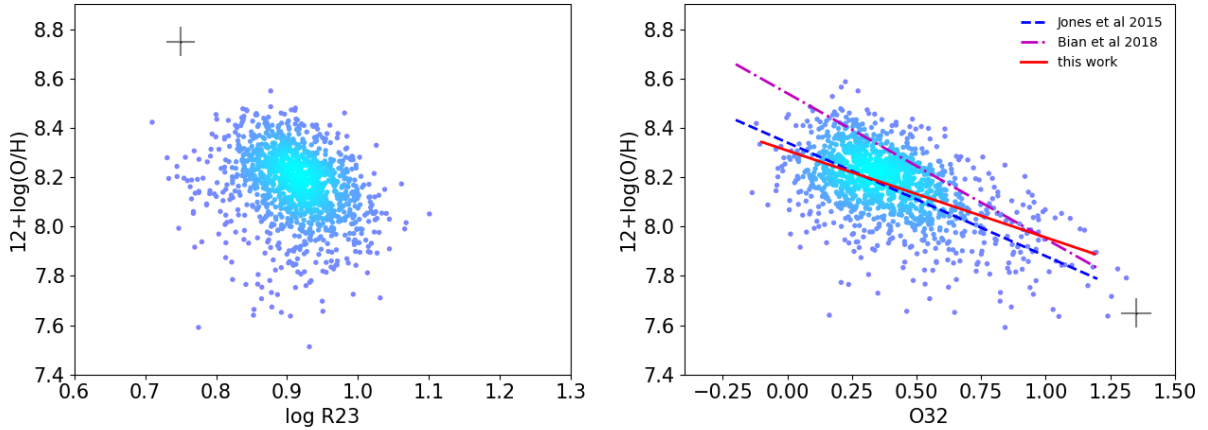
One of these dependencies is the strong relation with ionization parameter that is related with the metal content of the gas. In this way, we can define the Ne3O2 parameter as

$$\text{Ne3O2} = \log \frac{I([\text{Ne III}]3869)}{I([\text{O II}]3727)} \quad (5)$$

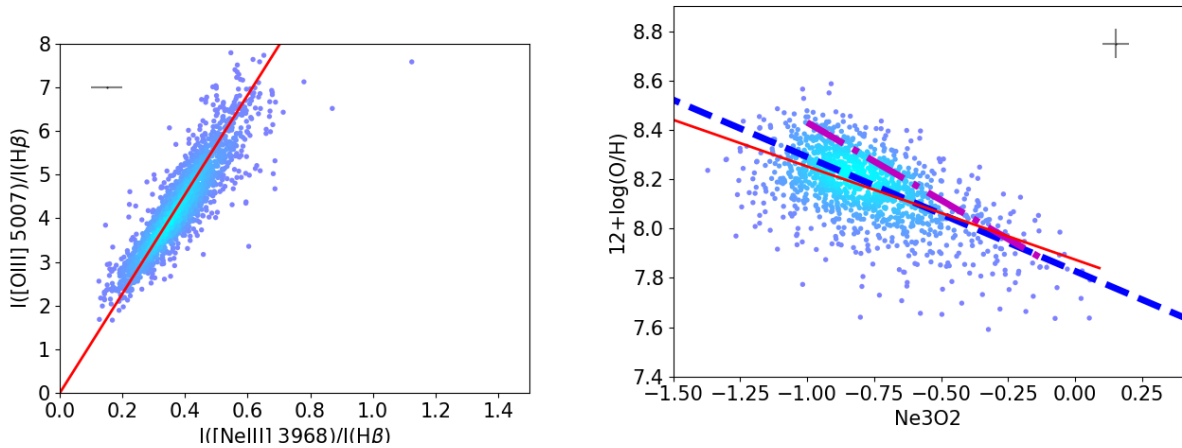
In the right panel of figure 5 we show the relation between this parameter and  $12+\log(\text{O}/\text{H})$  for those EELGs studied in this work for which this abundance was available through the direct method. As in the case of O32 the correlation is not very marked ( $\rho = -0.56$ ), but we can provide a linear fit to this sample that gives,

$$12 + \log(\text{O}/\text{H}) = (7.823 \pm 0.009) - (0.455 \pm 0.013) \times \text{Ne3O2} \quad (6)$$

This fit presents an uncertainty of 0.12 dex, calculated as the standard deviation of the residuals in the range  $-1.3 < \text{Ne3O2} < 0.0$ , what implies a validity range of  $7.8 \lesssim 12+\log(\text{O}/\text{H}) \lesssim 8.4$ . As in the case of O32, no significant difference is found with respect to the fit by Jones et al. (2015) (i.e. the mean offset with respect to the data is -0.02 dex) for a sample of intermediate-redshift galaxies, although the derived slope is slightly lower in our case. Nevertheless, as in the case of O32, a larger difference is found in relation to the linear fit provided by Bian et al. (2018) (i.e. the mean offset is -0.12 dex) based on local SDSS analogs, as they found a much pronounced slope for the fit.



**Figure 4.** Left panel: relation between  $12+\log(\text{O}/\text{H})$  and  $\log(\text{R}23)$  for those EELGs for which an estimation of  $\text{O}/\text{H}$  based on the electron temperature was possible. Right panel: Relation with  $\text{O}32$  for the same set of EELGs. The blue dashed line represents the linear fit given by Jones et al. (2015), the dotted-dashed magenta line represents the fit given by Bian et al (2018), and the red solid line is a linear fit to the plotted data. In both panels the mean errors are represented with a cross.



**Figure 5.** Left panel: The relation between the reddening-corrected relative-to- $\text{H}\beta$  intensities of  $[\text{O III}] \lambda 5007 \text{ \AA}$  and  $[\text{Ne III}] \lambda 3869 \text{ \AA}$  in our sample of EELGs. Right panel: relation between  $12+\log(\text{O}/\text{H})$  derived following the direct method and the  $\text{Ne}3\text{O}2$  parameter. The blue dashed line represents the linear fit given in Jones et al. (2015), the dotted-dashed magenta line represents the fit given by Bian et al (2018), and the solid red line is a linear fit to the sample analyzed in this work. The crosses in both panels represent the mean errors in abscissae and ordinates.

#### 4.2.3 Parameters based on $[\text{N II}]$ emission lines

Alternatively, other strong-line calibrators based on  $[\text{N II}]$  lines have been proposed to overcome the issue of the double-valued behavior of  $\text{R}23$ . The  $\text{N}2$  parameter was defined by Storchi-Bergmann et al. (1994) as,

$$\text{N}2 = \log \left( \frac{I([\text{N II}]6583)}{I(\text{H}\alpha)} \right) \quad (7)$$

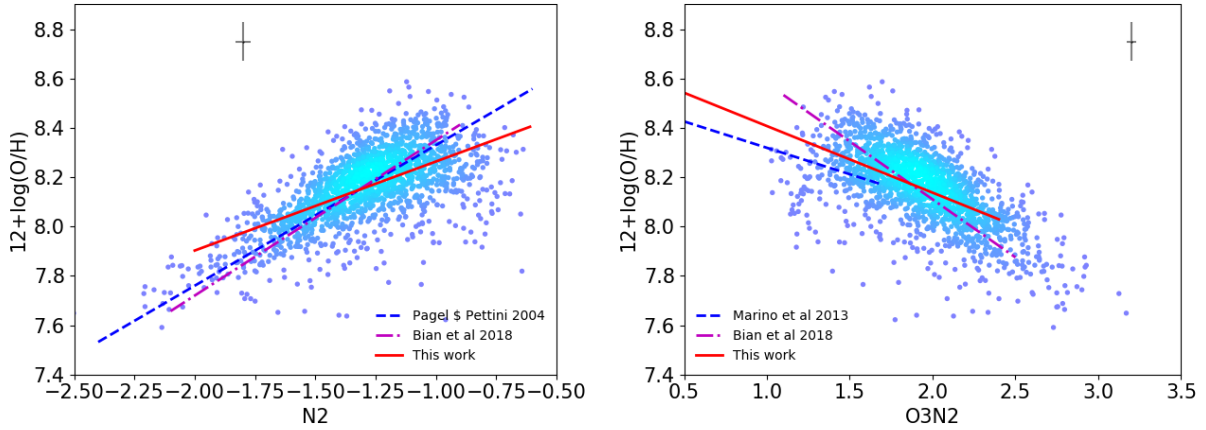
This parameter presents a monotonic relation with  $\text{O}/\text{H}$  and has the advantage of being easily measurable in the rest-frame red part of the optical spectrum with very low dependence on reddening or flux calibration. On the other hand, according to Pérez-Montero & Díaz (2005) it has a large dispersion due to its dependence on ionization parameter and  $\text{N}/\text{O}$  abundance ratio.

