



<b>Publication Year</b>	2017
<b>Acceptance in OA</b>	2021-01-20T16:34:54Z
<b>Title</b>	New fully empirical calibrations of strong-line metallicity indicators in star-forming galaxies
<b>Authors</b>	Curti, M., CRESCI, GIOVANNI, MANNUCCI, FILIPPO, Marconi, Alessandro, Maiolino, R., ESPOSITO, Simone
<b>Publisher's version (DOI)</b>	10.1093/mnras/stw2766
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/29888">http://hdl.handle.net/20.500.12386/29888</a>
<b>Journal</b>	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY
<b>Volume</b>	465

# New fully empirical calibrations of strong-line metallicity indicators in star-forming galaxies

M. Curti,<sup>1,2★</sup> G. Cresci,<sup>2</sup> F. Mannucci,<sup>2</sup> A. Marconi,<sup>1,2</sup> R. Maiolino<sup>3,4</sup>  
and S. Esposito<sup>2</sup>

<sup>1</sup>Dipartimento di Fisica e Astronomia, Università di Firenze, Via G. Sansone 1, I-50019 Sesto Fiorentino (Firenze), Italy

<sup>2</sup>INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy

<sup>3</sup>Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge CB3 0HE, UK

<sup>4</sup>Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

Accepted 2016 October 25. Received 2016 September 15; in original form 2016 April 29

## ABSTRACT

We derive new empirical calibrations for strong-line diagnostics of gas-phase metallicity in local star-forming galaxies by uniformly applying the  $T_e$  method over the full metallicity range probed by the Sloan Digital Sky Survey (SDSS). To measure electron temperatures at high metallicity, where the auroral lines needed are not detected in single galaxies, we stacked spectra of more than 110 000 galaxies from the SDSS in bins of  $\log[\text{O II}]/\text{H}\beta$  and  $\log[\text{O III}]/\text{H}\beta$ . This stacking scheme does not assume any dependence of metallicity on mass or star formation rate, but only that galaxies with the same line ratios have the same oxygen abundance. We provide calibrations which span more than 1 dex in metallicity and are entirely defined on a consistent absolute  $T_e$  metallicity scale for galaxies. We apply our calibrations to the SDSS sample and find that they provide consistent metallicity estimates to within 0.05 dex.

**Key words:** ISM: abundances – galaxies: abundances – galaxies: evolution – galaxies: ISM.

## 1 INTRODUCTION

The accurate determination of gas-phase metallicity represents a challenging topic for studies that aim at understanding the chemical evolution of galaxies. The metal content of a galaxy is regulated by complex interactions between physical processes occurring on different spatial and time-scales: heavy elements produced by stellar activity contribute to the enrichment of the interstellar medium (ISM), while cosmological infall of pristine gas from the intergalactic medium (IGM) and outflows due to active galactic nuclei (AGN) and supernovae could dilute the ISM and at the same time trigger new star formation episodes (Davé, Finlator & Oppenheimer 2011). These processes directly impact the global baryon cycle and thus affect other physical quantities strictly related to the history of chemical enrichment in galaxies like stellar mass ( $M_*$ ) and star formation rate (SFR); therefore, relationships between these parameters and gas-phase metallicity are expected. Indeed, in the last decades strong observational evidences of a correlation between  $M_*$  and gas-phase metallicity (the so-called mass–metallicity relation,  $M-Z$ ) have been reported by several studies, both in the local Universe (e.g. Tremonti et al. 2004; Lee et al. 2006; Liang et al. 2007) and at higher redshift, where signatures of a cosmic evolution have been found (e.g. Erb et al. 2006; Maiolino et al. 2008; Mannucci et al. 2009; Cresci et al. 2012; Zahid et al. 2012; Troncoso et al. 2014). Furthermore, Mannucci et al. (2010) showed that the intrinsic scatter in the  $M-Z$

could be reduced when SFR is taken into account, introducing the concept of a fundamental metallicity relation (FMR) that reduces the  $M-Z$  to a two-dimensional projection of a three-dimensional surface. The FMR appears to be *more fundamental* in the sense that it does not seem to present clear signs of evolution up to  $z \sim 2.5$ . Even though the physical origin of these relations is still debated, the knowledge of the main properties of the  $M-Z$  and the exact form of its dependence upon the SFR is important to investigate the processes regulating star formation and to assess the role of outflows in ejecting metals out of the galaxy (Davé, Finlator & Oppenheimer 2011; Dayal, Ferrara & Dunlop 2013; Lilly et al. 2013); this could provide crucial observational constraints for models aimed at reproducing the chemical evolution of galaxies across cosmic time.

Investigating the properties of these relationships and their redshift evolution requires precise and robust metallicity estimates. Since the scatter in the FMR is of the order of 0.05 dex (Mannucci et al. 2010), such a level of precision in metallicity determination would be desirable. There are several ways to measure abundances in galaxies, but unfortunately none of them is considered reliable or applicable over the whole metallicity range covered by large galaxy samples. The most commonly used method relies on the determination of the electron temperature of the nebulae responsible for emission lines in galaxies: in fact, electron temperature is known to be strongly correlated with metallicity, such that higher metallicities are associated with lower  $T_e$ , because forbidden emission lines from metals are the primary coolants in H II regions. Electron temperatures can be inferred by exploiting the temperature sensitive auroral

\* E-mail: [mcurti@arcetri.astro.it](mailto:mcurti@arcetri.astro.it)

to nebular line ratios of particular ions (e.g.  $[\text{O III}] \lambda 4363/5007$  is one of the most widely used); in fact, the atomic structure of these ions is such that auroral and nebular lines originate from excited states that are well spaced in energy and thus their relative level populations depend heavily on electron temperature. This so-called  $T_e$  method is widely accepted as the preferred one to estimate abundances since it is a direct probe of the processes that regulate the physics of ionized nebulae. Unfortunately, auroral lines are weak in most of individual galaxy spectra, especially for metal-rich objects, which typically prevents the  $T_e$  from being used as a method to determine abundances of metal-enriched galaxies. A different technique is based instead on exploiting the ratio between oxygen and hydrogen recombination lines (RLs): since these lines show a very weak dependence on electron temperature and density (Esteban et al. 2009, 2014; Peimbert & Peimbert 2014), this is probably the most reliable method because it is unaffected by the typical biases of the  $T_e$  method associated with temperature fluctuations. Typical discrepancies between  $T_e$ - and RLs-based abundances are found to be of the order of 0.2–0.3 dex, with the first ones underestimating the latter (García-Rojas & Esteban 2007; Esteban et al. 2009). Recently, Bresolin et al. (2016) showed that metal RLs yield nebular abundances in excellent agreement with stellar abundances for high-metallicity systems, while in more metal-poor environments they tend to underestimate the stellar metallicities by a significant amount. However, RLs are extremely faint (even hundred times fainter than oxygen auroral lines) and cannot be detected in galaxies more distant than a few kiloparsecs (Peimbert et al. 2007). For this reason, different methods have been developed to measure abundances in faint, distant and high-metallicity galaxies. In particular, it is known that some line ratios between strong collisionally excited lines (CELs) and Balmer lines show a dependence on metallicity, which can be either directly motivated or indirectly related to other physical quantities (e.g. the ionization parameter). Thus, it has been proposed that these line ratios could be calibrated against the oxygen abundance and used as metallicity indicators for galaxies in which the application of the  $T_e$  method is not possible due to the extreme faintness of auroral lines (Alloin et al. 1979; Pagel et al. 1979): these are referred to as the strong-line methods. Calibrations can be obtained either empirically, for samples in which metallicity have been previously derived with the  $T_e$  method (e.g. Pettini & Pagel 2004; Pilyugin & Thuan 2005; Pilyugin, Vílchez & Thuan 2010b; Pilyugin, Grebel & Mattsson 2012; Marino et al. 2013; Pilyugin & Grebel 2016), or theoretically, in which oxygen abundance have been inferred via photoionization models (e.g. McGaugh 1991; Zaritsky, Kennicutt & Huchra 1994; Kewley & Dopita 2002; Kobulnicky & Kewley 2004; Tremonti et al. 2004; Dopita et al. 2013, 2016), or be a combination of the two. Unfortunately, comparisons among metallicities estimated through different calibrations reveal large discrepancies, even for the same sample of objects, with variations up to  $\sim 0.6$  dex (Kewley & Ellison 2008; Moustakas et al. 2010). In fact, theoretical calibrations are known to produce higher metallicity estimates with respect to empirical calibrations based on the  $T_e$  method. The origin of these discrepancies is still unclear, but they could be attributed on one hand to oversimplified assumptions made in most of the photoionization models, e.g. about the geometry of the nebulae and the age of the ionizing source (Moustakas et al. 2010) and on the other hand to temperature gradients and fluctuations that may cause an overestimate of the electron temperature and a consequent underestimate of the true metallicity with the  $T_e$  method (Peimbert 1967; Stasińska 2002, 2005). Great care is therefore needed when using composite calibrations built with different methods over differ-

ent metallicity ranges, due to the large uncertainties introduced on the absolute metallicity scale. Empirical calibrations are generally preferable because they are based on the  $T_e$  method abundance scale, which is directly inferred from observed quantities. Moreover, on the abundance scale based on photoionization models, the Milky Way, where abundances can be precisely measured, would represent a very peculiar galaxy, falling well below the  $M-Z$  defined by similar star-forming galaxies. The discrepancy is reduced by more empirical metallicity calibrations that provide lower abundances. At the same time, one of their main limitations is that they are often calibrated for samples of objects that do not properly cover all the galaxy parameters space; this means, for example, that empirical calibrations obtained from a sample of low excitation H II regions could give unreliable results when applied to global galaxy spectra. Recently, the application of integral field spectroscopy allowed to study galaxy properties in great detail and to extend the H II regions data base for compiling abundances in order to obtain calibrations based on the  $T_e$  method [e.g. Marino et al. 2013 for the Calar Alto Legacy Integral Field Area (CALIFA) survey]. However, self-consistent calibrations obtained from integrated galaxy spectra and covering the entire metallicity range are still scarce.

In this work we derive a set of new empirical calibrations for some of the most common strong-line metallicity indicators, thanks to a uniform application of the  $T_e$  method over the full metallicity range covered by Sloan Digital Sky Survey (SDSS) galaxies. We combined a sample of low-metallicity galaxies with  $[\text{O III}] \lambda 4363$  detection from the SDSS together with stacked spectra of more than 110 000 galaxies in bins of  $\log[\text{O II}]/\text{H}\beta - \log[\text{O III}]/\text{H}\beta$  that allowed us to detect and measure the flux of the crucial auroral lines needed for the application of the  $T_e$  method also at high metallicity. Other studies demonstrated the potentiality and reliability of the stacking technique (Liang et al. 2007; Andrews & Martini 2013; Brown, Martini & Andrews 2016); compared to these works, our approach differs in the sense that we do not rely on any assumption regarding the nature and the form of the relationships between metallicity, mass and SFR, but only on the hypothesis that oxygen abundance can be determined from a combination of  $[\text{O II}]$  and  $[\text{O III}]$  emission line ratios.

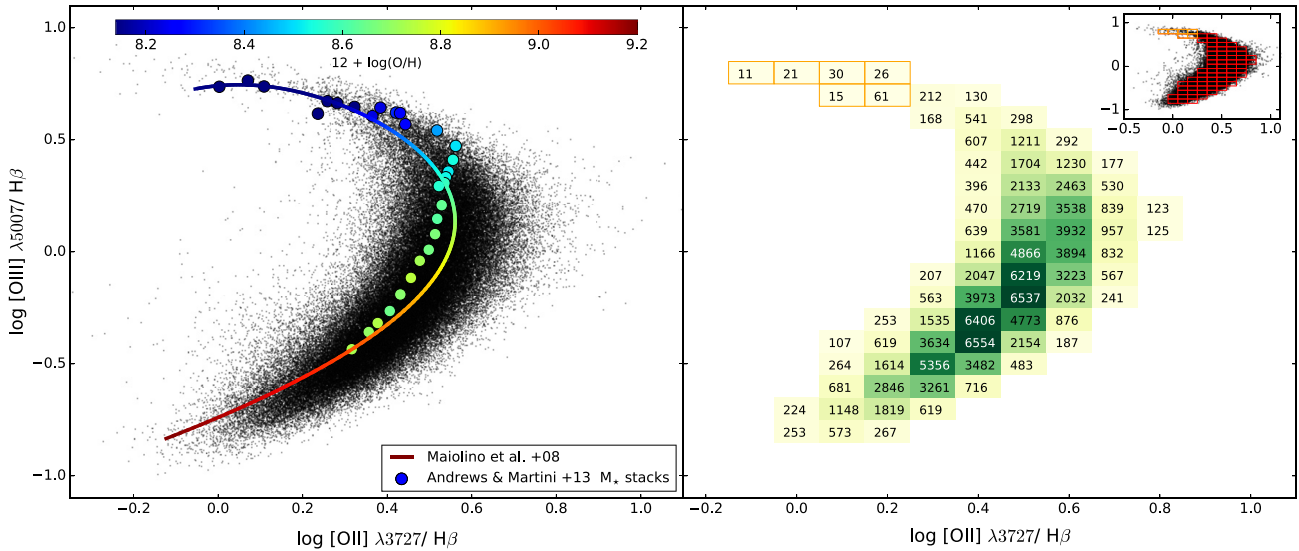
The paper is organized as follows. In Section 2 we describe the sample selection and the procedure used to stack the spectra, subtract the stellar continuum and fit the emission lines of interest. Section 3 describes the method we used to derive electron temperatures and chemical abundances. We then discuss the relations between different temperature diagnostics and between temperatures of different ionization zones. In Section 4 we report some tests we performed to verify the consistency of our hypothesis and stacking procedure. In Section 5, we present our new empirical calibrations for some of the most common strong-line abundance diagnostics and we compare them with previous ones from literature. We then apply them to the original SDSS sample as a test of their self-consistency. Section 6 summarizes our main results.

Publicly available tools to apply these new calibrations can be found on the web page <http://www.arcetri.astro.it/metallicity/>.

## 2 METHOD

### 2.1 Sample selection

Our galaxy sample comes from the SDSS Data Release 7 (DR7; Abazajian et al. 2009), a survey including  $\sim 930\,000$  galaxies in an area of  $8423 \text{ deg}^2$ . Emission line data have been taken from



**Figure 1.** Left-hand panel: the distribution of our galaxy sample in the  $\log [O\text{II}] \lambda 3727 / H\beta - \log [O\text{III}] \lambda 5007 / H\beta$  diagram. The curve represents the combined calibrations for the  $[O\text{II}] / H\beta$  and  $[O\text{III}] / H\beta$  metallicity indicators from Maiolino et al. (2008), colour coded by the metallicity inferred from the combination of the two indicators. The Andrews & Martini (2013) stacks of stellar mass are shown as circle points and colour coded for their direct metallicity measurement. Right-hand panel: stacking grid for our sample of SDSS galaxies in the  $\log [O\text{II}] \lambda 3727 / H\beta - \log [O\text{III}] \lambda 5007 / H\beta$  diagram. Each square represents a  $0.1 \times 0.1 \text{ dex}^2$  bin, colour coded by the number of galaxies included in it, which is also written for each bin. Orange boxes represent stacks of low-metallicity galaxies for which we relaxed the 100-object threshold in the definition of our grid. In the upper right box of the panel our stacking grid is shown superimposed on the distribution of galaxies in the diagram.

the MPA/JHU<sup>1</sup> catalogue, in which also stellar masses (Kauffmann et al. 2003a), SFRs (Brinchmann et al. 2004; Salim et al. 2007) and metallicities (Tremonti et al. 2004) are measured. We chose only galaxies with redshifts in the range  $0.027 < z < 0.25$ , to ensure the presence of the  $[O\text{II}] \lambda 3727$  emission line and of the  $[O\text{II}] \lambda \lambda 7320, 7330$  doublet within the useful spectral range of the SDSS spectrograph (3800–9200 Å). We selected only galaxies classified in the MPA/JHU as star forming, discarding galaxies dominated by AGN contribution according to criteria for BPT-diagram classification illustrated in Kauffmann et al. (2003b), in order to avoid any effect on the emission line ratios that could cause spurious metallicity measurements. We also used a signal-to-noise ratio (S/N) threshold of 5 on the  $H\alpha$ ,  $H\beta$ ,  $[O\text{III}] \lambda 5007$  and  $[O\text{II}] \lambda 3727$  emission line fluxes. After applying these selection criteria, the total number of galaxies in our sample was reduced to 118 478, with a median redshift of  $z = 0.072$ . At this redshift, the 3 arcsec diameter of the SDSS spectroscopic fibre corresponds to  $\sim 3$  kpc.

## 2.2 Stacking procedure

Our primary goal is to perform accurate measurements of galaxy metallicity in order to obtain more consistent calibrations for the main strong-line indicators, thanks to a uniform application of the  $T_e$  method. Unfortunately, in distant galaxies the  $[O\text{III}] \lambda 4363$  and  $[O\text{II}] \lambda \lambda 7320, 7330$  auroral lines are too weak to be detected in the individual spectra at metallicities higher than  $12 + \log(O/H) \gtrsim 8.3$ . Thus, we decided to stack spectra for galaxies that are expected to have similar metallicities.

Galaxies are stacked according to their values of reddening corrected  $[O\text{II}] \lambda 3727 / H\beta$  and  $[O\text{III}] \lambda 5007 / H\beta$  flux ratios. This is based on the assumption that the so-called strong-line methods can be used

to discriminate the metallicities of star-forming galaxies when multiple line ratios are simultaneously considered. We stress that we are not assuming that a particular combination of these line ratios, such as  $R_{23}$ , is related to metallicity, but only that galaxies with simultaneously the same values of both  $[O\text{III}] \lambda 5007 / H\beta$  and  $[O\text{II}] \lambda 5007 / H\beta$  have approximately the same oxygen abundance. In fact, these are the two line ratios directly proportional to the main ionization states of oxygen and are thus individually used as metallicity diagnostics (Nagao, Maiolino & Marconi 2006; Maiolino et al. 2008). Moreover, their ratio  $[O\text{III}] / [O\text{II}]$  is sensitive to the ionization parameter and it is also used as an indicator of oxygen abundance, especially in metal-enriched galaxies, due to the physical link between ionization and gas-phase metallicity (e.g. Nagao et al. 2006; Masters, Faisst & Capak 2016). This means that the location of a galaxy on the  $[O\text{II}] \lambda 3727 / H\beta - [O\text{III}] \lambda 5007 / H\beta$  diagram is primarily driven by the metal content and the ionization properties of galaxies. Since the scatter in a given line ratio at fixed metallicity is often regarded as driven by variations in the ionization parameter (e.g. Kewley & Dopita 2002; López-Sánchez et al. 2012; Blanc et al. 2015) our binning choice takes into account this possible source of scatter.

The left-hand panel of Fig. 1 shows the distribution of our selected SDSS galaxies in the  $\log [O\text{II}] \lambda 3727 / H\beta - \log [O\text{III}] \lambda 5007 / H\beta$  diagnostic diagram. We overplot the semi-empirical calibration of Maiolino et al. (2008) for the  $[O\text{II}] \lambda 5007 / H\beta$  and  $[O\text{III}] \lambda 5007 / H\beta$  indicators in order to better visualize how the position on the 2D diagram given by the combination of these line ratios represent a metallicity sequence. The curve, colour coded for the metallicity inferred from the combination of the two indicators, follows quite tightly the distribution of galaxies on the map, showing how metallicity increases from the upper left region of the diagram to the bottom left one. To further illustrate how metallicity varies along this diagram, we can also use the metallicity obtained with the  $T_e$  method from composite spectra in bins of stellar mass by Andrews & Martini (2013), whose stacks are shown as circled points in the left-hand panel of the figure. Also in this case we can recognize

<sup>1</sup> <http://www.mpa-garching.mpg.de/SDSS/DR7/>

a pattern in which their mass stacks, each point being representative of the line ratios measured from the associated composite spectra, increase monotonically in metallicity following the galaxy sequence on the diagram. Thus, both methods reveal a clean variation of oxygen abundance with location on the diagram, though being based on different and independent approaches; this strengthens the idea of using the combination of  $[\text{O II}]/\text{H}\beta$  and  $[\text{O III}]/\text{H}\beta$  as a metallicity indicator. We note that differences among metallicity values predicted by the Maiolino et al. (2008) calibrations and the Andrews & Martini (2013) stacks (with the first ones predicting higher abundances than the latter) are only due to the different abundance scale upon which the two methods are defined, being the Maiolino et al. (2008) indicators calibrated with photoionization models at high metallicities and the Andrews & Martini (2013) stacks based on  $T_e$  method metallicities.

In Section 4 we test our assumptions by comparing  $T_e$  metallicities inferred from single galaxy spectra belonging to the same bin; this allows also to evaluate the main issues related to the stacking technique (see also the discussion in Section 3.2). We refer to these sections for an exhaustive discussion on this topic.

We thus created stacked spectra in bins of 0.1 dex of  $\log [\text{O II}]/\text{H}\beta$  and  $\log [\text{O III}]/\text{H}\beta$ . The choice of the 0.1 dex width in the binning grid represents a good compromise between keeping a high enough number of galaxy in each bin to ensure auroral line detection and at the same time avoid wider bins in which we could have mixed objects with too different properties. We performed some tests in stacking spectra and computing oxygen abundance with different bin sizes, finding no relevant differences.

We adopted the emission line values provided by the MPA/JHU catalogue to create the set of galaxy stacks. All line fluxes have been corrected for Galactic reddening, adopting the extinction law from Cardelli, Clayton & Mathis (1989) and assuming an intrinsic ratio for the Balmer lines  $\text{H}\alpha/\text{H}\beta = 2.86$  (as set by case B recombination theory for typical nebular temperatures of  $T_e = 10\,000$  K and densities of  $n_e \approx 100\text{ cm}^{-3}$ ).

In the right-hand panel of Fig. 1, the binning grid for our galaxy sample in the space defined by  $\log([\text{O II}] \lambda 3727/\text{H}\beta)$  and  $\log([\text{O III}] \lambda 5007/\text{H}\beta)$  is shown, colour coded by the number of objects in each bin. In the upper right-hand corner of the figure we show the distribution of the galaxy sample in the diagnostic diagram, with our binning grid superimposed. In the construction of our binning grid we required a minimum of 100 sources per bin: this was a conservative choice in order to average enough galaxy spectra to ensure the required S/N (i.e. at least 3) on auroral line detection after the stacking procedure. Since we imposed a threshold of 100 sources per bin, the upper left-hand corner of the diagram, occupied by the galaxies of lower metallicity in the sample, is not well covered by our stacking grid. For this reason, we extended our grid to include also low-metallicity galaxies by reducing the threshold to 10 sources in that area of the diagram, enough to detect auroral lines in stacked spectra with an S/N higher than 3 in this metallicity regime. This extension of the grid is marked with orange borders in the figure. This allows our grid to entirely cover the region occupied by SDSS galaxies, probing the largest possible combination of physical parameters in the sample. Throughout this paper we will refer to a particular stack by indicating the centre of the corresponding bin in both the line ratios considered (e.g. 0.5; 0.2 corresponds to the bin centred in  $\log[\text{O II}] \lambda 3727/\text{H}\beta = 0.5$  and  $\log[\text{O III}] \lambda 5007/\text{H}\beta = 0.2$ ).

Before creating the composite spectrum from galaxies belonging to the same bin, each individual spectrum has been corrected for reddening with a Cardelli et al. (1989) extinction law and normalized

to the extinction corrected  $\text{H}\beta$ . We have verified that the final results do not depend on the choice of the extinction law, by alternatively using the Calzetti, Kinney & Storchi-Bergmann (1994) extinction law in a few random bins. Then, each spectrum has been re-mapped on to a linear grid (3000–9200 Å), with wavelength steps of  $\Delta\lambda = 0.8$  Å, and shifted at the same time to the rest frame to compensate for the intrinsic redshift of the sources. This procedure may cause a redistribution of the flux contained in a single input channel to more than one output channel; in order to take into account this effect, the incoming flux is weighted on the overlap area between the input and output channels. Finally, to create the stacked spectra we took the mean pixel by pixel between the 25th and the 75th percentile of the flux distribution in each wavelength bin; in this way we could avoid biases introduced by the flux distribution asymmetry clearly visible in every flux channel as a right-end tail.

### 2.3 Stellar continuum subtraction

Stacking the spectra improves significantly the S/N of the auroral lines, but we must also fit and subtract the stellar continuum to accurately measure their fluxes. To perform the stellar continuum fit and subtraction on our stacked spectra, we have created a synthetic spectrum using the MIUSCAT library of spectral templates (Ricciardelli et al. 2012; Vazdekis et al. 2012), an extension of the previous MILES library (Cenarro et al. 2001; Sánchez-Blázquez et al. 2006; Vazdekis et al. 2010; Falcón-Barroso et al. 2011) in which both Indo-US and CaT libraries have been added to fill the gaps in wavelength coverage. The new MIUSCAT library covers a wavelength range of 3525–9469 Å, although the useful spectral window for this work is entirely covered by MILES templates, whose resolution is 2.51 Å [full width at half-maximum (FWHM)]. Stellar templates have been retrieved from the MILES web site<sup>2</sup> for a wide range of ages and metallicities, assuming a unimodal initial mass function (IMF) with a 1.3 slope (i.e. a Salpeter IMF). The stellar continuum subtraction in the  $[\text{S II}] \lambda 4069$  spectral window (close to  $\text{H}\delta$ ) has been performed using a different kind of stellar templates, the PÉGASE HR<sup>3</sup> (Le Borgne et al. 2004), a library which covers a wavelength range of 4000–6800 Å with a spectral resolution of  $R = 10\,000$  at  $\lambda = 5500$  Å; this allowed a better stellar continuum fit in the proximity of the  $[\text{S II}] \lambda 4069$  auroral line. To further improve emission line fluxes measurements, stellar continuum fits and subtractions have been performed selecting subregions of the spectrum centred on the lines of interest, each subregion being large a few hundred angstrom. During the procedure the location of the emission lines has been masked out in order to prevent the fit to be affected by non-stellar features. We performed the fit exploiting the IDL version of the penalized pixel-fitting (pPXF) procedure by Cappellari & Emsellem (2004). In Table 1 is reported, for each emission lines whose flux have been measured in this work, the spectral window of the stellar continuum fit and the wavelength range that has been masked out. Fig. 2 shows examples of the results of the stacking procedure and stellar continuum subtraction for the 0.5; 0.5 bin, in particular in spectral windows including  $\text{H}\beta$ ,  $[\text{O III}]$ ,  $\text{H}\alpha$ ,  $[\text{N II}]$  and  $[\text{S II}]$  nebular lines and  $[\text{O III}] \lambda 4363$ ,  $[\text{O II}] \lambda \lambda 7320, 7330$  auroral lines, respectively. For latter emission lines, a single galaxy spectrum from the same stack is shown for comparison, to underline the dramatic increase in S/N that allows to reveal the otherwise invisible auroral

<sup>2</sup> <http://miles.iac.es>

<sup>3</sup> <http://www2.iap.fr/pegase/>

**Table 1.** Spectral windows and mask ranges of measured emission lines. Column (1): emission lines; column (2): wavelength range of stellar continuum fit; column (3): spectral range that was masked out.