In Fig. 6, we show the relation between the  $\text{N}2$  parameter and the total oxygen abundance for the 1675 objects of our sample of EELGs for which  $[\text{N II}]$  has a  $\text{SNR} > 3$  and there is an estimation of  $\text{O}/\text{H}$  based on the electron temperature, either from the direct

method or from the empirical calibration with  $t_e$  studied above (Eq. 1), and with an uncertainty lower than 0.2 dex. In this case the correlation is slightly better than for  $\text{O}32$  ( $\rho = 0.66$ ). However, for metal-poor galaxies (i.e.  $12+\log(\text{O}/\text{H}) < 8.0$ ), this correlation worsens. The lack of correlation between  $\text{N}2$  and  $\text{O}/\text{H}$  in metal-poor galaxies has been already reported (e.g. Morales-Luis et al. 2014) and it is related to the relatively high  $\text{N}/\text{O}$  ratios found in some of these galaxies. Actually, in our sample  $\log(\text{N}/\text{O})$  does not depend much on metallicity:  $\log(\text{N}/\text{O}) = -1.34 \pm 0.26$  for  $12+\log(\text{O}/\text{H}) < 7.9$  and  $-1.31 \pm 0.16$  for  $12+\log(\text{O}/\text{H}) > 7.9$ . A more thorough discussion on this issue will be provided elsewhere (Amorín et al., in preparation). Thus, we only provide a linear fit in the range  $-2.0 < \text{N}2 < -0.6$  which gives the following expression,

$$12 + \log(\text{O}/\text{H}) = (8.625 \pm 0.015) + (0.361 \pm 0.012) \times \text{N}2 \quad (8)$$

, which holds for  $7.9 \lesssim 12+\log(\text{O}/\text{H}) \lesssim 8.4$ , and presents a standard deviation of the residuals of 0.12 dex. As it can be seen in Figure 6, the obtained fit gives a less pronounced slope and a lower intercept



**Figure 6.** Relations between  $12+\log(\text{O}/\text{H})$  derived from electron temperature and strong line ratios using  $[\text{N II}]$  emission lines fluxes ( $\text{N2}$  in the left panel and  $\text{O3N2}$  in the right panel). The solid red line represents linear fits to the plotted data. Note that this fit is only valid for  $\text{O3N2} < 2.5$  in the case of the  $\text{O3N2}$  parameter. The dashed blue line represents the linear fit by Pettini & Pagel (2004) for  $\text{N2}$ , and the linear calibration from Marino et al (2013) for  $\text{O3N2} < 1.8$ . The dotted-dashed magenta line represent in both panels the linear fit given by Bian et al (2018). In both panels the crosses represent the mean errors.

than the linear relation given by Pettini & Pagel (2004) for a sample of  $\text{H II}$  regions and  $\text{H II}$  galaxies in the Local Universe. This implies that the relation for the EELGS leads to lower values of  $\text{O}/\text{H}$  for the same  $\text{N2}$ , in the high- $Z$  regime, even if the mean offset with respect to the Pettini & Pagel (2004) calibration is  $-0.01$  dex for this sample. This different slope can possibly be caused by the different mean ionization parameter of the two samples. Since  $\log U$  is much larger in EELGs, low-excitation lines such as  $[\text{N II}]$  are fainter for a fixed  $\text{O}/\text{H}$ , so the EELG calibration of this parameter leads to higher values of  $\text{O}/\text{H}$  for the same  $[\text{N II}]$  intensities at lower  $\text{O}/\text{H}$ . This difference is also found in the case of the local SDSS analogs studied in Bian et al. (2018), who obtain a very similar fit to the rest of local  $\text{H II}$  regions for this parameter and presents a higher slope.

Another combination of collisionally excited strong emission-lines related to  $[\text{N II}] \lambda 6583 \text{ \AA}$  used to derive  $\text{O}/\text{H}$  is the  $\text{O3N2}$  parameter, defined by Alloin et al. (1979) as,

$$\text{O3N2} = \log \left( \frac{I([\text{O III}]\lambda 5007)}{I(\text{H}\beta)} \cdot \frac{I(\text{H}\alpha)}{I([\text{N II}]\lambda 6583)} \right) \quad (9)$$

This parameter is thought to reduce the dependence on ionization parameter as it follows the ionization sequence of  $[\text{O III}]/\text{H}\beta$  and  $[\text{N II}]/\text{H}\alpha$  emission-line ratios observed in one of the BPT diagrams, as shown in Figure 2. The main drawback of using  $\text{O3N2}$  is the fact that it does not vary at low  $\text{O}/\text{H}$  and according to Pérez-Montero & Contini (2009) no clear correlation can be obtained for  $\text{O3N2} > 2$ .

The right panel of Fig. 6 shows the relation between  $\text{O3N2}$  and total oxygen abundance for our sample of EELGs with  $\text{SNR} > 3$  in all the required lines and also show a derivation of the total oxygen abundance based on the electron temperature. As it can be seen,  $\text{O}/\text{H}$  and  $\text{O3N2}$  show a very good correlation in our sample of EELGs, even for large  $\text{O3N2}$ . Thus, it is possible to provide a linear fit giving (with  $\rho = -0.63$ ),

$$12 + \log(\text{O}/\text{H}) = (8.677 \pm 0.018) - (0.270 \pm 0.009) \times \text{O3N2} \quad (10)$$

that is obtained in the range  $1.0 < \text{O3N2} < 2.7$ , which corresponds to the range  $7.9 \lesssim 12+\log(\text{O}/\text{H}) \lesssim 8.4$ . The standard deviation of the residuals is  $0.12$  dex. This fit can be compared with the one obtained by Marino et al. (2013) for a sample of local  $\text{H II}$  regions and  $\text{H II}$  galaxies spanning a range of  $\text{O3N2} < 1.8$ . In this case, the fit for

the EELGS has a slightly larger slope and larger intercept, so this calibration leads to higher values of  $\text{O}/\text{H}$  for a same value of  $\text{O3N2}$  as compared with the calibration for the local star-forming sample (i.e. the mean offset for this calibration is  $0.05$  dex). In addition, it reaches a larger  $\text{O3N2}$  value. As in the case of  $\text{N2}$ , this can be interpreted on the basis of a larger average ionization parameter of the EELGs sample. Larger values of  $\log U$  make the  $\text{O3N2}$  to be larger for a same  $\text{O}/\text{H}$ , so the calibration leads to larger values of  $\text{O}/\text{H}$ . Conversely, our fit presents again a smaller slope as compared to that given by Bian et al. (2018) for their sample of local analogs calibrated using bins of the  $\text{N2}$  parameter but, in this case, no offset is found between the abundances derived using this calibration and the mean value derived for  $\text{O}/\text{H}$ .

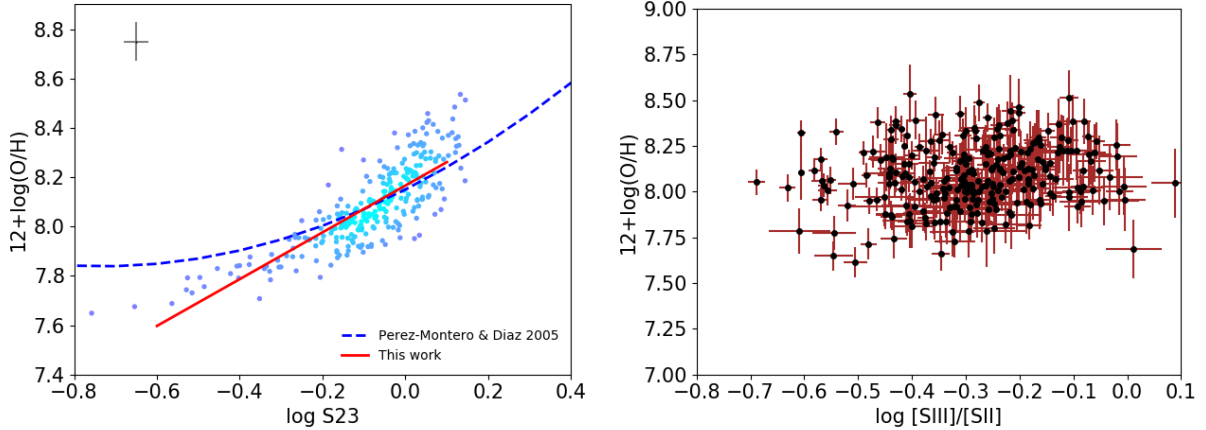
#### 4.2.4 Parameters based on $[\text{S II}]$ and $[\text{S III}]$ emission lines

We also explored strong-line indices sensitive to  $\text{O}/\text{H}$  based on  $[\text{S II}]$  and  $[\text{S III}]$  lines. The  $\text{S23}$  parameter (Vilchez & Esteban 1996; Díaz & Pérez-Montero 2000) is defined as,