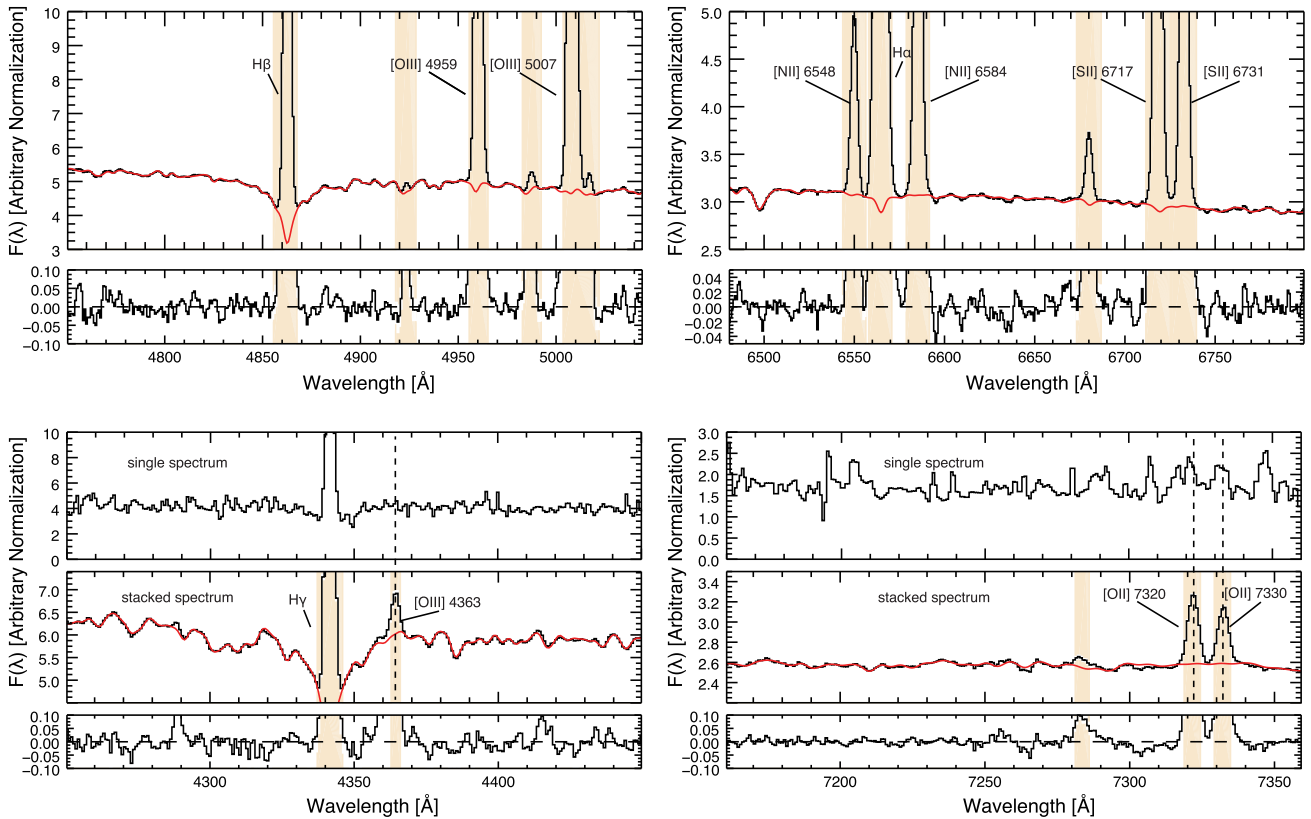
Line	Fit range (Å)	Mask range (Å)
(1)	(2)	(3)
[O II] $\lambda$ 3727	3650–3830	3723.36–3733.60
[Ne III] $\lambda$ 3870	3850–4150	3866.29–3874.03
[S II] $\lambda$ 4069	4000–4150	4068.39–4071.11
H $\delta$ $\lambda$ 4102	3850–4150	4098.79–4107.00
H $\gamma$ $\lambda$ 4340	4250–4450	4337.34–4346.03
[O III] $\lambda$ 4363	4250–4450	4362.98–4365.89
H $\beta$ $\lambda$ 4861	4750–5050	4857.82–4867.55
[O III] $\lambda$ 4960	4750–5050	4955.33–4965.26
[O III] $\lambda$ 5007	4750–5050	5003.23–5013.25
[N II] $\lambda$ 5756	5650–5850	5754.32–5758.16
[N II] $\lambda$ 6549	6480–6800	6543.30–6556.40
H $\alpha$ $\lambda$ 6563	6480–6800	6558.05–6571.17
[N II] $\lambda$ 6584	6480–6800	6578.69–6591.87
[S II] $\lambda$ 6717	6480–6800	6711.57–6725.01
[S II] $\lambda$ 6731	6480–6800	6725.94–6739.40
[O II] $\lambda$ 7320	7160–7360	7318.50–7323.28
[O II] $\lambda$ 7330	7160–7360	7329.24–7334.12

lines. The orange regions mark the spectral range masked out from the stellar fitting around nebular lines. In each plot, the lower panel shows the residual spectrum of the fit.

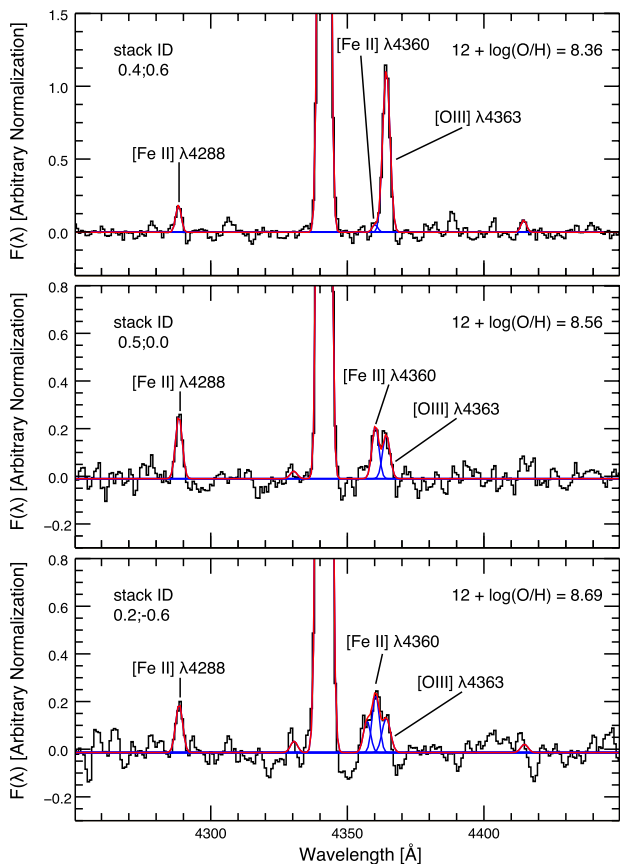
#### 2.4 Line flux measurement and iron contamination of [O III] $\lambda$ 4363 auroral line at high metallicity

We fit emission lines with a single Gaussian profile, fixing velocities and widths of the weak auroral lines by linking them to the strongest line of the same spectral region. For doublets, we fixed the velocity width of the weaker lines to the stronger ones ([O II]  $\lambda$ 3727 to [O II]  $\lambda$ 3729, [O III]  $\lambda$ 4960 to [O III]  $\lambda$ 5007, [N II]  $\lambda$ 6548 to [N II]  $\lambda$ 6583, [S II]  $\lambda$ 6731 to [S II]  $\lambda$ 6717 and [O II]  $\lambda$ 7330 to [O II]  $\lambda$ 7320).

During the fitting procedure, an emission feature close to 4360 Å has been detected and become blended with the [O III]  $\lambda$ 4363 auroral line, especially in the high-metallicity stacks. A similar contamination was previously found also by Andrews & Martini (2013) in their composite spectra. The nature of this feature is unknown, but it may reasonably be associated with emission lines from [Fe II]  $\lambda$ 4360. In fact, many others features from the same ion are clearly observable both in the same (e.g. [Fe II]  $\lambda$ 4288) and in different spectral windows; this particular emission has been reported also in studies on the Orion nebula (see e.g. table 2 of Esteban et al. 2004). Moreover, the strength of the line increases with increasing metallicity, as well as the other [Fe II] lines in the spectra. In Fig. 3 we



**Figure 2.** Fit and subtracted spectra for wavelength ranges relative to H $\beta$  and [O III] nebular lines (upper left-hand panels), H $\alpha$ , [N II] and [S II] nebular lines (upper right-hand panels), [O III]  $\lambda$ 4363 auroral line (lower left-hand panels) and [O II]  $\lambda$ 7320, 7330 auroral lines (lower right-hand panels), respectively, for the 0.5; 0.5 stack. For strong nebular lines, the upper panel shows the stacked spectrum (black) and the stellar continuum best-fitting component (red), while the bottom panel shows the residual spectrum after the stellar continuum subtraction. For auroral line boxes, a single galaxy spectrum is shown in the upper panel for comparison, while the stacked spectrum is shown in the middle one. The yellow shaded regions mark the spectral interval masked out during the stellar fitting procedure.



**Figure 3.** Left: composite spectra for the 0.4; 0.6 (upper panel), 0.5; 0.0 (middle panel) and 0.2; -0.6 (lower panel) stack, in the wavelength range relative to [O III]  $\lambda 4363$ , after the stellar continuum subtraction. The different components of the fit are reported in blue while the red curve represents the total fit. The metallicity of each stack is reported in the right-upper part of the corresponding panel. The contamination of the [O III]  $\lambda 4363$  line becomes more relevant with increasing metallicity (in the last case we can fit up to three components), as well as the intensity of the [Fe II] emission line at 4288 Å.

show three stacked spectra corresponding to different metallicities, namely 0.4; 0.6, 0.5; 0.0 and 0.2; -0.6, after performing the stellar continuum subtraction in the spectral window that contains the [O III]  $\lambda 4363$  line. The metallicity of each stack (see Section 5) is reported on every panel. The figure clearly shows how [O III]  $\lambda 4363$  becomes more contaminated as the metallicity increases, with the [Fe II] emission being just a few per cent of the flux of the oxygen one in the upper panel and then completely blending with it in the other two. The [Fe II] emission line at 4288 Å is also clearly visible in all the composite spectra, with increasing strength for increasing metallicity, as expected.

Therefore we simultaneously fit the  $\lambda 4360$  feature and [O III]  $\lambda 4363$ , linking both velocity widths and central wavelengths to  $H\gamma$ . The different components of the fit are shown in blue in Fig. 3, with the red line representing the total fit. Similarly to Andrews & Martini (2013), we consider the fit not sufficiently robust when the  $\lambda 4360$  emission flux resulted  $\geq 0.5$  times the flux measured for [O III]  $\lambda 4363$ . The use of contaminated [O III]  $\lambda 4363$  line would result in totally non-physical temperatures, which result overestimated by a factor of 10. According to this criterion, 42 out of 69 bins have been flagged for undetected [O III]  $\lambda 4363$ . We note that many previous detection of the [O III]  $\lambda 4363$  may be possibly

contaminated by this feature, resulting in unreliable measurements for this crucial auroral line; therefore, we recommend great care in using [O III]  $\lambda 4363$ , when detected, to measure electron temperature from high metallicity ( $12 + \log(O/H) \geq 8.3$ ) galaxy spectra. In the next section we discuss how to derive the electron temperature for the high-ionization zone for those stacks where the [O III]  $\lambda 4363$  was not considered sufficiently robust.

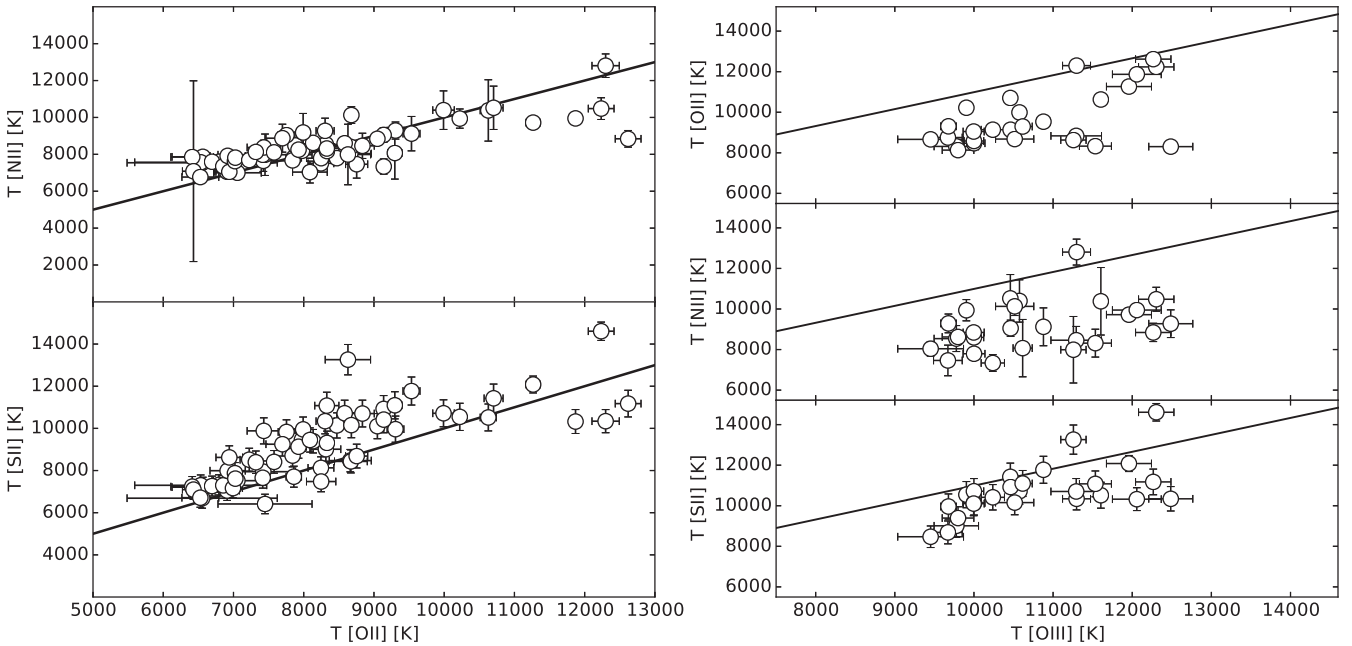
Despite the large number of galaxies in each bin and the great care in the fitting procedure, in some of our stacks we were unable to measure both [O II] and [O III] auroral line fluxes with sufficient precision. These are the stacked spectra corresponding to the 0.0; -0.8, 0.1; -0.4, 0.1; -0.7, 0.1; -0.8, 0.8; 0.2 and 0.6; -0.3 bins and we decided to exclude this stacks from all the forthcoming analysis.

### 3 ELECTRON TEMPERATURES AND IONIC ABUNDANCES DETERMINATION

#### 3.1 Electron temperatures

In principle, to measure electron temperatures and densities of different zones, the complete ionization structure of a H II region is needed. This is actually not possible and simpler approximations are always used. Usually a two-zone (of low and high ionization) or even a three-zone (of low, intermediate and high ionization) structure is adopted to model the H II regions responsible for emission lines in galaxies. In this work we consider a two-zone H II region: in this scenario, the high-ionization zone is traced by the  $O^{++}$  ion, while the low-ionization zone could be traced by different ionic species, e.g.  $O^+$ ,  $N^+$  and  $S^+$ . Thus, given the SDSS spectral coverage, in our case we have three different diagnostics for the temperature of the low-ionization zone (which we will refer to as  $t_2$  from now on), namely [O II]  $\lambda\lambda 3727, 3729$ /[O II]  $\lambda\lambda 7320, 7330$ , [N II]  $\lambda 6584$ /[N II]  $\lambda 5755$  and [S II]  $\lambda\lambda 6717, 6731$ /[S II]  $\lambda 4969$ , but only one for the temperature of the high-ionization zone ( $t_3$ ), namely [O III]  $\lambda 5007$ /[O III]  $\lambda 4363$ . Other collisionally excited lines probing the temperature of the intermediate- and high-ionization region are either too weak and thus undetectable even in galaxy stacks (e.g. [Ar III]  $\lambda 5192$ ) or fall outside the spectral range of the SDSS spectrograph (e.g. [S III]  $\lambda 9069$ , [Ne III]  $\lambda 3342$ ) and we could not use them.

We computed electron temperatures exploiting PyNEB (Luridiana, Morisset & Shaw 2012, 2015), the PYTHON-based version of the STSDAS NEBULAR routines in IRAF, using the new atomic data set presented in Palay et al. (2012). These routines, which are based on the solution of a five-level atomic structure following De Robertis, Dufour & Hunt (1987), determine the electron temperature of a given ionized state from the nebular to auroral flux ratio assuming a value for electron density. The electron density  $n_e$  can be measured from the density sensitive [S II]  $\lambda\lambda 6717, 6731$  or [O II]  $\lambda\lambda 3727, 3729$  doublets. In the majority of our stacks we fall in the low-density regime ( $n_e < 100 \text{ cm}^{-3}$ ), for we measure for example a [S II] ratio close to the theoretical limit of 1.41; in this case the dependence of our temperature diagnostics upon density is small. We note that using older atomic data sets (e.g. Aggarwal & Keenan 1999), instead on the new ones by Palay et al. (2012), would result in similar  $t_2$  but  $t_3$  systematically higher on average by 400 K. This is consistent with expectations given the updated effective collision strengths for [O III] lines, as pointed out e.g. in Nicholls et al. (2013) where a discrepancy of  $\sim 500$  K is expected at  $T[\text{O III}] \sim 10^4$  K (see e.g. section 7 and figs 2 and 12 of their paper for further details). We also note that the collision strengths presented in Palay et al. (2012) for the



**Figure 4.** Left-hand panels: electron temperatures derived from the [N II] and [S II] line ratios as a function of the electron temperature derived from [O II]; the equality line is shown in black. While the [N II] temperatures are consistent with the [O II] ones, the [S II] provides temperatures systematically higher. Right-hand panels: electron temperatures of the low ionization zone derived with all three different diagnostics as a function of the electron temperature of the high-ionization zone derived from [O III] line ratio. The black line represents the  $t_2$ – $t_3$  relation from equation (1), which does not provide a good representation of the data.

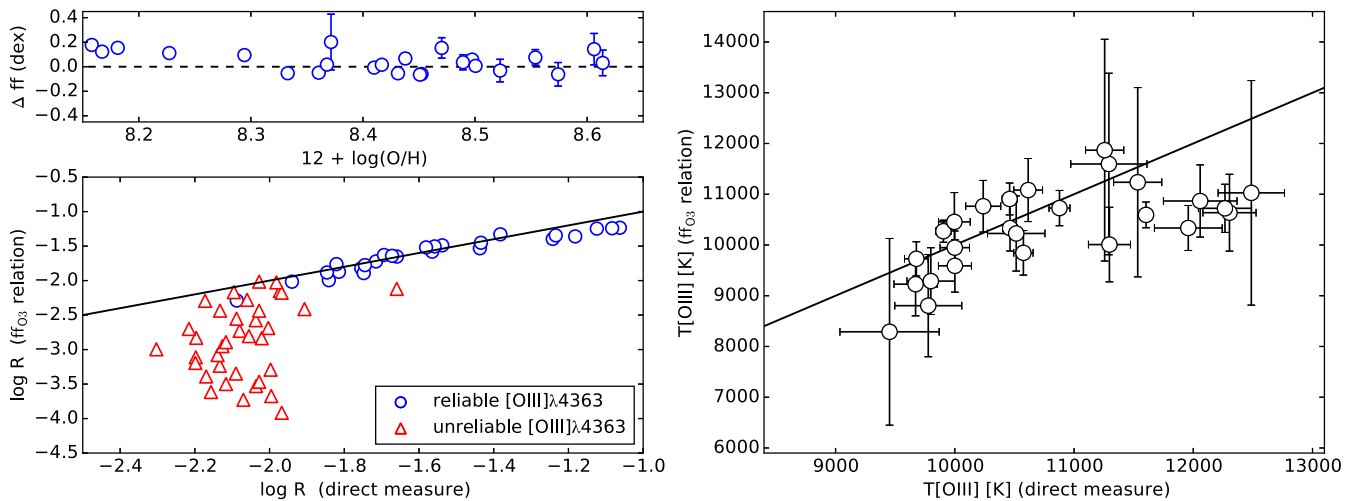
[O III] optical transitions are tabulated for a wide range of temperatures typical of nebular environments (from 100 to 30 000 K), which include all the temperatures we expect to find given the metallicity range spanned by the SDSS galaxies. Temperatures uncertainties were computed with Monte Carlo simulations. We generated 1000 realizations of the flux ratios, following a normal distribution with  $\sigma$  equal to the errors associated with the flux measurement by the fitting procedure and propagated analytically, and for each of them a temperature value was calculated. Then, we took the standard deviation of the resulting distribution as the error to associate to our temperature measure.

The left-hand panels of Fig. 4 show the relations between the temperatures of the low-ionization zone of our stacks inferred through different diagnostics, i.e.  $T_e$ [N II] (upper panel) and  $T_e$ [S II] (lower panel) as a function of  $T_e$ [O II]; the black line represents the line of equality. In the upper panel, we can see how the electron temperatures derived from nitrogen line ratios are consistent with  $T_e$ [O II], although with a large scatter, while in the lower panel we show that  $T_e$ [S II] is larger than  $T_e$ [O II] for almost all of our points, thus overpredicting  $t_2$  with respect to that derived through oxygen lines. Evidences of similar temperatures discrepancies have been reported by several works in the literature aimed at studying the physical properties of single H II regions (see e.g. Kennicutt, Bresolin & Garnett 2003; Bresolin et al. 2005; Esteban et al. 2009; Pilyugin et al. 2009; Binette et al. 2012; Berg et al. 2015). Interestingly, when  $T_e$ [S II] and  $T_e$ [O II] are considered in these papers, average offsets are usually found in the direction of larger  $T_e$ [O II], differently from what we found for our stacks. The most likely explanation resides in the different atomic data set for energy levels and collision strengths used among these works and ours. In fact, when computing  $T_e$ [S II] for our stacks exploiting different data sets, we find variations up to thousands of kelvins even at fixed diagnostic ratio.

Temperature fluctuations and inhomogeneities as well as shocks propagating within the photoionized gas have been proposed as the main sources of discrepancies between  $T_e$  inferred through different ionic tracers. Moreover, we are here considering the simple case of Maxwell–Boltzmann distributed electrons, while recent studies suggested that  $k$ -distributions could better represent the behaviour of free electrons inside single H II regions (Nicholls, Dopita & Sutherland 2012; Dopita et al. 2013; Nicholls et al. 2013). In particular, when considering only Maxwell–Boltzmann distributions, the electron temperature inferred using the most common diagnostics could be overestimated and this effect is more relevant for ions in which the excitation temperature of the upper level involved in the transitions results different from the kinetic temperature of the distribution (Nicholls et al. 2012): the effect of  $k$ -distributed electrons could therefore affect different temperature diagnostics in different ways and thus explain the observed discrepancies in  $T_e$  estimates. For an in-depth discussion on how  $k$ -distributed electrons could affect the main metallicity diagnostics in H II regions, see Dopita et al. (2013). However, it is not clear to what extent these processes can affect the determination of electron temperature when considering global galaxy spectra and, in particular, a stacking of many galaxies, as we do in this work.

### 3.2 The $t_2$ – $t_3$ relation

In the right-hand panels of Fig. 4 the relations between the temperatures of the low- and high-ionization zones are shown for those stacks in which we have been able to determine  $T_e$ [O III] directly from the spectra. Many works in literature report the existence of a relation between the temperatures of the different ionization zones. The linear form of this relation, called  $t_2$ – $t_3$  relation, has been proposed for the first time by Campbell, Terlevich & Melnick (1986) and then revised in several studies (Garnett 1992;



**Figure 5.** Left-hand panel:  $\log R$  (i.e.  $\log [\text{O III}] \lambda 4363/\text{H}\beta$ ) directly measured from the stacked spectra as a function of the same quantity obtained through the ff relation. Blue circles represent stacks whose  $[\text{O III}] \lambda 4363$  detection was considered robust according to the criteria described in the text, while red triangles are stacks whose  $[\text{O III}] \lambda 4363$  detection was considered unreliable. The black line represents the ff<sub>O3</sub> relation of equation (2). In the upper box the offset of  $\log(R)$  from the ff<sub>O3</sub> relation for the stacks with reliable  $[\text{O III}] \lambda 4363$  measurements is plotted as a function of the metallicity of the stacks. Right-hand panel:  $T_e[\text{O III}]$  derived from the ff<sub>O3</sub> relation as a function of direct measure  $T_e[\text{O III}]$  for the stacks with detected  $[\text{O III}] \lambda 4363$ ; black line represents equality.

Izotov et al. 2006; Pilyugin 2006; Pilyugin, Vílchez & Thuan 2006b; Pilyugin et al. 2009, 2010a). This relation is of great interest in the context of nebular studies since it is generally used to compute the electron temperature for unseen ionization states. In this work we will consider the linear relation suggested by Pilyugin et al. (2009) for temperatures derived through oxygen lines in their sample of H II regions, given by the equation

$$t_2 = 0.264 + 0.835 t_3, \quad (1)$$

where both temperatures are in units of  $10^4$  K. The  $t_2$ – $t_3$  relation of equation (1) is shown for comparison in each plot as a black line. We can see that the great majority of our stacks fall below the  $t_2$ – $t_3$  relation independently of the type of ion tracer, although with different median offsets from the relation corresponding to different tracers. In any case, it is clear that this  $t_2$ – $t_3$  relation underestimates the temperature of the high-ionization zone (or overestimates the temperature of the low-ionization zone) for our composite spectra. A very similar result was found by Andrews & Martini (2013) for their stacked spectra in bins of stellar mass, even though they used a different, but quite similar, form for the  $t_2$ – $t_3$  relation. The median offsets from the  $t_2$ – $t_3$  relation for our stacks are  $-2642$ ,  $-2712$  and  $-1055$  K for  $T_e[\text{O II}]$ ,  $T_e[\text{N II}]$  and  $T_e[\text{S II}]$ , respectively.

The offset between the electron temperatures of the stacks and the  $t_2$ – $t_3$  relation is in agreement with the trend found by Andrews & Martini (2013) for galaxy stacks but also by Pilyugin et al. (2010a) for individual galaxies, suggesting that this effect is not a product of the stacking procedure but rather reflects the intrinsic properties of global galaxy spectra. The most likely explanation indeed is that galaxy spectra are the result of several contributions from H II regions that could present very different physical properties, in terms of both chemical composition and hardness of their ionizing sources: this may affect the auroral line fluxes in the sense that they are weighted differently in H II regions of different temperatures. Since the auroral line flux does not scale linearly with metallicity, the effect of a luminosity-weighted average towards warmer H II regions on their total flux can be substantial and difficult to account for, in a way similar to temperature fluctuations for single H II regions described by Peimbert (1967); therefore, one can ob-

tain results that do not agree with the observed  $t_2$ – $t_3$  relation for single H II regions (Kobulnicky, Kennicutt & Pizagno 1999; Kennicutt, Bresolin & Garnett 2003). For example, Pilyugin et al. (2010a) showed that the  $t_2$ – $t_3$  relation offset can be substantially reproduced considering composite spectra obtained mixing contributions from few H II regions of very different temperatures. Moreover, the variation of the relative contribution of each H II region for different ionic species, together with the contribution from diffuse ionized gas (Moustakas & Kennicutt 2006), can explain the different distributions in the  $t_2$ – $t_3$  plane for different temperature diagnostics as well as the offset between different estimations of the temperature of the low-ionization zone. Despite these difficulties, the  $t_2$ – $t_3$  relation has been widely used in literature to compute electron temperatures of unseen ionization states.

### 3.3 The ff relations

Another possibility to solve the problem of determining the temperature of the high-ionization zone for stacks with unmeasured  $[\text{O III}] \lambda 4363$  is to rely on the relationship between the strong emission lines and the auroral line itself. As pointed out by Pilyugin (2005), a relation between auroral and nebular oxygen line fluxes has been demonstrated for H II regions of metallicity higher than  $12 + \log(\text{O}/\text{H}) \sim 8.25$ . This so-called ff relation allows to estimate the auroral line flux from the measured nebular line fluxes when the first is not available. In this work we employed the following ff relation (which we will refer to from now on as the ff<sub>O3</sub> relation) obtained by Pilyugin, Thuan & Vílchez (2006a) to infer the  $[\text{O III}] \lambda 4363$  flux for stacks where the flux of this line was not properly measured:

$$\log R = -4.151 - 3.118 \log P + 2.958 \log R_3 - 0.680 (\log P)^2, \quad (2)$$

where  $R = I_{[\text{O III}] \lambda 4363}/I_{\text{H}\beta}$ ,  $R_3 = I_{[\text{O III}] \lambda \lambda 4949, 5007}/I_{\text{H}\beta}$  and  $P$  (i.e. the excitation parameter)  $= R_3/(R_3 + R_2)$ , with  $R_2 = I_{[\text{O II}] \lambda \lambda 3727, 3729}/I_{\text{H}\beta}$ .