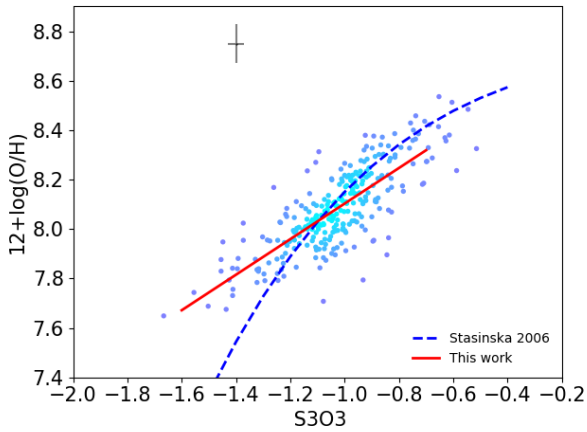
$$\text{S23} = \left( \frac{I([\text{S III}]\lambda 9069, 9532) + I([\text{S II}]\lambda 6717, 6731)}{I(\text{H}\beta)} \right). \quad (11)$$

This parameter shows a monotonic relation with  $\text{O}/\text{H}$  up to much higher  $Z$  than in the case of  $\text{R23}$ , so it partially fixes the problem of the selection of the appropriate branch in the case of  $\text{R23}$ . Its dependence on reddening can be overcome using the theoretical relation between  $\text{H}\alpha$  or  $\text{H}\beta$  with the Paschen  $\text{H}\gamma$  recombination lines close to the  $[\text{S III}]$  lines. In the case of our EELG sample, the observed SDSS spectra do not cover the redder  $[\text{S III}] \lambda 9532 \text{ \AA}$  line, but we can use a theoretical relation with the weaker  $[\text{S III}] \lambda 9069 \text{ \AA}$  line (i.e.  $I([\text{S III}]\lambda 9532) = 2.44 I([\text{S III}]\lambda 9069)$ ). On the other hand, even this line is only detected in the SDSS spectra up to a very low redshift ( $z < 0.02$ ), and hence we could only measure it in 349 objects of our sample. Moreover, the measurement of the  $[\text{S III}] \lambda 9069 \text{ \AA}$  can only be carried out in objects where the  $[\text{O II}] \lambda 3727 \text{ \AA}$  lies out of the observed spectral range, so a determination of the total oxygen abundance in these objects is not possible. Alternatively, the empirical calibration between  $[\text{O III}]$  electron temperature and  $12+\log(\text{O}/\text{H})$  presented in subsection 4.1 can be used. In left panel





**Figure 7.** Relations between  $\log(S23)$  (left panel), and  $\log([S\text{ III}]/[S\text{ II}])$  (right panel) with  $12+\log(O/H)$  of our EELG sample as derived from a previous estimate of the electron temperature. The solid red line represents a linear fit to the data. In left panel, the blue dashed line represents the polynomial fit given by Pérez-Montero & Díaz (2005). Crosses represent the mean errors.



**Figure 8.** Relation between  $S3O3$  with  $12+\log(O/H)$  of our EELG sample as derived from a previous estimate of the electron temperature. The solid red line represents a linear fit to the data. The blue dashed line represents the polynomial fit given by Stasińska (2006). The crosses represents the mean errors.

of Figure 7, we show the obtained relation between  $\log S23$  and  $12+\log(O/H)$  as derived using this method. The resulting correlation is very good ( $\rho = 0.82$ ), although the  $O/H$  range for this subsample is somehow restricted. A linear fit yields,

$$12+\log(O/H) = (8.166 \pm 0.07) + (0.949 \pm 0.004) \times \log(S23). \quad (12)$$

This fit was derived in the range  $-0.6 < \log(S23) < 0.2$ , which corresponds to the range  $7.6 \lesssim 12+\log(O/H) \lesssim 8.35$ , and the standard deviation of the residuals is of 0.09 dex. There are no significant differences with respect to the quadratical fit by Pérez-Montero & Díaz (2005) (i.e. the mean offset of the data for this calibration is -0.01 dex) for a sample of local star-forming objects, with the exception of very low values of  $O/H$ . In our case, a non-linear fit does not provide a better agreement with the data, as the high- $Z$  range is not properly covered.

As shown in the right panel of Figure 7 and contrary to other high-to-low excitation emission-line intensity ratios such as  $O3O2$  or  $Ne3O2$ , no correlation between oxygen abundance and the emis-

sion line ratio  $[S\text{ III}]/[S\text{ II}]$  is observed ( $\rho = 0.06$ ). This different behaviour could be due to the fact that  $O32$  and  $Ne3O2$  are not directly tracing the excitation of the gas, but the hardness of the ionizing radiation (Dors et al. 2017; Pérez-Montero et al. 2019a). On the contrary, the emission-line ratio  $[S\text{ III}]/[S\text{ II}]$  shows very little dependence on the equivalent effective temperature of the ionizing source and strongly depends on the ionization parameter (e.g. Mathis 1985; Díaz et al. 1991; Morisset et al. 2016). Thus, the observed relation of  $O32$  or  $Ne3O2$  with  $O/H$  could be more the result of a connection between  $Z$  and the hardness of the ionizing source, rather than a relation with  $U$ .

On the other hand, it is possible to derive  $O/H$  using the  $S3O3$  parameter, defined as,

$$S3O3 = \log \left( \frac{I([S\text{ III}]9069, 9532)}{I([O\text{ III}]4959, 5007)} \right). \quad (13)$$

This parameter was proposed as estimator for  $O/H$  by Stasińska (2006). The relation between this parameter and  $12+\log(O/H)$  is represented in Figure 8. The correlation in this case is also good ( $\rho = 0.81$ ) and the linear fit to the data yields the following expression,

$$12 + \log(O/H) = (8.822 \pm 0.034) + (0.719 \pm 0.033) \times S3O3. \quad (14)$$

The standard deviation of the residuals is only of 0.10 dex in the range  $-1.6 < S3O3 < -0.6$ , which corresponds to  $7.6 \lesssim 12+\log(O/H) \lesssim 8.4$ . There is a good agreement as compared to the polynomial relation proposed by Stasińska (2006) (i.e. the mean offset of the data with this calibration is -0.01 dex).

A possible reason for the correlation between this emission-line ratio and  $O/H$  is probably the same as that for the cases of  $O23$  or  $O2Ne3$ , and contrary to the case of  $[S\text{ III}]/[S\text{ II}]$ , mostly due to a correlation with the shape of the spectral energy distribution (SED) than to a dependence on excitation.

### 4.3 Strong-line methods to derive nitrogen-to-oxygen abundance ratio

Several different efforts have been made to also provide strong-line calibrators for the derivation of the nitrogen-to-oxygen abundance ratio (e.g. Thurston et al. 1996; Liang et al. 2006; Pilyugin &

Grebel 2016), but little has been explored on calibrations based on high-redshift objects (e.g. Strom et al. 2018). According to Pérez-Montero & Contini (2009), N/O can be derived using collisional strong emission lines combined to form the N2O2 parameter, defined as,

$$\text{N2O2} = \log \left( \frac{I([\text{NII}]6583)}{I([\text{OII}]3727)} \right). \quad (15)$$

This parameter has been also suggested as calibrator for O/H (Kewley et al. 2002), given its very low dependence on  $U$ , and the expected relation between N and O abundances at high- $Z$ , where most of the N production has a secondary origin. In left panel of Fig. 9 we show the relation between this emission-line ratio and  $\log(\text{N/O})$  as derived following the direct method for those EELGs whose emission lines could be measured with enough S/R to yield N/O with an uncertainty lower than 0.2 dex. The observed points present a very tight correlation ( $\rho = 0.96$ ) and can be fitted by a linear relation. The fit gives the following expression,

$$\log(\text{N/O}) = (-0.316 \pm 0.009) + (0.870 \pm 0.007) \times \text{N2O2}. \quad (16)$$

This fit has been calculated in the range  $-1.7 < \text{N2O2} < -0.5$ , which corresponds to  $-1.8 \lesssim \log(\text{N/O}) \lesssim -0.75$ , and the standard deviation of the residuals is 0.04 dex. As compared with the linear relation derived by Pérez-Montero & Contini (2009) for a local sample of star-forming regions, the difference is not larger than the derived errors (i.e. the mean offset of the data to this calibration is -0.03 dex). Since N2O2 does not depend on ionization parameter the obtained relations based on this parameter do not vary much even for families of objects whose mean excitation is quite different. We also compare our calibration with the model-based linear relation with N/O given for this parameter for high-redshift objects by Strom et al. (2018), which shows a smaller slope.

The two lines involved in N2O2 have a significant wavelength difference, which makes it more sensitive to observational issues, like calibration errors or the fact that a single observational setup may not yield them simultaneously. To overcome the problem, Pérez-Montero & Contini (2009) also suggests the use of the N2S2 parameter, defined as,

$$\text{N2S2} = \log \left( \frac{I([\text{NII}]6583)}{I([\text{SII}]6717, 6731)} \right). \quad (17)$$

This has been also suggested as calibrator for O/H in high-redshift galaxies (Dopita et al. 2016), on the basis of a relation between O and N in the secondary N production regime. The right panel of Fig. 9 shows the relation between this parameter and the nitrogen-to-oxygen ratio as derived following the direct method for our sample of EELGs. The correlation is slightly worse than in the case of N2O2 ( $\rho = 0.88$ ) and a linear fit to the data gives the following expression,

$$\log(\text{N/O}) = (-1.005 \pm 0.005) + (0.857 \pm 0.013) \times \text{N2S2}. \quad (18)$$

The fit is valid for the range covered by the data  $-0.8 < \text{N2S2} < 0.3$ , which corresponds to the range  $-1.7 \lesssim \log(\text{N/O}) \lesssim -0.75$ . The standard deviation of the residuals is 0.08 dex. As in the case of N2O2, there is not large difference compared to the linear fit derived by Pérez-Montero & Contini (2009) for a local sample (there is almost no mean offset), as N2S2 does not present a strong dependence on ionization parameter. In addition, differently to the

case of N2O2, our linear fit is nearly identical to the model-based one proposed by Strom et al. (2018) for high-redshift galaxies.

## 5 DETERMINATION OF CHEMICAL ABUNDANCES BASED ON MODELS

An alternative method to calibrate strong lines to derive chemical abundances is based on the direct comparison with photoionization models. This approach allows to examine the parameter space in detail avoiding possible observational biases (e.g. the O/H range or the cut-off in luminosity of the used emission lines). In this section we explore an adapted version of the code HCM (Pérez-Montero 2014), to the conditions of our sample of EELGs and we check the accuracy in the determination of both O/H and N/O as compared to the values derived from the direct method or from the direct calibration of the electron temperature. It can be useful for different observed spectral ranges and different emission line sets.

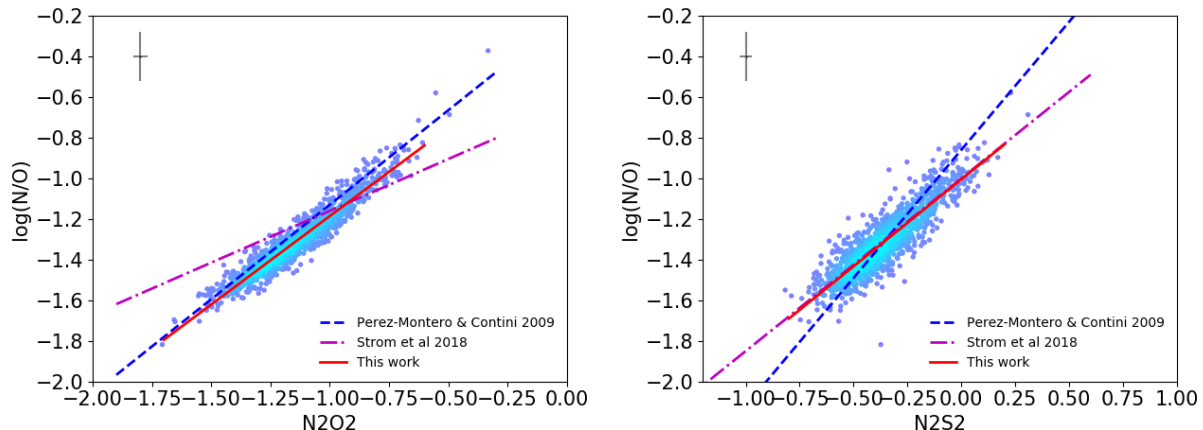
### 5.1 Description of the models

The grid of models used by HCM for EELGs was calculated using the photoionization code CLOUDY v.17.00 (Ferland et al. 2017). This code calculates the line fluxes emitted by a gas distribution irradiated by a central source. We considered input properties of the models thought to reproduce some of the expected and observed properties in our sample of EELGs, including synthetic model cluster atmospheres and gas densities typical in extreme starbursts at low- and intermediate-redshift. In this case we considered the SEDs from BPASS v.2.1 (Eldridge et al. 2017), assuming an initial mass function with binaries and a Salpeter slope of  $x = -1.35$  and an upper stellar mass limit of  $300 M_{\odot}$ . The  $Z$  of the stars was chosen to match that of the gas as scaled from the total oxygen abundance, covering a range  $12 + \log(\text{O/H}) = [6.9, 9.1]$  in bins of 0.1 dex. In terms of the solar abundance, the range goes from 1/60 to 2.6 times the solar oxygen abundance in Asplund et al. (2009). A standard dust-to-gas mass ratio was considered in all models, along with a filling factor of 0.1. Two different electron densities were assumed with values  $n_e = 100$  and  $500$  particles per  $\text{cm}^{-3}$ , covering the values found in our sample of EELGs. The model calculation was stopped when the proportion of electrons in the gas in relation to hydrogen atoms was lower than 98%. The resulting geometry of all models was plane-parallel. To complete the grid, for each O/H we considered values of  $\log(\text{N/O})$  in the range  $[-2.0, 0.0]$  in bins of 0.125 dex, and values of the ionization parameter  $\log U = [-4.0, -1.5]$  in bins of 0.25 dex. This gives 4301 models for each density.

From each model we extracted the emission line fluxes relative to  $\text{H}\beta$  relevant for the calculation of the abundances. These are  $[\text{O II}] \lambda 3727 \text{ \AA}$ ,  $[\text{Ne III}] \lambda 3869 \text{ \AA}$ ,  $[\text{O III}] \lambda 4363, \lambda 5007 \text{ \AA}$ ,  $[\text{N II}] \lambda 6584 \text{ \AA}$ , and  $[\text{S II}] \lambda 6717+6731 \text{ \AA}$ .

### 5.2 Description of the code

The code HCM calculates the total oxygen abundance, nitrogen-to-oxygen abundance ratio and ionization parameter in gaseous nebulae ionized by massive stars using optical emission-lines and comparing them using a bayesian-like method with the predictions from a large grid of photoionization models covering a wide range of input conditions. The code is described in Pérez-Montero (2014), and the different later modifications and improvements appear in version 2 (use of an interpolated grid of models, Pérez-Montero et al. 2016); version 3 (use of Monte Carlo iterations to estimate the derived



**Figure 9.** Density plots for the relation between  $\log(N/O)$ , as derived following the direct method, and the parameters  $N2O2$  (on the left) and  $N2S2$  (on the right) for the sample of EELGs in our study. The blue dashed line represent the linear fits obtained in Pérez-Montero & Contini (2009) and the magenta pointed-dashed lines represent the linear fits based on models given by Strom et al (2018). The solid red lines correspond to the linear fits to the sample of EELGs. Crosses represent the mean errors.

error, Pérez-Montero et al. 2019a); or version 4 (application for the Narrow Line Region of Seyfert 2 galaxies, Pérez-Montero et al. 2019b). Other versions are also devoted to the calculation of the carbon-to-oxygen abundance ratio and the oxygen total abundances using ultraviolet emission lines (i.e.  $H_{CM-UV}$ , Pérez-Montero & Amorín 2017). All these versions, along with the one described here for EELGs are publicly available<sup>4</sup>.

In brief, the code firstly calculates  $N/O$  considering that the emission-line ratios  $N2O2$  and  $N2S2$ , described in previous sections, are almost independent of the ionization parameter. This calculation is performed creating a distribution of  $\chi^2$  values with the difference between the observed emission-line ratios and the prediction of all models in the grid. The final  $N/O$  value is the weighted-mean of this distribution. Its error is calculated as the quadratical sum of the standard deviation of the same distribution and the dispersion of a iterative series of Monte Carlo simulations. For each iteration, the observed emission line fluxes are perturbed using the observed errors.

Once the grid of models is reduced to the calculated  $N/O$  value within errors, a new iteration through the resulting grid constrained in the  $N/O$  space is done to calculate  $O/H$  and  $\log U$ . In this case, the code calculates the  $\chi^2$  weights using as observables different emission-line ratios known to have some dependence on  $O/H$  and  $U$  (e.g.  $R23$ ,  $O32$ ,  $N2$ ,  $O3N2$ ). The resulting values and their corresponding errors are calculated following the same procedure as described above for  $N/O$ .

The code allows the use of the default grid of models or of an interpolated grid that enhances the resolution in the three studied dimensions by a factor of 10. This option minimizes the clustering of the solutions around the values of the grid. The code also allows to perform calculations attending to different SEDs in the models, including an instantaneous burst with POPSTAR (Mollá et al. 2009), with a Chabrier IMF and an age of 1 Myr, or a double peak power law for Seyfert 2 galaxies. For this work we use the models described in the above subsection based on BPASS v. 2.1 (Eldridge et al. 2017) with binaries, an IMF with upper stellar mass cut-off at  $300 M_{\odot}$  and a slope of -1.35 at an age of 1 Myr. The grid with density of

$100 \text{ cm}^{-3}$  is the one available in the public version of the code, but we also comment the results obtained using the higher density of  $500 \text{ cm}^{-3}$ .

### 5.3 Comparison with the direct method

The chemical abundances for the EELGs obtained with  $H_{CM}$  AND THE GRID DESCRIBED ABOVE IS COMPARED HERE WITH THE ABUNDANCES FROM THE DIRECT METHOD OR FROM ELECTRON TEMPERATURE AS PARAMETERIZED IN EQ. (1).