Inspection of the left-hand panel of Fig. 5, where  $\log(R)$  obtained through the ff<sub>O3</sub> relation is plotted against its direct measure from

the spectra, reveals that our stacks which satisfy the criteria for good [O III]  $\lambda$ 4363 detection (blue circle points) are in good agreement with equation (2), represented by the black line. We also plot as red triangles the points representing the composite spectra whose [O III]  $\lambda$ 4363 detection was flagged as unreliable due to [Fe II] contamination. Almost all of these points do not follow the  $\text{ff}_{\text{O3}}$  relation, falling well below the black line of Fig. 5. This was expected and corroborates the fact that the [O III]  $\lambda$ 4363 flux measurements in those stacks cannot be considered reliable. In the upper box of the same figure the deviations of  $\log(R)$  from the  $\text{ff}_{\text{O3}}$  relation (defined as  $\Delta\text{ff} = \log(R)_{\text{direct}} - \log(R)_{\text{ffrelation}}$ ) for the stacks with good [O III]  $\lambda$ 4363 detection are plotted as a function of metallicity derived with the  $T_e$  method; the points scatter around zero with a  $\sigma = 0.09$  dex, showing no trends with metallicity. The error bars in the upper box of Fig. 5 represent the uncertainties on  $\Delta\text{ff}$ , derived propagating the errors on the line flux measurements through equation (2); with the exception of a few points, this source of uncertainty (0.05 dex on average) cannot account for the total dispersion shown, being the larger part due to the intrinsic dispersion of the  $\text{ff}_{\text{O3}}$  relation itself.

In the right-hand panel of Fig. 5, we compare  $T_e[\text{O III}]$  derived through the  $\text{ff}_{\text{O3}}$  relation with the one directly measured from the spectra, for stacks with good detection of [O III]  $\lambda$ 4363; black line represents equality in this plot. Temperatures predicted by the  $\text{ff}_{\text{O3}}$  relation are in good agreement with direct measurements within the uncertainties, and there is no evident and systematic trend unlike what happens with the  $t_2$ – $t_3$  relation (see, for comparison, the upper right-hand panel of Fig. 4). From the above considerations and since all of our stacks, given the construction of our stacking grid, have a well-defined value for  $R_2$  and  $R_3$  (and thus  $P$ ), we decided to use equation (2) instead of the  $t_2$ – $t_3$  relation to determine the flux of [O III]  $\lambda$ 4363, and consequently the  $t_3$ , for stacks with no reliable detection of this auroral line. This allows to minimize the systematic offset introduced on abundance determination (see also Section 3.5).

### 3.4 Defining an ff relation for [O II] auroral lines

Following the same idea of Pilyugin (2005), we can exploit the direct measurements of [O II]  $\lambda\lambda$ 7320, 7330 in our stacks to define an analogous ff relation for the [O II] auroral doublet, which we will refer to as the  $\text{ff}_{\text{O2}}$  relation. Pilyugin et al. (2009) manage to obtain a similar relation for their sample of single H II regions in the low- $R_3$  range (i.e.  $\log R_3 < 0.5$ ). In particular, here we search for a combination of [O II]/ $H\beta$  and [O III]/ $H\beta$  (which define our stacking grid) that predicts the flux of the [O II]  $\lambda\lambda$ 7320, 7330 auroral doublet. Inspection of the upper panels of Fig. 6 reveals that since our stacks appear to lie on a surface in the 3D space defined by  $\log[\text{O II}] \lambda 3727/H\beta - \log[\text{O III}] \lambda 5007/H\beta - \log[\text{O II}] \lambda\lambda 7320, 7330/H\beta$ , we can search for the projection that minimizes the scatter in our sample and gives the combination of the first two indices that predict the value of the latter; such a combination could be easily formalized with a linear fit, which we show as a black line in the upper right-hand panel of the same figure. In order to better constrain the definition of our new  $\text{ff}_{\text{O2}}$  relation, we included the sample of low-metallicity SDSS DR6 galaxies from Pilyugin et al. (2010a) with detected [O II]  $\lambda\lambda$ 7320, 7330; these objects lie in the upper left-hand zone of our original diagram, which is characterized by high excitation galaxies. Even though these objects are characterized by a larger scatter than the stacks, they do not show any extreme offset from the surface defined by the stacks in the 3D space.

The functional form of our linear fit is the following:

$$\log R_{[\text{O II}]} = -1.913 + 0.806 \log R_2 + 0.374 \log R_3, \quad (3)$$

where  $R_{[\text{O II}]} = I_{[\text{O II}] \lambda\lambda 7320, 7330} / I_{H\beta}$ ; the results of the fit are shown in the upper right-hand panel of Fig. 6 as the black line. The dispersion around the  $\text{ff}_{\text{O2}}$  relation is 0.04 dex for the stacks and 0.06 dex for individual galaxies.

We can test the consistency of our new  $\text{ff}_{\text{O2}}$  relation by comparing the electron temperatures predicted and those directly measured from the spectra. The bottom right-hand panel of Fig. 6 shows that our  $\text{ff}_{\text{O2}}$  relation predicts  $T_e[\text{O II}]$  with good precision for both the stacks (red circles) and the single galaxies (blue points), even though single galaxies show a larger scatter from the equality line (in black) as the result of their intrinsic dispersion in the plane which define the  $\text{ff}_{\text{O2}}$  relation, with a few points whose temperature predictions deviate more than 1000 K from those observed. In addition, we can also compare the temperature prediction of the  $\text{ff}_{\text{O2}}$  relation with that from the  $t_2$ – $t_3$  relation of equation (1) (applied only to stacks with direct measurement of  $T_e[\text{O III}]$ ), for which the comparison with the direct  $T_e[\text{O II}]$  is shown in the bottom left-hand panel of Fig. 6. Our new  $\text{ff}_{\text{O2}}$  relation clearly reproduces the observed  $T_e[\text{O II}]$  better than the  $t_2$ – $t_3$  relation for both our stacks and the Pilyugin et al. (2010a) galaxies, as expected given the considerations made in the previous section about how the  $t_2$ – $t_3$  relation underestimates the temperature of the low-ionization zone when measured from global galaxy spectra. For these reasons, in this work we decided to use the new  $\text{ff}_{\text{O2}}$  relation defined by equation (3), instead of the  $t_2$ – $t_3$  relation, to infer the temperature of the low-ionization zone in single, low-metallicity galaxies where a direct measurement of  $T_e[\text{O II}]$  was not available (see Section 5).

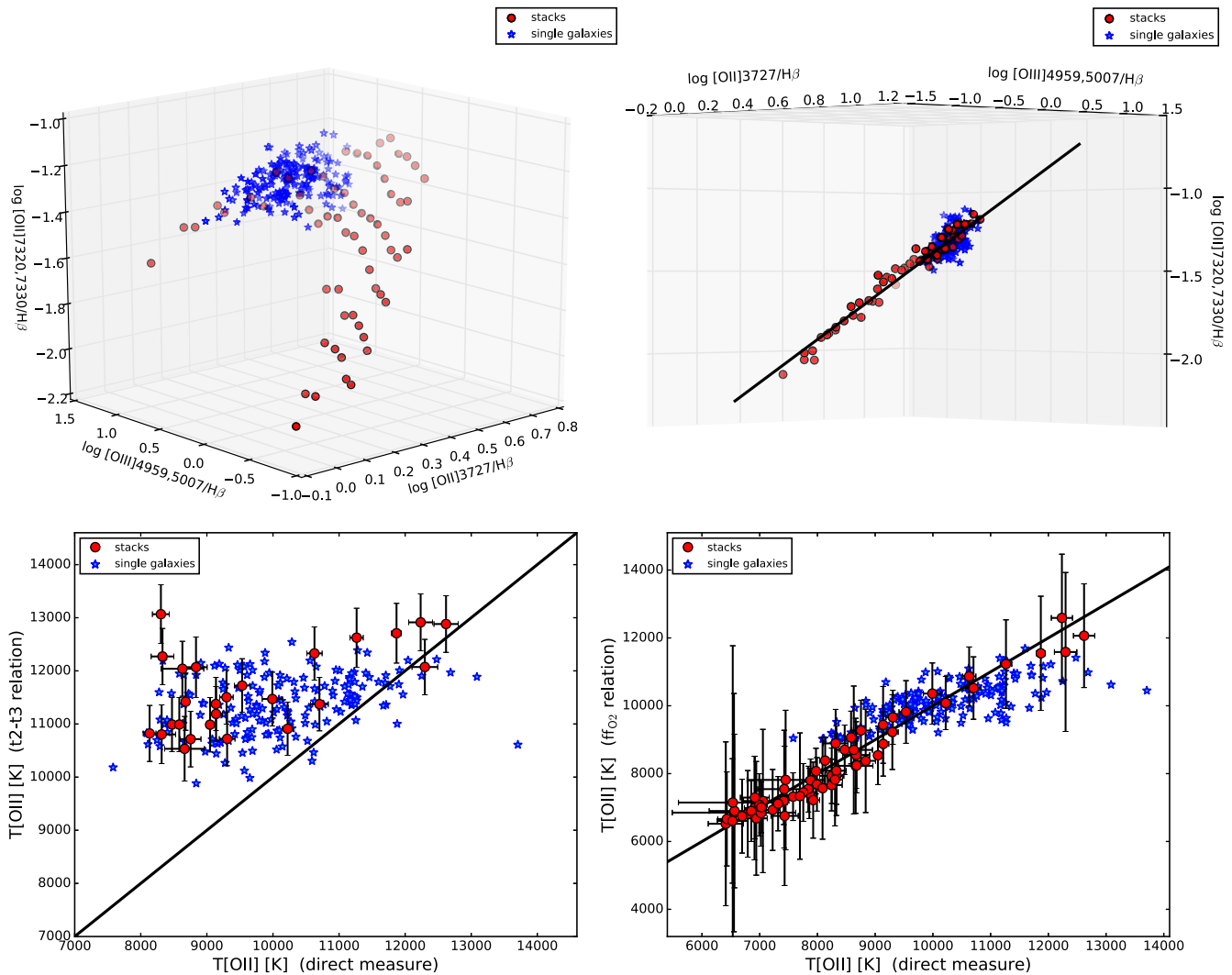
Summarizing, temperatures are derived as follows: when both [O III]  $\lambda$ 4363 and [O II]  $\lambda$ 7320 are detected,  $t_2$  and  $t_3$  are computed directly from the diagnostic ratios involving these auroral lines; when one of the two lines is missing, we use the relative ff relation to infer the flux of that line, compute the diagnostic ratio and derive  $T_e$ . We do not rely in this work on any relation, either empirically derived or based on photoionization models calculations, which links the temperatures of the different ionization zones.

### 3.5 Ionic abundances

We have calculated the ionic abundances of  $\text{O}^+$  and  $\text{O}^{++}$  for our stacks with the PYNEB version of the IRAF NEBULAR.IONIC routine, which determines the abundance of a ionic species given the electron temperature, electron density and the flux ratio of the relative strong emission line with respect to  $H\beta$ . We then assume that the total oxygen abundance is the sum of the two species considered,

$$\frac{\text{O}}{\text{H}} = \frac{\text{O}^+}{\text{H}^+} + \frac{\text{O}^{++}}{\text{H}^+}, \quad (4)$$

neglecting the contribution from  $\text{O}^{3+}$ , that can be found in highly ionized gas but it is typically minimal (Andrews & Martini 2013). In calculating the  $\text{O}^+$  abundance, we used the electron temperature derived from the [O II] diagnostic ratios, while to derive the  $\text{O}^{++}$  abundance we used the electron temperature derived from the [O III] diagnostic ratios. For stacks with undetected [O III]  $\lambda$ 4363, we used the  $\text{ff}_{\text{O3}}$  relation of equation (2) to infer  $T_e[\text{O III}]$  and compute the  $\text{O}^{++}$  abundance of unseen ionization states. The systematic offset introduced in abundance determination is small, as we can see by comparing the total oxygen abundance inferred both from  $T_e[\text{O III}]$  and from the  $\text{ff}_{\text{O3}}$  relation in stacks with [O III]  $\lambda$ 4363 detection. The mean offset in metallicity is 0.028 dex, smaller than the typical abundance uncertainty. For comparison, the  $t_2$ – $t_3$  relation introduces an average metallicity overestimation of 0.19 dex, as a direct consequence of the underestimation of the  $T_e[\text{O III}]$ .



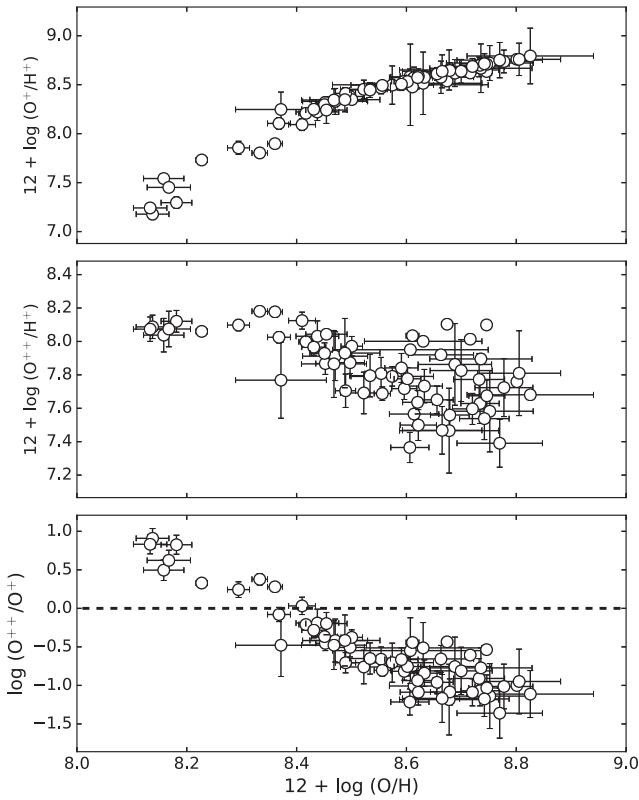
**Figure 6.** Upper panels:  $\log [\text{O II}] \lambda\lambda 7320, 7330/\text{H}\beta$  as a function of  $\log [\text{O II}] \lambda 3727/\text{H}\beta$  and  $\log [\text{O III}] \lambda 5007/\text{H}\beta$  for the sample of our stacks (red circles) and the Pilyugin et al. (2010a) galaxies (blue stars). All the points lie on a tight surface and in the right-hand panel we show the 2D projection that minimizes the scatter and predicts the flux of the oxygen auroral doublet from a combination of the two strong-line ratios; the black line represents the linear fit which defines our new ff relation. Bottom panels:  $T_e[\text{O II}]$  inferred through the  $t_2-t_3$  relation (left-hand panel) and through the  $\text{ff}_{\text{O}_2}$  relation (right-hand panel) as a function of the direct measure  $T_e[\text{O II}]$ . Symbols are the same as in the upper panels. The equality line is shown in black in both panels.