In Table 1, we list the mean errors for  $O/H$  and  $N/O$  as obtained by the code using different combinations of emission lines. We also give the mean ( $\Delta$ ) and the standard deviation ( $\sigma$ ) of the residuals as compared with the empirically derived abundances. These  $\Delta$  and  $\sigma$  values can be taken as empirical corrections and additional uncertainties, respectively, to the chemical abundances derived by  $H_{CM}$  when different sets of emission-lines are used as input.

Figure 10 shows the comparison between the abundances derived following the direct method and those obtained by our code when all lines are available and the derived abundances have an uncertainty smaller than 0.2 dex. These criteria include 1269 EELGs. The mean offset with respect to the  $O/H$  and  $N/O$  derived by  $H_{CM}$  is in average smaller than the mean errors or the standard deviation of the residuals. However, for values of  $12+\log(O/H) < 8.0$ , the code tends to overestimate the resulting  $O/H$  (i.e. 0.1 dex in average). This offset for this regime was also found in the version of the code for  $H_{II}$  regions. Regarding  $N/O$ , for values of  $\log(N/O)$  lower than -1.6 the code tends to overestimate abundances above  $1\sigma$ , but this deviation is reduced when  $[S_{II}]$  is not considered. We also checked the impact of increasing electron density in the models from 100 to  $500 \text{ cm}^{-3}$ , but we did not found any significant deviation outside the obtained errors for none of the derived quantities.

Contrary to other published abundance determinations based on models (e.g. Kewley & Dopita 2002; Blanc et al. 2015; Vale Asari et al. 2016), the agreement between the results obtained from the direct method and from  $H_{CM}$  is good within the errors. This difference between the predictions of some models and the abundances derived from the direct method, also known as Abundance Discrepancy Factor (ADF), has been explained invoking several physical causes, such as shocks (e.g. Dopita et al. 1984; Peim-

<sup>4</sup> <http://www.iaa.csic.es/~epm/HII-CHI-mistry.html>.

bert et al. 1991), temperature fluctuations (e.g. Peimbert & Costero 1969; García-Rojas & Esteban 2007), density or abundance inhomogeneities (e.g. Tsamis & Péquignot 2005), and  $\kappa$  distributions (e.g. Dopita et al. 2013). However, these physical conditions should be introduced in the models as input conditions to explain the obtained ADF and, if not, models and empirical derivations should lead to the same results under the same assumptions and atomic data.

The general agreement between the abundances derived from the direct method and from HCM also implies that the fraction of unseen oxygen in the optical spectra (e.g.  $O^{3+}$ ) is negligible in this sample of EELGs.

The comparison between the abundances derived from models with those empirically obtained can be extended to all the sources with an estimate of electron temperature via the emission-line ratio  $[O\text{ III}] 5007/4363$ , even if the  $[O\text{ II}] \lambda 3727 \text{ \AA}$  is not detected due to the blue cutoff in the SDSS spectral range. As it can be seen in Table 1, the analysis of the 1714 objects in our sample with an accurate determination of temperature leads to O/H in agreement with the direct method within the mean uncertainty, even when  $[O\text{ II}] \lambda 3727 \text{ \AA}$  is not taken into account. Regarding the N/O obtained without  $[O\text{ II}]$ , there is also a general agreement within error, although both the mean uncertainty (0.17 dex) and the mean offset (0.08 dex) are higher than those for the N/O derived using  $[O\text{ II}]$ .

Since HCM gives as result both O/H and  $U$ , we can explore the relation between these two properties for the sample of EELGs. It must be kept in mind that this relation can be also interpreted in terms of an empirical relation between O/H and the hardness of the incident radiation, but our grid of models does not consider these, as a single SED is assumed for all the models. The left panel of Figure 11 represents this relation as calculated by HCM when the  $[O\text{ III}] \lambda 4363 \text{ \AA}$  emission-line is used. As in the star-forming regions discussed by Pérez-Montero (2014) there is a trend to find lower excitation in objects with higher O/H, but this is less evident and there is also a mild trend to find metal-rich objects with larger values of  $U$ . In fact,  $U$  in EELGs tends to be larger than in the rest of star-forming regions. This can be seen in the same Figure, taking as a reference the grid of models encompassing the local SF regions analyzed by HCM. EELGs present  $\log U$  larger than average, although with significant dispersion. A new empirical relation between O/H and  $U$  can be hence taken as a constraint in the grid of models to derive abundances in absence of the  $[O\text{ III}]$  auroral line. In the case of EELGs, this relation adopt values of  $\log U$  slightly higher than those considered for other star-forming objects at all  $Z$  regimes.

Table 1 also includes the results of the comparison between the abundances derived by the code using a limited set of emission lines in absence of the  $[O\text{ III}]$  auroral line and those obtained using the direct method or the empirical relation with the electron temperature. As shown in Figure 12, the deviations are in general within the standard deviation of the residuals and, as expected, they increase as the number of used emission lines is smaller. Regarding the mean errors derived by the code, the rule is not strict. For example, when all lines are included, the errors are among the largest (Table 1). This can be understood keeping in mind that HCM accounts for statistical errors, namely, evaluates how the uncertainties of the fluxes propagate into abundances. Many sources of systematic error are not included here, and can be significant. Thus, the use of many lines allows bringing out systematic errors bypassed when fewer lines are used and, all in all, may increase HCM error when increasing the number of lines. In any case, as can be seen in Figure 12, HCM tends to overestimate the abundances derived from the direct method at

low O/H and to underestimate O/H at high  $Z$ . The best way to find an accurate determination of the abundance relies always on the detection of the auroral line to derive the electron temperature, however, when this is not possible, HCM leads to values consistent at first order with the direct method and allows us to discriminate, with confidence, between low- and high- $Z$  objects.

Finally, when no previous determination of N/O can be performed because neither  $N2O2$  nor  $N2S2$  are measured in the observed spectra, the code has to make a priori assumption about the relation between O/H and N/O in order to use  $[N\text{ II}]$  lines to derive O/H. The right panel of Figure 11 shows the relation adopted in Pérez-Montero (2014) compared to the values obtained in our sample of EELGs. As it can be seen, many of the sampled EELGs present combinations of N/O and O/H that lie outside the expected relation, as they show higher N/O values at low O/H than other low-metallicity samples (e.g. Pilyugin et al. 2012; Dopita et al. 2016; Guseva et al. 2020). This plot will be thoroughly discussed in a forthcoming work (Amorín et al., in preparation), but here it illustrates the risk of assuming a specified O/H vs N/O relation for these objects in order to derive O/H abundances using N emission-lines.

We also show in Table 1 the values obtained from our analysis when the lines necessary to derive N/O are not available. As an example, Figure 13 compares the O/H derived using HCM and the direct method for two combinations of lines where a O/H versus N/O relation has to be assumed. The results based on  $[O\text{ II}]$  and  $[Ne\text{ III}]$  are slightly better than when only  $[N\text{ II}]$  is used. In any case, the offset is always at the limit of the standard deviation of the residuals and, as in previous cases, it is sensibly worse in the low- $Z$  regime.

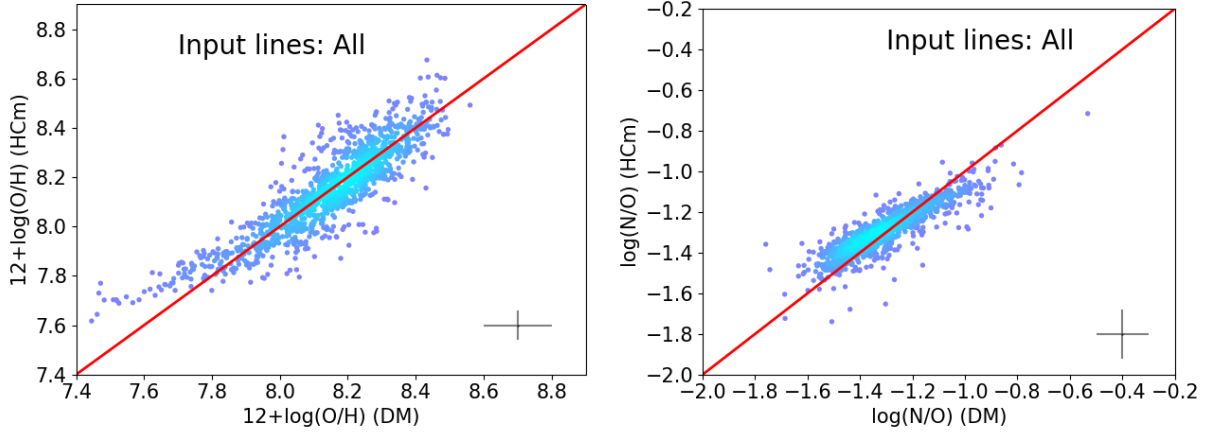
## 6 SUMMARY AND CONCLUSIONS

We presented the analysis of a sample of 1969 EELGs at redshifts  $0.00 < z < 0.49$  (mean  $z = 0.08$ ) selected from the SDSS using the procedure described in Sánchez Almeida et al. (2010). The sample of EELGs contains compact, high-SFR objects with high  $H\beta$  and  $[O\text{ III}]$  equivalent widths. The selected objects have on average very high values of  $[O\text{ III}]/H\beta$  and low values of  $[N\text{ II}]/H\alpha$ , so they are close to the separation curve between star-forming and AGN galaxies in the BPT diagram. The EELG sample gathers analogs of high- $z$  galaxies such as green-pea galaxies (Cardamone et al. 2009) and their cohort described by Izotov et al. (2011).

Our main goal is to take advantage of the very similar properties of this sample with those of the starbursts at higher- $z$ . We derive their chemical abundances using the direct method and then provide empirical calibrations of the fluxes of their strong emission-lines to be applied to distant objects in the study of the evolution of the main scaling relations involving metallicity such as the MZR or the FMR.

A direct derivation of the total O/H and N/O could be carried out in around 64% of the sample, as for  $z < 0.02$  the  $[O\text{ II}] 3727 \text{ \AA}$  line lies outside the observed spectral range in SDSS. The average O/H turns out to be 0.3 times the solar value i.e. the  $3\sigma$  range around the average is  $7.7 < 12+\log(O/H) < 8.6$ , while the average N/O value is 0.36 times the solar value (i.e. covering the range  $-1.8 < \log(N/O) < -0.8$ ). Very small differences are found in the average chemical abundances when we restrict the sample to higher  $z$  or EW( $[O\text{ III}]$ ) values in the sample and these are consistent with an average higher SFR in these subsamples.

For those objects of the studied sample with a good estimation of the electron temperature but for which it is not possible to



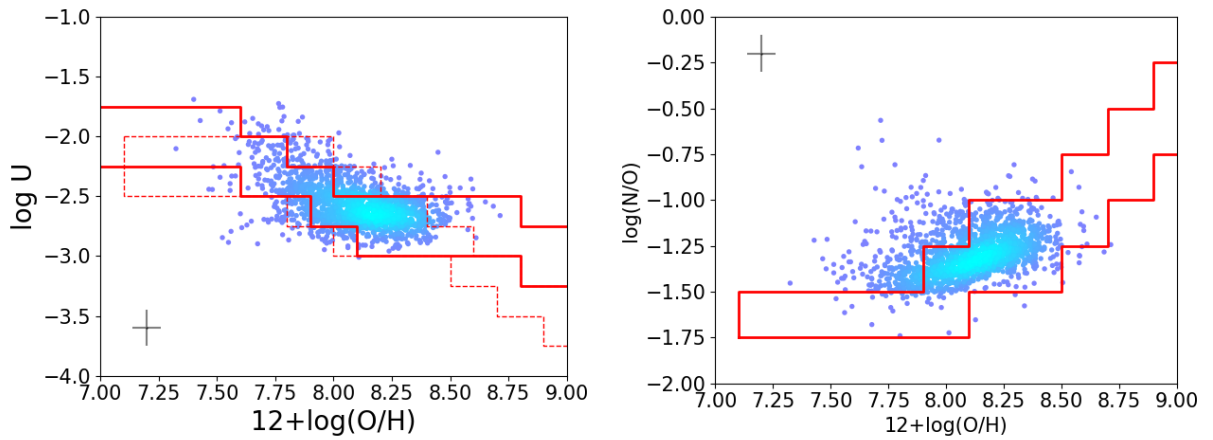
**Figure 10.** Comparison between chemical abundances from the direct method (DM) and those obtained from HCM using all lines for  $12+\log(\text{O}/\text{H})$  (left panel) and for  $\log(\text{N}/\text{O})$  (right panel). The solid red line represents the one-to-one relation. Crosses represent the mean errors.

**Table 1.** Mean errors, and mean ( $\Delta$ ) and standard deviation ( $\sigma$ ) of the residuals in dex units when comparing  $12+\log(\text{O}/\text{H})$  and  $\log(\text{N}/\text{O})$  derived from HCM and the direct method applied to the studied EELGs. The rows correspond to the different combinations of extinction-corrected relative-to- $\text{H}\beta$  emission lines used by HCM. The mean errors obtained by HCM in each case is also given. [O II] stands for  $\lambda 3727 \text{ \AA}$ , [Ne III] for  $\lambda 3869 \text{ \AA}$ , [O III]a for  $\lambda 4363 \text{ \AA}$ , [O III]n for  $\lambda 5007 \text{ \AA}$ , [N II] for  $\lambda 6583 \text{ \AA}$ , and [S II] for  $\lambda 6717+6731 \text{ \AA}$ .

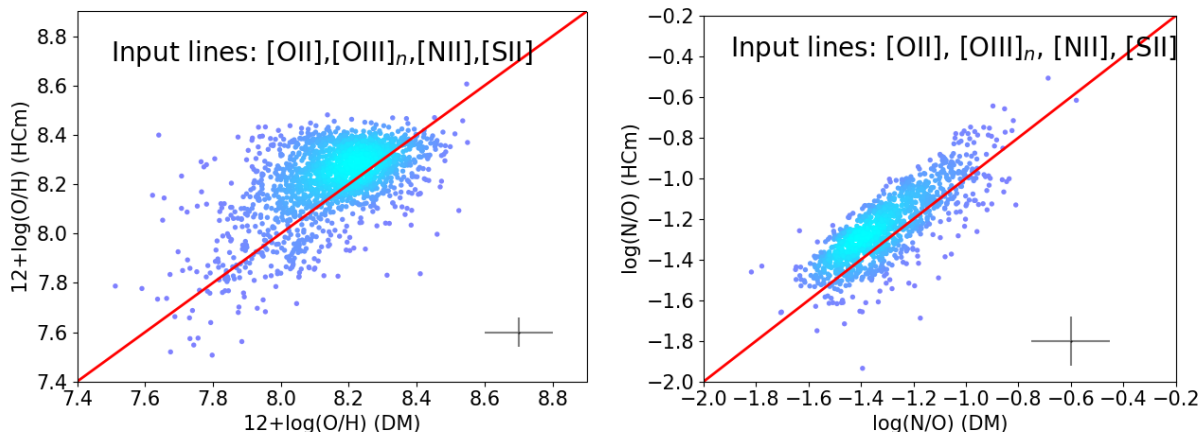
Set of input emission lines	Mean O/H error	$\Delta(\text{O}/\text{H})$	$\sigma(\text{O}/\text{H})$	Mean N/O error	$\Delta(\text{N}/\text{O})$	$\sigma(\text{N}/\text{O})$
All lines	0.07	+0.00	0.08	0.13	-0.01	0.08
[O III]a, [O III]n, [N II], [S II]	0.06	-0.04	0.09	0.17	+0.08	0.08
[O II], [O III]n, [N II], [S II] <sup>a</sup>	0.05	+0.06	0.14	0.03	+0.06	0.11
[O II],[O III]n, [N II] <sup>a</sup>	0.05	+0.08	0.12	0.03	-0.05	0.20
[O III]n, [N II], [S II] <sup>a</sup>	0.04	+0.06	0.15	0.01	+0.07	0.11
[N II], [S II] <sup>a</sup>	0.06	+0.16	0.16	0.01	+0.07	0.11
[O III]n, [N II] <sup>a,b</sup>	0.03	+0.10	0.12	-	-	-
[N II] <sup>a,b</sup>	0.03	+0.10	0.12	-	-	-
[O II], [O III]n <sup>a,b</sup>	0.10	+0.09	0.14	-	-	-
[O II], [Ne III] <sup>a,b</sup>	0.05	-0.04	0.16	-	-	-

<sup>a</sup> Assuming an empirical relation between O/H and  $U$

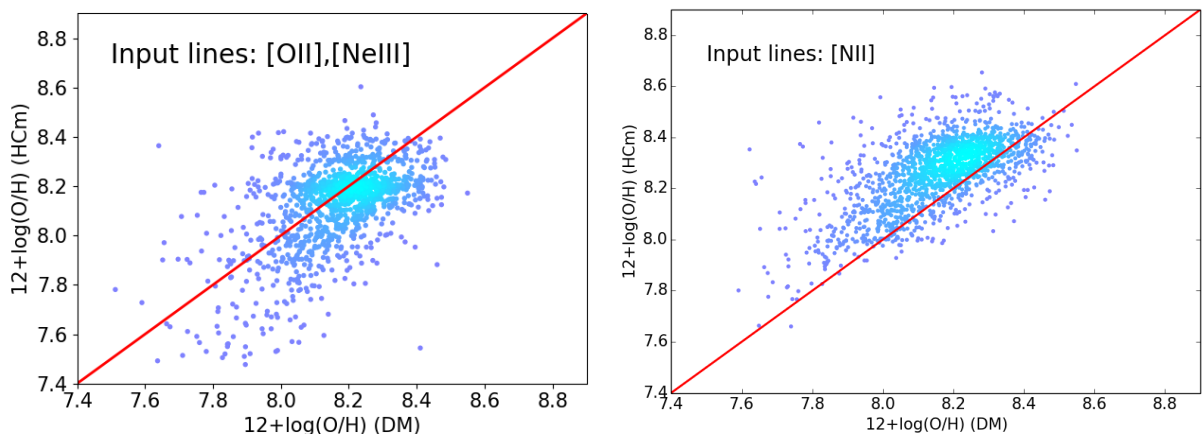
<sup>b</sup> Assuming an empirical relation between O/H and N/O



**Figure 11.** Relation between  $\log U$  (left panel) and  $\log(\text{N}/\text{O})$  (right panel) and  $12+\log(\text{O}/\text{H})$  calculated using HCM for those EELGs having [O III]  $\lambda 4363 \text{ \AA}$  emission line. In both panels, the area encompassed by the solid red lines represent the set of models considered by the code in the absence of [O III]  $\lambda 4363 \text{ \AA}$  or the emission-line ratios N2O2 and N2S2. The dashed red line in left panel shows the assumption made in the case of the star-forming objects described in Pérez-Montero (2014). Crosses represent the mean errors.