However, since the relative contribution of the  $\text{O}^+$  state is dominant in almost all the stacks, especially in the high-metallicity region, the inferred  $\text{O}^{++}$  represents only a small contribution to the total oxygen abundance. This is shown in Fig. 7, where the single ionic abundances for the  $\text{O}^{++}$  and  $\text{O}^+$  species and their ratio are plotted as a function of the total oxygen abundance: a large part of our stacks above  $12 + \log(\text{O}/\text{H}) = 8.5$  presents an  $\text{O}^{++}$  contribution to total oxygen abundance that does not exceed the 10–20 per cent. The upper and middle panels of Fig. 7 show how the  $\text{O}^+$  abundance is seen increasing monotonically in our stacks with the total metallicity while the  $\text{O}^{++}$  seems to remain constant or slightly decrease, being in any case affected by a large scatter. A very similar trend was found also by Andrews & Martini (2013) (see fig. 5 of their paper). We note that in almost all the stacks with significant contribution of the  $\text{O}^{++}$  abundance (i.e.  $\geq 50$  per cent) we were able to measure  $T_e[\text{O III}]$  directly. Uncertainties on ionic abundances were evaluated following the same Monte Carlo simulations used to compute errors on electron temperatures.

#### 4 TESTS ON THE METHOD

In this work we stacked spectra of several hundreds of galaxies per bin in order to enhance the S/N and detect the auroral lines needed for the application of the  $T_e$  method. Of course, physical properties like electron temperatures and metallicities inferred from stacked spectra are meaningful only if they are a good representation of the average properties of objects that went into the stack. In particular, the risk is that few objects could dominate the contribution on auroral line fluxes, thus biasing the estimate of electron temperature, and consequently of metallicity, from stacked spectra. This work is also based on the assumption that galaxies with similar values for both  $[\text{O II}] \lambda 3727/\text{H}\beta$  and  $[\text{O III}] \lambda 5007/\text{H}\beta$  have similar metallicities.

In order to test our hypothesis and the stacking procedure, we took the sample of galaxies from Pilyugin et al. (2010a) with detected  $[\text{O III}] \lambda 4363$  and  $[\text{O II}] \lambda\lambda 7320, 7330$ , and stacked these spectra in bins of  $[\text{O II}]/\text{H}\beta$  and  $[\text{O III}]/\text{H}\beta$  according to our pipeline. For this analysis we considered only those bins with at least 15 objects, which namely are 0.4; 0.6, 0.4; 0.7, 0.5; 0.5 and 0.5; 0.6, as reported



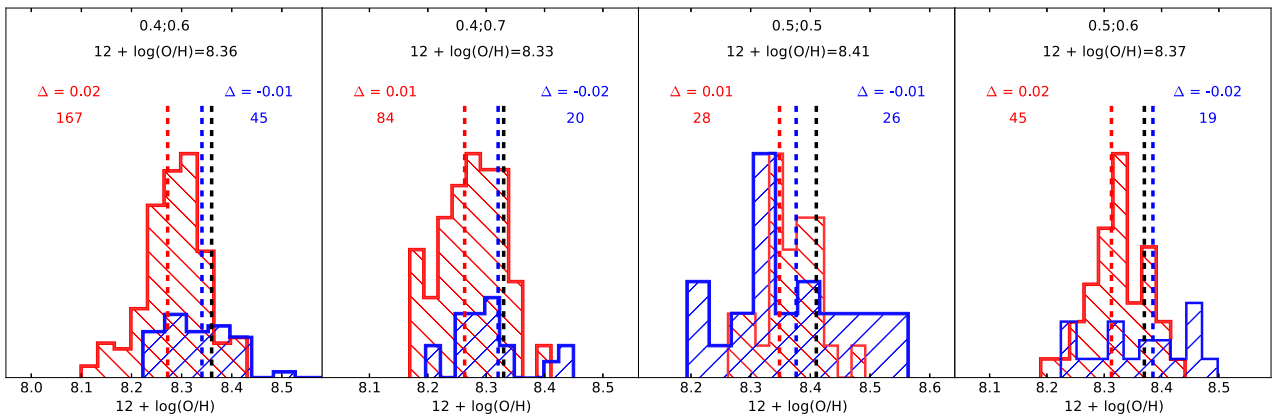
**Figure 7.** The  $O^+$  abundance (upper panel), the  $O^{++}$  abundance (middle panel) and the relative ionic abundance of the two species (bottom panel) are shown as a function of the total oxygen abundance.

at the top of Fig. 8. Furthermore, we searched for those galaxies in our original sample (described in Section 2), with detection of  $[O\text{ III}]\lambda 4363$  in the MPA/JHU catalogue, falling into the same bins. In particular, we selected only galaxies with  $[O\text{ III}]\lambda 4363$  detected at  $\geq 10\sigma$ .

Temperatures and metallicities were computed for each single galaxy in both samples. The sample of Pilyugin et al. (2010a) galaxies has both the oxygen auroral lines detected, thus we were able to

directly infer the temperatures and abundances of both oxygen ionic species. Since only the  $[O\text{ III}]\lambda 4363$  auroral line is instead available for objects selected from the MPA/JHU catalogue, we used the  $ff_{O_2}$  relation of equation (3) to determine the flux of the  $[O\text{ II}]\lambda\lambda 7320, 7330$  auroral doublet and compute  $T_e[O\text{ II}]$  and the  $O^+$  abundance. Then, we generated, for each bin, separate composite spectra for both samples and we measured  $T_e$  and metallicities with the  $T_e$  method from the stacked spectra.

Fig. 8 shows the histograms of metallicity distribution of individual galaxies in each bin for the Pilyugin et al. (2010a) sample (from now on: the blue sample) and the sample selected from the MPA/JHU catalogue (from now on: the red sample). We note the quite small range of metallicities spanned by single galaxies in each bin, with typical dispersions of 0.1 dex, consistently with the width of our binning grid. Even though we cannot perform the same test for higher metallicity stacks due to the lack of auroral line detection in single galaxies, this corroborates the assumption that galaxies belonging to a given bin of fixed  $[O\text{ III}]\lambda 5007/H\beta$  and  $[O\text{ II}]\lambda 3727/H\beta$  have similar metallicities and that we are thus stacking objects with similar properties in terms of oxygen abundance. The dashed lines in Fig. 8 indicate instead the metallicity inferred from the associated stacked spectrum for both samples. The difference between the average metallicity of single galaxies in a given bin and the one inferred from the stacked spectrum is reported as  $\Delta$ ; the number of objects per bin is also written. We note that abundances estimated from stacks are well matched to the average of the metallicity distributions in every bin, with offsets being at most 0.02 dex for both samples. However, both the red and blue samples could not be fully representative of the galaxy population inside each bin, which consist also of a large number of galaxies with no detection of auroral lines. Therefore, we compare the metallicity inferred from the stacked spectra of both subsamples with the one derived from the global composite spectrum, i.e. the spectrum obtained stacking all the galaxies included in that bin according to the procedure described in Section 2. These values are reported at the top of each box of Fig. 8 and indicated by the black dashed lines. We find good agreement between the global stack metallicity and the one inferred from the stacked spectra of the two different subsamples, with typical offsets on average of 0.04 dex, even though we note a systematic metallicity underestimation when considering the two subsamples



**Figure 8.** Histograms of  $T_e$  method metallicities for the subsample of galaxies selected from Pilyugin et al. (2010a) with detected  $[O\text{ III}]\lambda 4363$  and  $[O\text{ II}]\lambda\lambda 7320, 7330$  auroral lines (blue sample) and for galaxies with  $[O\text{ III}]\lambda 4363$  detected at  $>10\sigma$  from the MPA/JHU catalogue (red sample) for the 0.4; 0.6, 0.4; 0.7, 0.5; 0.5 and 0.5; 0.6 bin. The dashed lines indicate the metallicity inferred from the composite spectra obtained stacking the relative sample of galaxies. In every panel is also reported, for both subsamples, the difference between the average metallicity of the distribution and the value inferred from the associated stacked spectrum ( $\Delta$ ) and the number of objects per stack. The metallicity of the global stack, i.e. the stack obtained from the full sample of galaxies that fall in that bin, is written in the upper part of each panel and indicated by the dashed black line.

with respect to the global one. This is probably due to the fact that, when creating the stacked spectra for the different subsamples, we are averaging upon the most metal-poor galaxies in the bin, which in fact have the auroral lines detected. This could bias the subsample stacks towards lower metallicities, but this effect is smaller both than our bin size and than the average uncertainty associated with abundance measurements in our stacks. We therefore conclude that different subsampling criteria inside the same bin do not dramatically affect the metallicity estimation from composite spectra and therefore that stacked spectra are effectively representative of the average properties, in terms of oxygen abundance, of the objects from which they are generated.

## 5 CALIBRATIONS OF STRONG-LINE METALLICITY INDICATORS

In order to extend the metallicity range covered by our calibrations, we add to our stacks a sample of single galaxies with robust detection of [O III]  $\lambda$ 4363. We selected galaxies from our original SDSS DR7 sample with [O III]  $\lambda$ 4363 detection at  $>10\sigma$ , and we recomputed the oxygen abundance for these galaxies according to the procedure described in the previous section. In particular, we derive  $T_e[\text{O III}]$  directly exploiting the [O III]  $\lambda$ 4363 value reported on the MPA/JHU catalogue and used the  $\text{ff}_{\text{O}_2}$  relation of equation (3) to infer  $T_e[\text{O II}]$  and the  $\text{O}^+$  ionic abundance. Even though a part of these galaxies, although not all of them, is already included into our stacking grid, we are able in this way to directly account for some of the most metal poor galaxies of our sample, without averaging them into the stacking bins; thus, we can better constrain the low-metallicity region of our calibrations.

In Fig. 9 we plot the relations between some of the most widely used strong-line metallicity indicators and gas-phase oxygen abundance for our full sample. In particular, we recalibrate  $R_2([\text{O II}] \lambda 3727/\text{H}\beta)$ ,  $R_3([\text{O III}] \lambda 5007/\text{H}\beta)$ ,  $R_{23}([\text{O II}] \lambda 3727 + [\text{O III}] \lambda \lambda 4959, 5007)/\text{H}\beta)$ ,  $\text{O}_{32}([\text{O III}] \lambda 5007/[\text{O II}] \lambda 3727)$ ,  $\text{N}_2([\text{N II}] \lambda 6584/\text{H}\alpha)$  and  $\text{O}_3\text{N}_2([\text{O III}] \lambda 5007/\text{H}\beta)/([\text{N II}] \lambda 6584/\text{H}\alpha)$ . Green small stars represent single galaxies, while circles represent our stacked spectra, colour coded by the number of objects that went into each stack. To derive our new calibrations, we performed a polynomial fitting whose general functional form is

$$\log R = \sum_N c_n x^n, \quad (5)$$

where  $R$  is a given diagnostic and  $x$  is the oxygen abundance normalized to the solar value ( $12 + \log(\text{O}/\text{H})_{\odot} = 8.69$ ; Allende Prieto, Lambert & Asplund 2001). Since the indicators based on the ratio between oxygen forbidden lines over hydrogen recombination lines exhibit the well-known double branch behaviour, a high-order polynomial fitting is required. Assuming that the uncertainty on the auroral line flux, which represents the main contribution to the error in the  $T_e$  abundance determination in our stacks, decreases as the square root of the number of galaxies, in our fitting procedure we assigned a weight equal to this value to each point representing a stack; points associated with single galaxies have been weighted as they were stacks of only one object. In this way we also avoid our fit to be dominated by the low-metallicity single galaxies which are far more numerous than the stacks. In Fig. 9 our new calibrations are shown with the blue curve and in Table 2 the best-fitting coefficients and the rms of the residuals of the fit are reported for each of them.

We then applied each calibration to our total sample of single galaxies and stacks and computed the differences between  $T_e$

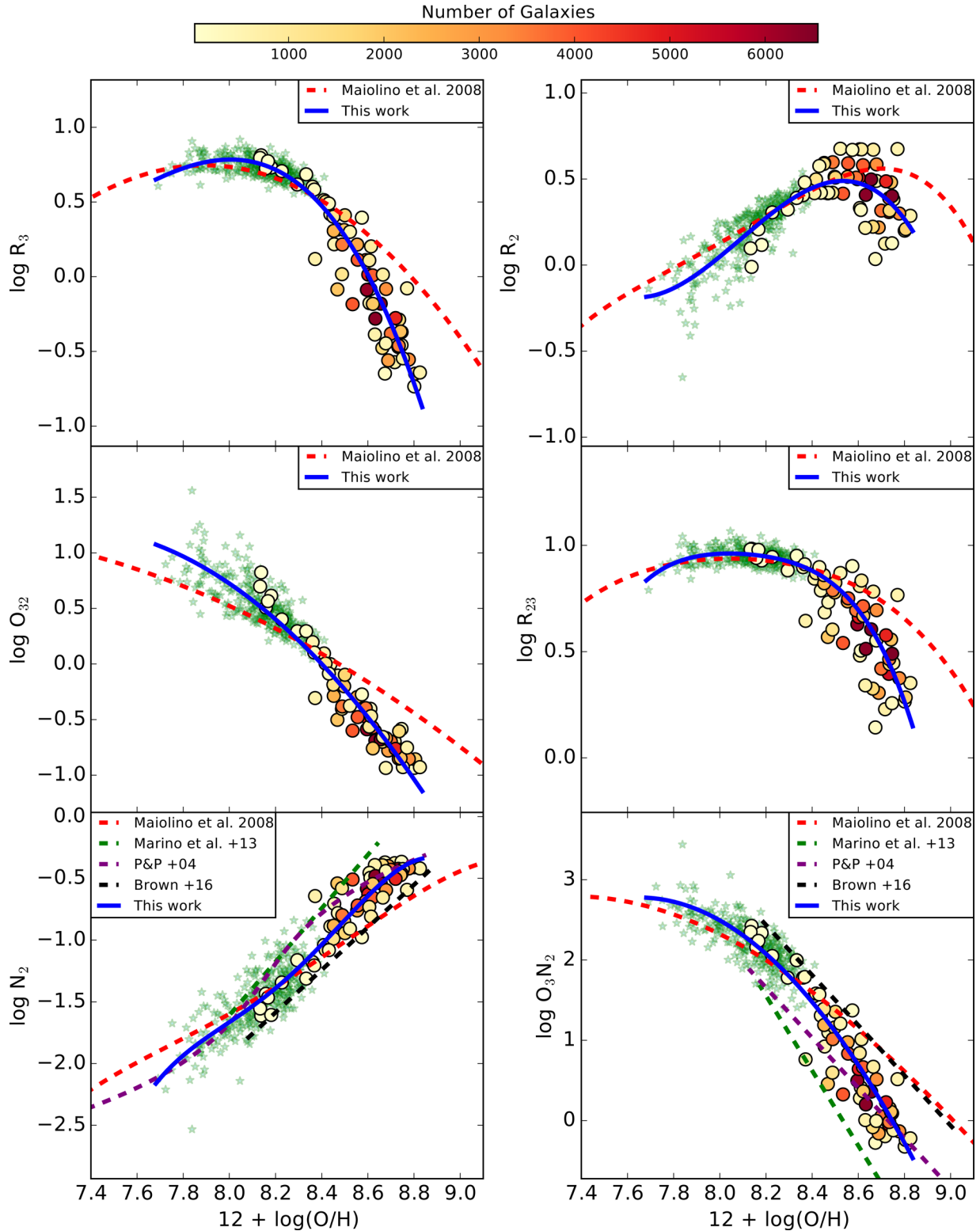
method metallicity and metallicity predicted by the calibration, in order to give an estimate of the dispersion along the  $\log(\text{O}/\text{H})$  direction, which is reported as  $\sigma$  in Table 2. For double-branched diagnostics (i.e.  $R_3$ ,  $R_2$  and  $R_{23}$ ) this estimate is provided only considering the metallicity range where they show monotonic dependence on  $\log(\text{O}/\text{H})$ , which is reported in the *Range* column of Table 2. This column represents indeed the range of applicability for a given diagnostics when used as a single metallicity indicator. We note that  $\sigma$  should not be directly interpreted as the uncertainty to associate with metallicity determination with our calibrations, since uncertainties in emission line ratios could introduce comparable errors.

Since our calibrations are built from a non-homogeneous combination of single galaxies and stacks, dispersion in our diagrams is due to different contributions. In the range covered by single SDSS galaxies, it is the consequence of the intrinsic spread in a given strong-line ratio at fixed metallicity and of the uncertainty on the auroral line fluxes measurement. For the high-metallicity region covered by our stacks, since we are averaging on a large number of objects, the scatter due to the intrinsic dispersion should be in principle reduced. However, we must consider the effects associated with the particular choice of our stacking grid. Every stack has, by definition, a defined value of  $[\text{O II}]/\text{H}\beta$  and  $[\text{O III}]/\text{H}\beta$ ; therefore, in the  $R_2$  and  $R_3$  calibration diagrams the residual dispersion reflects the segregation in a given diagnostic when the other is fixed. This means that for any given value of one line ratio, different metallicities can be found varying the other one. This is particularly clear in the  $R_2$  calibration, where different sequences for different  $[\text{O III}]/\text{H}\beta$  values appear at metallicities above 8.2. Therefore, this diagnostic shows a clear dependence on oxygen abundance only in the low-metallicity regime, revealing how most of SDSS galaxies are falling in the transition zone between the two branches of this indicator. Thus, for the majority of our stacks the metallicity dependence is driven by  $[\text{O III}]/\text{H}\beta$ , and indeed for this diagnostic the segregation in sequences of  $[\text{O II}]/\text{H}\beta$  is much less prominent.

For other indicators, the dispersion mainly reflects the scatter for a given diagnostic line ratio inside each  $[\text{O II}]/\text{H}\beta$ – $[\text{O III}]/\text{H}\beta$  bin. For each diagnostic the distribution of the corresponding line ratio inside our bins is generally strongly peaked, even though we are affected by different dispersions when considering different positions on our stacking grid. This means that a given line ratio, as measured from the stacked spectra, can be respectively more or less representative of the distribution of galaxies inside a given bin for different positions on the diagram. However, for every diagnostic ratio here considered, the typical dispersion of its distribution inside a given bin is of the order of 0.1 dex (or less), thus being consistent with the choice of our bin size.

In Fig. 9 we compare our new calibrations with those from Maiolino et al. (2008). They obtained semi-empirical calibrations combining direct abundance determination for galaxies from the Nagao et al. (2006) sample with metallicity estimation from theoretical models by Kewley & Dopita (2002). The two calibrations agree well, as expected, for most of the indicators at low metallicities, the main discrepancies arising in the high-metallicity regime where  $T_e$  method metallicities of our stacks result lower than those predicted by photoionization models. This introduces a clear deviation in the slope in all our calibrations, that change significantly their steepness after  $12 + \log(\text{O}/\text{H}) \sim 8.2$ . In fact, we note that the highest metallicities inferred from our composite spectra are only slightly higher ( $\sim 0.1$  dex) than the solar value.

For the  $\text{O}_3\text{N}_2$  and  $\text{N}_2$  indicators we can compare our calibrations also with empirical ones from Pettini & Pagel (2004) and Marino



**Figure 9.** Strong line diagnostics as a function of oxygen abundance for our full sample: small green stars represent the sample of single SDSS galaxies with  $[\text{O III}] \lambda 4363$  detected at  $S/N > 10$ , while circles are the stacks colour coded by the number of galaxies in each bin. Our best-fitting polynomial functions are shown as solid blue curves, while the dashed red line represents the Maiolino et al. (2008) semi-empirical calibrations. In the  $N_2$  and  $O_3N_2$  diagrams also the Pettini & Pagel (2004) (dashed purple curve), Marino et al. (2013) (dashed green curve) and Brown et al. (2016) for  $\Delta(\text{SSFR}) = 0$  (dashed black curve) calibrations are shown. A publicly available routine to apply these calibrations can be found at <http://www.arcetri.astro.it/metallicity/>.

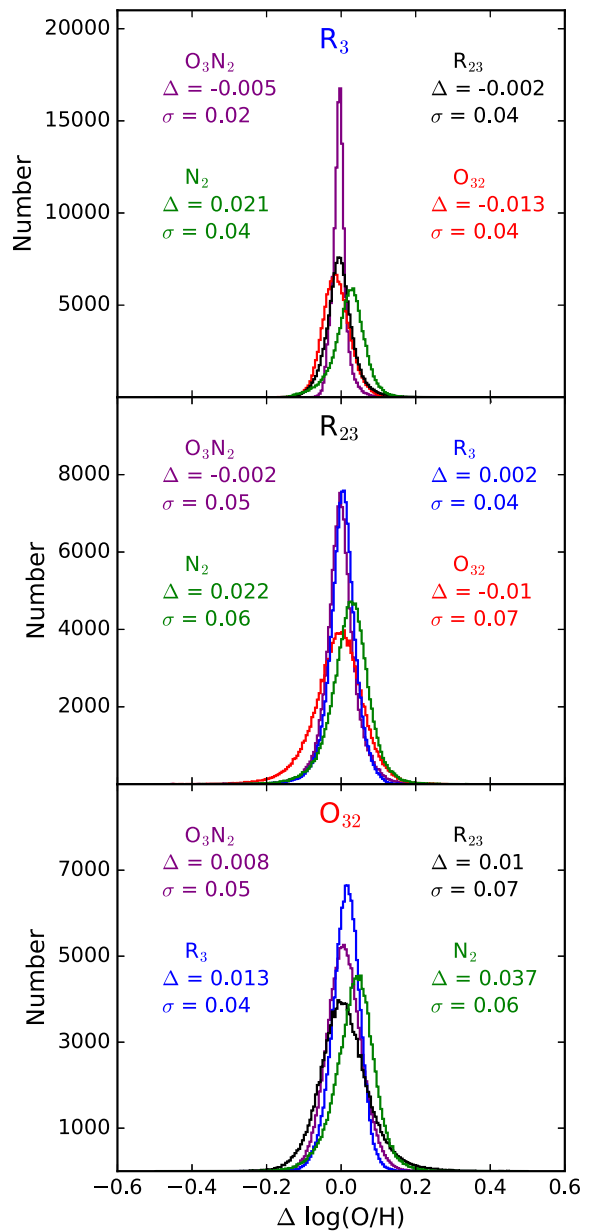
**Table 2.** Best-fitting coefficients and rms of the residuals for calibrations of metallicity diagnostics given by equation (5). The  $\sigma$  parameter is an estimate of the dispersion along the  $\log(\text{O}/\text{H})$  direction in the interval of applicability given in the *Range* column.

Diagnostic	$c_0$	$c_1$	$c_2$	$c_3$	$c_4$	rms	$\sigma$	Range
$R_2$	0.418	-0.961	-3.505	-1.949		0.11	0.26	$7.6 < 12 + \log(\text{O}/\text{H}) < 8.3$
$R_3$	-0.277	-3.549	-3.593	-0.981		0.09	0.07	$8.3 < 12 + \log(\text{O}/\text{H}) < 8.85$
$\text{O}_{32}$	-0.691	-2.944	-1.308			0.15	0.14	$7.6 < 12 + \log(\text{O}/\text{H}) < 8.85$
$R_{23}$	0.527	-1.569	-1.652	-0.421		0.06	0.12	$8.4 < 12 + \log(\text{O}/\text{H}) < 8.85$
$\text{N}_2$	-0.489	1.513	-2.554	-5.293	-2.867	0.16	0.10	$7.6 < 12 + \log(\text{O}/\text{H}) < 8.85$
$\text{O}_3\text{N}_2$	0.281	-4.765	-2.268			0.21	0.09	$7.6 < 12 + \log(\text{O}/\text{H}) < 8.85$

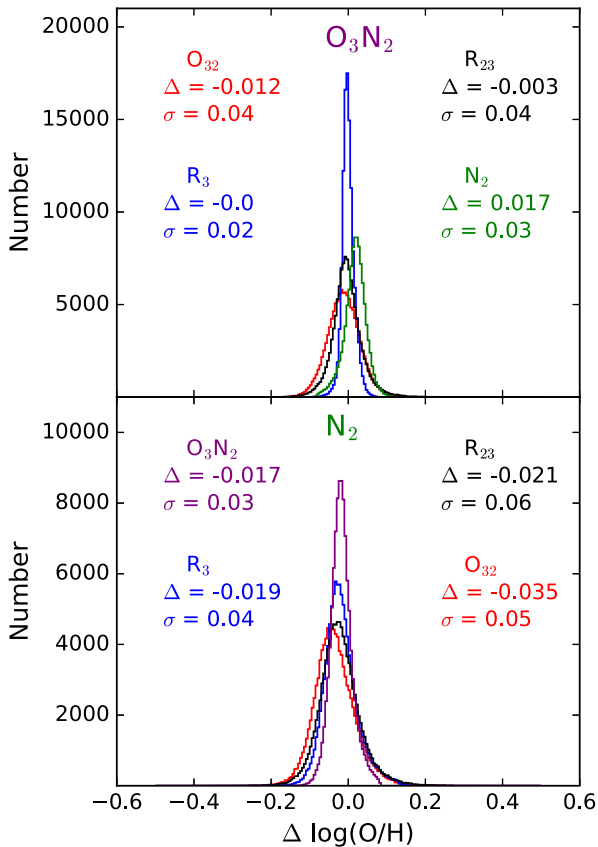
et al. (2013), who used single H II regions and not integrated galaxy spectra to calibrate these line ratios against metallicity. Our calibrations have comparable slopes to those of Marino et al. (2013), but they present a systematic offset towards higher metallicities. This is probably due to the fact that calibrations entirely based on H II regions like Marino et al. (2013) are biased towards high excitation conditions and low metallicities. Our  $\text{N}_2$  calibration is in good agreement with Pettini & Pagel (2004) at low metallicities but diverge, in the direction of predicting higher abundances, in the middle region. At metallicities close to the solar value this diagnostic begin to saturate, as expected from the fact that nitrogen becomes the dominant coolant of the ISM: the two calibrations then become comparable again. The  $\text{O}_3\text{N}_2$  calibration instead presents a different slope than the Pettini & Pagel (2004) since the slope of their calibration is determined by the use of photoionization models at high metallicities due to the lack in their sample of H II regions with direct abundances in that region of the diagram. We note that our calibrations are better constrained to be used for integrated galaxy spectra, since single H II regions upon which most of the empirical calibrations are based on do not properly and fully cover the parameter space where many galaxies lie.

In Fig. 9 we also compare our calibrations for the  $\text{N}_2$  and  $\text{O}_3\text{N}_2$  indicators with those derived by Brown et al. (2016), who derived oxygen abundances with the  $T_e$  method from stacked spectra in bins of stellar mass and  $\Delta(\text{SSFR})$ , i.e. the deviation of the specific star formation rate from the star-forming main sequence (SFMS; Noeske et al. 2007). Since they include  $\Delta(\text{SSFR})$  as a second parameter in their calibrations, we decide to plot here (in black) only the curves representative of SFMS galaxies, i.e. those obtained assuming  $\Delta(\text{SSFR}) = 0$ . In fact, our galaxy sample is distributed around their SFMS representation (see equation 6 of Brown et al. 2016), with a small median offset of 0.009 dex. Their calibrations show an offset of  $\sim 0.1$  dex towards higher metallicities for both indicators with respect to ours. We note that our calibrations are more consistent with Brown et al. (2016) calibrations when considering their curves for  $\Delta(\text{SSFR}) = -0.75$ .

In order to test the self-consistency of our calibrations, we applied them individually to our original sample of SDSS galaxies. Diagnostics behaving monotonically (i.e.  $\text{N}_2$ ,  $\text{O}_3\text{N}_2$  and  $\text{O}_{32}$ ) can be compared over the full metallicity range spanned by our sample, to which we can straightly apply the calibration, while for those double valued we restricted our analysis to the interval given in the *Range* column of Table 2. Given that our  $R_2$  calibration mostly cover the transition zone for such indicator in our SDSS galaxy sample (i.e. for  $12 + \log(\text{O}/\text{H}) > 8.2$ ), we decide not to include it in this analysis. In each panel of Figs 10 and 11, we show the histograms of the differences in metallicity estimation between a given diagnostic and all the others. Every strong-line indicator is identified by a different colour, and in each panel the name of the



**Figure 10.** Upper panel: histograms of the difference between metallicities of the SDSS galaxies derived with the  $R_3$  calibration and through the other diagnostics. Each diagnostic is identified by a different colour: blue for  $R_3$ , red for  $\text{O}_{32}$ , black for  $R_{23}$ , purple for  $\text{O}_3\text{N}_2$  and green for  $\text{N}_2$ . The average offset and  $\sigma$  of the  $\Delta(\text{O}/\text{H})$  distributions are written for every diagnostic with the associated colour. Middle panel: same as upper panel, with  $R_{23}$  as reference diagnostic. Bottom panel: same as upper panel, with  $\text{O}_{32}$  as reference diagnostic.