**Figure 12.** Comparison between chemical abundances from the direct method and those obtained from HCM using all lines except  $\text{O III } \lambda 4363 \text{ \AA}$ . The left and right panels correspond to  $12+\log(\text{O}/\text{H})$  and  $\log(\text{N}/\text{O})$ , respectively. The solid red line represents the one-to-one relation. Crosses represent the mean errors estimated by HCM.



**Figure 13.** Comparison between  $12+\log(\text{O}/\text{H})$  from the direct method and those obtained from HCM using different sets of emission lines:  $[\text{O II}]$  and  $[\text{Ne III}]$ , in the left panel, and  $[\text{N II}]$  in the right panel, and assuming a relation between  $\text{N}/\text{O}$  and  $\text{O}/\text{O}$ . The solid red line represents the one-to-one relation. Crosses represent the mean errors estimated by HCM.

calculate the total oxygen abundance because  $[\text{O II}] 3727 \text{ \AA}$  is not observed, we applied an empirical relation between  $t([\text{O III}])$  and  $\text{O}/\text{H}$  yielding an uncertainty in  $12+\log(\text{O}/\text{H})$  of only 0.04 dex.

We then explored the empirical calibrations of oxygen abundance for the most commonly used strong-line ratios, in our EELG sample, including  $\text{O32}$ ,  $\text{Ne3O2}$ ,  $\text{N2}$ ,  $\text{O3N2}$ ,  $\text{S23}$ , and  $\text{S3O3}$ . It was not possible to provide calibration based on  $\text{R23}$  because our EELGs are in the so-called turnover region where the scatter of  $\text{R23}$  with  $\text{O}/\text{H}$  is large. Regarding  $\text{O32}$  or  $\text{Ne3O2}$ , our linear calibrations present very similar properties to those already provided for intermediate redshift galaxies (e.g. Jones et al. 2015).

In the case of strong-line methods based on  $[\text{N II}]$  lines, such as  $\text{N2}$  or  $\text{O3N2}$ , no robust calibration can be given for objects with  $12+\log(\text{O}/\text{H}) < 7.8$ , likely owing to the very high values of  $\text{N}/\text{O}$  derived in this sample in this regime. For higher  $Z$ , our calibrations differ from others based on local samples, possibly because of differences in the excitation properties.

We also provide empirical calibrations for  $\text{N}/\text{O}$  using the emission-line ratios  $\text{N2O2}$  and  $\text{N2S2}$ , that lead to good determinations of  $\text{N}/\text{O}$  with uncertainties of less than 0.10 dex. As neither

$\text{N2O2}$  nor  $\text{N2S2}$  sensibly depend on  $U$ , the derived linear calibrations for EELGs are not substantially different from those for other star-forming objects reported by Pérez-Montero & Contini (2009) or the fits based on models for high-redshift galaxies given by Strom et al. (2018).

We also investigated the behaviour of other strong-line parameters based on  $[\text{S III}]$  (i.e.  $\text{S23}$ ,  $\text{S3O3}$ ) for a subsample of 335 objects with  $z < 0.02$ , and we found that the correlation is much better than that for other parameters. In any case, no significant difference is found with respect to other calibrations based on other local samples (e.g. Pérez-Montero & Díaz 2005; Stasińska 2006), suggesting that these calibrations could be based on samples biased towards objects with extreme properties i.e., compatible with those of EELGs. On the other hand, the lack of correlation between  $\text{O}/\text{H}$  and the emission-line ratio  $[\text{S III}]/[\text{S II}]$ , which is strongly dependent on  $U$ , may indicate that the previously observed relations with other high-to-low excitation emission-line ratios (such as  $\text{O32}$  or  $\text{Ne3O2}$ ) results from the dependence of  $\text{O}/\text{H}$  on the hardness of the ionizing radiation rather than on  $U$ .

Finally, we provide and discuss a version of the code HII-CHI-

MISTRY adapted for the studied sample of EELGs using stellar model atmospheres from Eldridge et al. (2017). The agreement between the derived O/H and N/O from the models as compared to those obtained from the direct method are within the errors provided by the code, even in absence of the [O II]  $\lambda 3727$  Å line, although with a slight offset for very low values of  $Z$  that must be further investigated. The  $U$  in EELGs derived using HCM is slightly larger than the sample of local star-forming galaxies analyzed in Pérez-Montero (2014), so the empirical relation between O/H and  $U$  used by the code when [O III]  $\lambda 4363$  Å is not detected had to be re-evaluated. Given the high dispersion shown by this sample in the O/H vs N/O diagram, a previous determination of N/O to derive oxygen abundances based on [N II] lines is especially relevant.

The dispersions for different sets of emission-lines are given in Table 1 and can be consistently used within errors for different sets of galaxies at different redshift ranges in agreement with the direct method both for O/H and N/O. A caveat is in order. The error bars provided by HCM are estimative and should be handled with caution since they do not include sources of systematic errors that may be significant.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for constructing and helpful comments. This work has been partly funded by projects Estallidos6 AYA2016-79724-C4 (Spanish Ministerio de Economía y Competitividad), “Estallidos7 PID2019-107408GB-C44 (Spanish Ministerio de Ciencia e Innovación), and the Junta de Andalucía for grant EXC/2011 FQM-7058. This work has been also supported by the Spanish Science Ministry "Centro de Excelencia Severo Ochoa Program under grant SEV-2017-0709. RA acknowledges support from ANID FONDECYT Regular Grant 1202007. JSA acknowledges support from the Spanish Ministry of Science and Innovation, project PID2019-107408GB-C43 (ESTALLIDOS), and from Gobierno de Canarias through EU FEDER funding, project PID2020010050. RGB acknowledges support from grants PID2019-109067GB-I00 (Spanish Ministerio de Ciencia e Innovación) and P18-FRJ-2595 (Junta de Andalucía). EPM also acknowledges the assistance from his guide dog Rocko without whose daily help this work would have been much more difficult. Funding for SDSS, SDSS-II, and SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science.

## DATA AVAILABILITY

The data underlying this article are available in the SDSS-DR7 database at <http://classic.sdss.org/dr7/>. Data products for the EELG sample used in this article will be shared on reasonable request to the corresponding author.

## REFERENCES

Aguerri J. A. L., Huertas-Company M., Sánchez Almeida J., Muñoz-Tuñón C., 2012, *A&A*, **540**, A136  
 Ahumada R., et al., 2019, arXiv e-prints, p. arXiv:1912.02905  
 Alloin D., Collin-Souffrin S., Joly M., Vigroux L., 1979, *A&A*, **78**, 200  
 Amorín R. O., Pérez-Montero E., Vílchez J. M., 2010, *ApJ*, **715**, L128

Amorín R., Pérez-Montero E., Vílchez J. M., Papaderos P., 2012, *ApJ*, **749**, 185  
 Amorín R., et al., 2014, *A&A*, **568**, L8  
 Amorín R., et al., 2015, *A&A*, **578**, A105  
 Ascasibar Y., Sánchez Almeida J., 2011, *MNRAS*, **415**, 2417  
 Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, **47**, 481  
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, **93**, 5  
 Bian F., Kewley L. J., Dopita M. A., 2018, *ApJ*, **859**, 175  
 Bian F., Kewley L. J., Groves B., Dopita M. A., 2020, *MNRAS*, **493**, 580  
 Bishop C., 2006, Pattern recognition and machine learning  
 Blanc G. A., Kewley L., Vogt F. P. A., Dopita M. A., 2015, *ApJ*, **798**, 99  
 Brinchmann J., Pettini M., Charlot S., 2008, *MNRAS*, **385**, 769  
 Brisbin D., Harwit M., 2012, *ApJ*, **750**, 142  
 Bruzual G., Charlot S., 2003, *MNRAS*, **344**, 1000  
 Calabrò A., et al., 2017, *A&A*, **601**, A95  
 Cardamone C., et al., 2009, *MNRAS*, **399**, 1191  
 Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, **345**, 245  
 Chabrier G., 2003, *ApJ*, **586**, L133  
 Cid Fernandes R., Gu Q., Melnick J., Terlevich E., Terlevich R., Kunth D., Rodrigues Lacerda R., Joguet B., 2004, *MNRAS*, **355**, 273  
 Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, *MNRAS*, **358**, 363  
 Cresci G., Mannucci F., Maiolino R., Marconi A., Gnerucci A., Magrini L., 2010, *Nature*, **467**, 811  
 Daddi E., et al., 2005, *ApJ*, **626**, 680  
 Díaz A. I., Pérez-Montero E., 2000, *MNRAS*, **312**, 130  
 Diaz A. I., Terlevich E., Vilchez J. M., Pagel B. E. J., Edmunds M. G., 1991, *MNRAS*, **253**, 245  
 Dopita M. A., Binette L., Dodorico S., Benvenuti P., 1984, *ApJ*, **276**, 653  
 Dopita M. A., Kewley L. J., Heisler C. A., Sutherland R. S., 2000, *ApJ*, **542**, 224  
 Dopita M. A., Sutherland R. S., Nicholls D. C., Kewley L. J., Vogt F. P. A., 2013, *The Astrophysical Journal Supplement Series*, **208**, 10  
 Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, *Ap&SS*, **361**, 61  
 Dors Jr. O. L., Copetti M. V. F., 2003, *A&A*, **404**, 969  
 Dors O. L., Hägele G. F., Cardaci M. V., Krabbe A. C., 2017, *MNRAS*, **466**, 726  
 Edmunds M. G., 1990, *MNRAS*, **246**, 678  
 Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, *Publ. Astron. Soc. Australia*, **34**, e058  
 Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, *ApJ*, **672**, L107  
 Everitt B., 1995, Cluster analysis  
 Ferland G. J., et al., 2017, *Rev. Mex. Astron. Astrofis.*, **53**, 385  
 Freeman P., Doe S., Siemiginowska A., 2001, Sherpa: a mission-independent data analysis application. pp 76–87, doi:10.1117/12.447161  
 García-Rojas J., Esteban C., 2007, *ApJ*, **670**, 457  
 Guseva N. G., et al., 2020, *MNRAS*, **497**, 4293  
 Hayashi M., et al., 2015, *PASJ*, **67**, 80  
 Hoopes C. G., et al., 2007, *ApJS*, **173**, 441  
 Huertas-Company M., Aguerri J. A. L., Bernardi M., Mei S., Sánchez Almeida J., 2011, *A&A*, **525**, A157  
 Izotov Y. I., Guseva N. G., Thuan T. X., 2011, *ApJ*, **728**, 161  
 Jones T., Martin C., Cooper M. C., 2015, *ApJ*, **813**, 126  
 Kaasinen M., Bian F., Groves B., Kewley L. J., Gupta A., 2017, *MNRAS*, **465**, 3220  
 Kashino D., et al., 2017, *ApJ*, **843**, 138  
 Kauffmann G., et al., 2003, *MNRAS*, **346**, 1055  
 Kewley L. J., Dopita M. A., 2002, *ApJS*, **142**, 35  
 Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, **556**, 121  
 Kewley L. J., Geller M. J., Jansen R. A., Dopita M. A., 2002, *AJ*, **124**, 3135  
 Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013, *ApJ*, **774**, 100  
 Khochfar S., Silk J., 2011, *MNRAS*, **410**, L42

- Kisielius R., Storey P. J., Ferland G. J., Keenan F. P., 2009, *MNRAS*, **397**, 903
- Kojima T., Ouchi M., Nakajima K., Shibuya T., Harikane Y., Ono Y., 2017, *PASJ*, **69**, 44
- Köppen J., Hensler G., 2005, *A&A*, **434**, 531
- Lamareille F., et al., 2009, *A&A*, **495**, 53
- Lara-López M. A., et al., 2010, *A&A*, **521**, L53
- Le Borgne J. F., et al., 2003, *A&A*, **402**, 433
- Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, *A&A*, **80**, 155
- Liang G. Y., Zhao G., Zeng J. L., Shi J. R., 2006, *J. Quant. Spectrosc. Radiative Transfer*, **102**, 473
- Loaiza-Agudelo M., Overzier R. A., Heckman T. M., 2020, *ApJ*, **891**, 19
- Maier C., Lilly S. J., Ziegler B. L., Contini T., Pérez Montero E., Peng Y., Balestra I., 2014, *ApJ*, **792**, 3
- Maiolino R., Mannucci F., 2019, *A&ARv*, **27**, 3
- Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, *MNRAS*, **408**, 2115
- Marino R. A., et al., 2013, *Astronomy & Astrophysics*, **559**, A114
- Masters D., et al., 2014, *ApJ*, **785**, 153
- Mathis J. S., 1985, *ApJ*, **291**, 247
- Mollá M., García-Vargas M. L., Bressan A., 2009, *MNRAS*, **398**, 451
- Morales-Luis A. B., Pérez-Montero E., Sánchez Almeida J., Muñoz-Tuñón C., 2014, *ApJ*, **797**, 81
- Morisset C., et al., 2016, *A&A*, **594**, A37
- Nakajima K., Ouchi M., 2014, *MNRAS*, **442**, 900
- Page B. E. J., Edmunds M. G., Blackwell D. E., Chun M. S., Smith G., 1979, *MNRAS*, **189**, 95
- Peimbert M., Costero R., 1969, *Boletín de los Observatorios Tonantzintla y Tacubaya*, **5**, 3
- Peimbert M., Sarmiento A., Fierro J., 1991, *PASP*, **103**, 815
- Pérez-Montero E., 2014, *MNRAS*, **441**, 2663
- Pérez-Montero E., 2017, *PASP*, **129**, 043001
- Pérez-Montero E., Amorín R., 2017, *MNRAS*, **467**, 1287
- Pérez-Montero E., Contini T., 2009, *MNRAS*, **398**, 949
- Pérez-Montero E., Díaz A. I., 2005, *MNRAS*, **361**, 1063
- Pérez-Montero E., Hägele G. F., Contini T., Díaz Á. I., 2007, *MNRAS*, **381**, 125
- Pérez-Montero E., et al., 2013, *A&A*, **549**, A25
- Pérez-Montero E., et al., 2016, *A&A*, **595**, A62
- Pérez-Montero E., García-Benito R., Vílchez J. M., 2019a, *MNRAS*, **483**, 3322
- Pérez-Montero E., Dors O. L., Vílchez J. M., García-Benito R., Cardaci M. V., Hägele G. F., 2019b, *MNRAS*, **489**, 2652
- Pettini M., Page B. E. J., 2004, *MNRAS*, **348**, L59
- Pilyugin L. S., Grebel E. K., 2016, *MNRAS*, **457**, 3678
- Pilyugin L. S., Vílchez J. M., Mattsson L., Thuan T. X., 2012, *MNRAS*, **421**, 1624
- Sánchez Almeida J., Dalla Vecchia C., 2018, *ApJ*, **859**, 109
- Sánchez Almeida J., Aguerri J. A. L., Muñoz-Tuñón C., de Vicente A., 2010, *ApJ*, **714**, 487
- Sánchez Almeida J., Aguerri J. A. L., Muñoz-Tuñón C., Huertas-Company M., 2011, *ApJ*, **735**, 125
- Sánchez Almeida J., Terlevich R., Terlevich E., Cid Fernandes R., Morales-Luis A. B., 2012, *ApJ*, **756**, 163
- Sanders R. L., et al., 2018, *ApJ*, **858**, 99
- Sanders R. L., et al., 2020, *MNRAS*, **491**, 1427
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *The Astrophysical Journal*, **500**, 525
- Shapley A. E., Coil A. L., Ma C.-P., Bundy K., 2005, *The Astrophysical Journal*, **635**, 1006
- Shapley A. E., et al., 2015, *ApJ*, **801**, 88
- Shirazi M., Brinchmann J., Rahmati A., 2014, *ApJ*, **787**, 120
- Stasińska G., 2006, *A&A*, **454**, L127
- Steidel C. C., Strom A. L., Pettini M., Rudie G. C., Reddy N. A., Trainor R. F., 2016, *ApJ*, **826**, 159
- Storchi-Bergmann T., Calzetti D., Kinney A. L., 1994, *ApJ*, **429**, 572
- Storey P. J., Sochi T., Badnell N. R., 2014, *MNRAS*, **441**, 3028
- Strom A. L., Steidel C. C., Rudie G. C., Trainor R. F., Pettini M., 2018, *ApJ*, **868**, 117
- Terlevich R., Melnick J., Masegosa J., Moles M., Copetti M. V. F., 1991, *A&AS*, **91**, 285
- Thurston T. R., Edmunds M. G., Henry R. B. C., 1996, *MNRAS*, **283**, 990
- Tremonti C. A., et al., 2004, *ApJ*, **613**, 898
- Tsamis Y. G., Péquignot D., 2005, *MNRAS*, **364**, 687
- Vale Asari N., Stasińska G., Morisset C., Cid Fernandes R., 2016, *MNRAS*, **460**, 1739
- Vílchez J. M., Esteban C., 1996, *MNRAS*, **280**, 720
- Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz M. J., 2013, *ApJ*, **771**, L19

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