**Figure 11.** Upper panel: same as upper panel of Fig. 10, with  $O_3N_2$  as reference diagnostic. Bottom panel: same as upper panel of Fig. 10, with  $N_2$  as reference diagnostic.

reference indicator is written in the upper region. The  $\Delta \log(O/H)$  is then evaluated as the difference between the metallicity probed by the reference indicator and the metallicities estimated with the other four; in each panel the average offset of the  $\Delta \log(O/H)$  distribution and the associated  $\sigma$  is reported.

Inspection of the different panels of Figs 10 and 11 reveals that metallicities probed by different indicators are in good agreement among each other, with average offsets below 0.04 dex and typical dispersions below 0.1 dex. The little systematic metallicity overestimate reported for the  $O_{32}$  indicator and underestimate for the  $N_2$  indicator with respect to the others can be accounted for as a product of the fitting procedure. In this sense, the use of higher order polynomials allows to straighten the consistency of all our calibrations, since it minimizes the mutual disagreement between metallicities determined with different indicators. Thus, Figs 10 and 11 show that consistent metallicities are obtained in individual galaxies when using different calibrations, allowing to compare for example abundances obtained from diagnostics located in different spectral regions.

From the above considerations, we can say that our calibrations represent a self-consistent set totally based on the  $T_e$  metallicity scale. This is important since at the present time there is not an established absolute abundance scale for galaxies. Most of the calibrations found in literature either rely on the assumptions of photoionization models or are based on individual  $H\ II$  regions metallicities. In the first case the metallicity scale defined by models is inconsistent with the  $T_e$  scale. In the second case, the emission line properties of  $H\ II$  regions differ from those of integrated galaxy

spectra and higher excitation conditions affect  $T_e$  abundance estimation towards lower values. Until the number of high-metallicity galaxy with detected auroral lines will increase, allowing to build fully  $T_e$  method calibrations based on samples of individual objects, our stacking technique represents a valuable approach to define  $T_e$ -based calibrations. However, it should be stressed that metallicity estimates obtained from these calibrations are always affected by the particular choice of the stacking procedure. For example, the Brown et al. (2016) calibrations from stacked spectra in bins of stellar mass and  $\Delta(SSFR)$  rely on a particular parametrization of the local star-forming main sequence and their metallicity predictions could differ from ours despite the fact that abundances are evaluated with the  $T_e$  method in both cases.

In this work we chose to recalibrate the strong-line diagnostics relying only on the values assumed by galaxies on particular emission line ratios, thus assuming only the validity of the strong-line methods to infer metallicity from spectra of star-forming galaxies. In this way our calibrations could be in principle applied to a great variety of cases, in particular to large integral field unit (IFU) galaxy surveys that allow spatially resolved abundance studies [e.g. CALIFA, Sánchez et al. 2012; Mapping Nearby Galaxies at APO (MaNGA), Bundy et al. 2015 or Sydney-AAO Multi-Object Integral-Field Spectrograph (SAMI), Croom et al. 2012]. In such cases for example, the use of our calibrations allows to relax the assumption that scaling relations well assessed on global scales (e.g. the  $M-Z$  and the  $M-Z-SFR$  relations) still hold on smaller, local ones.

## 6 SUMMARY

We provided new and totally empirical calibrations for some of the most widely used strong-line diagnostics for the determination of oxygen abundance in star-forming galaxies. These relations have been derived combining a sample of single low-metallicity galaxies together with the stacking of more than 110 000 galaxies of the SDSS in bins of 0.1 dex in the  $\log [O\ III] \lambda 3737/H\beta - \log [O\ III] \lambda 5007/H\beta$  diagram, just assuming that galaxies with such similar strong-line ratios also show similar metallicity (i.e. assuming the validity of the so-called strong-line method). The increase in S/N provided by the stacking procedure allowed us to detect and measure both the  $[O\ III] \lambda 4363$  and  $[O\ II] \lambda \lambda 7320, 7330$  auroral lines necessary to compute electron temperatures of the different ionization zones and apply the  $T_e$  method for measuring metallicity on the full range of metal abundances spanned by galaxies in the SDSS survey. Here are summarized our main results.

(i) We found evidence for  $[Fe\ II]$  contamination of the  $[O\ III] \lambda 4363$  auroral line in high-metallicity stacks (Fig. 3). This is one of the crucial lines for the application of the  $T_e$  method for abundances estimation, thus we recommend care in using this line as electron temperature diagnostic when detected in high-metallicity ( $12 + \log(O/H) \gtrsim 8.3$ ) galaxy spectra.

(ii) We analysed the relations between electron temperatures of different ionization zones, finding that our stacks do not follow the established  $t_2-t_3$  relation for  $H\ II$  regions (Fig. 4). They show instead better agreement with a relation that correlates auroral and nebular line fluxes ( $ff_{O_3}$  relation). Exploiting the direct detection of the  $[O\ II] \lambda \lambda 7320, 7330$  auroral doublet in all of our composite spectra, we provided a new relation (the  $ff_{O_2}$  relation) for the determination of the flux of this auroral line (Fig. 6).

(iii) We then analysed the relations between some of the most common strong-line diagnostics and oxygen abundance in order to

obtain a reliable calibration (Fig. 9). Our global sample allowed us to construct a set of calibrations, spanning more than 1 dex in metallicity, which are based on the uniform application of the  $T_e$  method for oxygen abundance estimation on global galaxy spectra. All our calibrations are therefore defined, over their whole range, on a consistent absolute  $T_e$  metallicity scale for local star-forming galaxies. The scatter around the best-fitting calibration varies between 0.05 and 0.15 for different indicators.

(iv) Comparing our new calibrations with different ones from literature reveals how our calibrations deviate significantly both from empirical ones based on H II regions and from theoretical ones based on photoionization models, especially at high metallicities. In fact, we find that our most metal-rich stacks have oxygen abundance significantly lower than those predicted by models, and at most 0.14 dex higher than the solar one (i.e.  $\sim 1.4 Z_\odot$ ). On the other hand, classical empirical calibrations obtained from H II regions samples generally show lower metallicities for fixed line ratios, probably due to the fact that the single H II regions used for those calibrations are somehow biased towards high excitation conditions in order to ensure auroral line detection at high metallicity.

(v) We applied our calibrations to the original sample of SDSS galaxies. Metallicity estimates from different calibrations result in good agreement between each other, with typical average offsets lower than 0.04 dex and dispersions of the order of 0.05 dex. We do not find any systematic effect of metallicity overestimate or underestimate between the different diagnostics. Thus, our calibrations represent a self-consistent set that could be used in a variety of different cases depending on the availability of emission lines.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the referee for his/her constructive comments that improved this work.

MC acknowledges financial support from INAF for his PhD fellowship.

RM acknowledges support from the ERC Advanced Grant 695671 "QUENCH" and support from the Science and Technology Facilities Council (STFC).

We greatly appreciate the MPA/JHU group for making their catalog public.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

## REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543  
 Aggarwal K. M., Keenan F. P., 1999, *ApJS*, 123, 311  
 Allende Prieto C., Lambert D. L., Asplund M., 2001, *ApJ*, 556, L63  
 Alloin D., Collin-Souffrin S., Joly M., Vigroux L., 1979, *A&A*, 78, 200  
 Andrews B. H., Martini P., 2013, *ApJ*, 765, 140  
 Berg D. A., Croxall K. V., Skillman E. D., Pogge R. W., Moustakas J., Groh-Johnson M., 2015, *ApJ*, 806, 16  
 Binette L., Matadamas R., Hägele G. F., Nicholls D. C., Magris C. G., Peña-Guerrero M. Á., Morisset C., Rodríguez-González A., 2012, *A&A*, 547, A29  
 Blanc G. A., Kewley L., Vogt F. P. A., Dopita M. A., 2015, *ApJ*, 798, 99  
 Bresolin F., Schaerer D., González Delgado R. M., Stasińska G., 2005, *A&A*, 441, 981  
 Bresolin F., Kudritzki R.-P., Urbaneja M. A., Gieren W., Ho I., Pietrzynski G., 2016, *ApJ*, 830, 64  
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151  
 Brown J. S., Martini P., Andrews B. H., 2016, *MNRAS*, 458, 1529  
 Bundy K. et al., 2015, *ApJ*, 798, 7  
 Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, *ApJ*, 429, 582  
 Campbell A., Terlevich R., Melnick J., 1986, *MNRAS*, 223, 811  
 Cappellari M., Emsellem E., 2004, *PASP*, 116, 138  
 Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245  
 Cenarro A. J., Cardiel N., Gorgas J., Peletier R. F., Vazdekis A., Prada F., 2001, *MNRAS*, 326, 959  
 Cresci G., Mannucci F., Sommariva V., Maiolino R., Marconi A., Brusa M., 2012, *MNRAS*, 421, 262  
 Croom S. M. et al., 2012, *MNRAS*, 421, 872  
 Davé R., Finlator K., Oppenheimer B. D., 2011, *MNRAS*, 416, 1354  
 Dayal P., Ferrara A., Dunlop J. S., 2013, *MNRAS*, 430, 2891  
 De Robertis M. M., Dufour R. J., Hunt R. W., 1987, *J. R. Astron. Soc. Canada*, 81, 195  
 Dopita M. A., Sutherland R. S., Nicholls D. C., Kewley L. J., Vogt F. P. A., 2013, *ApJS*, 208, 10  
 Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, *Ap&SS*, 361, 61  
 Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, *ApJ*, 644, 813  
 Esteban C., Peimbert M., García-Rojas J., Ruiz M. T., Peimbert A., Rodríguez M., 2004, *MNRAS*, 355, 229  
 Esteban C., Bresolin F., Peimbert M., García-Rojas J., Peimbert A., Mesa-Delgado A., 2009, *ApJ*, 700, 654  
 Esteban C., García-Rojas J., Carigi L., Peimbert M., Bresolin F., López-Sánchez A. R., Mesa-Delgado A., 2014, *MNRAS*, 443, 624  
 Falcón-Barroso J., Sánchez-Blázquez P., Vazdekis A., Ricciardelli E., Cardiel N., Cenarro A. J., Gorgas J., Peletier R. F., 2011, *A&A*, 532, A95  
 García-Rojas J., Esteban C., 2007, *ApJ*, 670, 457  
 Garnett D. R., 1992, *AJ*, 103, 1330  
 Izotov Y. I., Stasińska G., Meynet G., Guseva N. G., Thuan T. X., 2006, *A&A*, 448, 955  
 Kauffmann G. et al., 2003a, *MNRAS*, 341, 33  
 Kauffmann G. et al., 2003b, *MNRAS*, 346, 1055  
 Kennicutt R. C., Jr, Bresolin F., Garnett D. R., 2003, *ApJ*, 591, 801  
 Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35  
 Kewley L. J., Ellison S. L., 2008, *ApJ*, 681, 1183  
 Kobulnicky H. A., Kewley L. J., 2004, *ApJ*, 617, 240  
 Kobulnicky H. A., Kennicutt R. C., Jr, Pizagno J. L., 1999, *ApJ*, 514, 544  
 Le Borgne D., Rocca-Volmerange B., Prugniel P., Lançon A., Fioc M., Soubiran C., 2004, *A&A*, 425, 881  
 Lee H., Skillman E. D., Cannon J. M., Jackson D. C., Gehrz R. D., Polonski E. F., Woodward C. E., 2006, *ApJ*, 647, 970  
 Liang Y. C., Hammer F., Yin S. Y., Flores H., Rodrigues M., Yang Y. B., 2007, *A&A*, 473, 411  
 Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, *ApJ*, 772, 119

- López-Sánchez Á. R., Dopita M. A., Kewley L. J., Zahid H. J., Nicholls D. C., Scharwächter J., 2012, *MNRAS*, 426, 2630
- Luridiana V., Morisset C., Shaw R. A., 2012, in Manchado A., Stanghellini L., Schönberner D., eds, *Proc. IAU Symp. 283, Planetary Nebulae: An Eye to the Future*. Cambridge Univ. Press, Cambridge, p. 422
- Luridiana V., Morisset C., Shaw R. A., 2015, *A&A*, 573, A42
- McGaugh S. S., 1991, *ApJ*, 380, 140
- Maiolino R. et al., 2008, *A&A*, 488, 463
- Mannucci F. et al., 2009, *MNRAS*, 398, 1915
- Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, *MNRAS*, 408, 2115
- Marino R. A. et al., 2013, *A&A*, 559, A114
- Masters D., Faisst A., Capak P., 2016, *ApJ*, 828, 18
- Moustakas J., Kennicutt R. C., Jr, 2006, *ApJS*, 164, 81
- Moustakas J., Kennicutt R. C., Jr, Tremonti C. A., Dale D. A., Smith J.-D. T., Calzetti D., 2010, *ApJS*, 190, 233
- Nagao T., Maiolino R., Marconi A., 2006, *A&A*, 459, 85
- Nicholls D. C., Dopita M. A., Sutherland R. S., 2012, *ApJ*, 752, 148
- Nicholls D. C., Dopita M. A., Sutherland R. S., Kewley L. J., Palay E., 2013, *ApJS*, 207, 21
- Noeske K. G. et al., 2007, *ApJ*, 660, L47
- Pagel B. E. J., Edmunds M. G., Blackwell D. E., Chun M. S., Smith G., 1979, *MNRAS*, 189, 95
- Palay E., Nahar S. N., Pradhan A. K., Eissner W., 2012, *MNRAS*, 423, L35
- Peimbert M., 1967, *ApJ*, 150, 825
- Peimbert A., Peimbert M., Delgado-Inglada G., García-Rojas J., Peña M., 2014, *Rev. Mex. Astron. Astrofis.*, 50, 329
- Peimbert M., Peimbert A., Esteban C., García-Rojas J., Bresolin F., Carigi L., Ruiz M. T., López-Sánchez A. R., 2007, *Rev. Mex. Astron. Astrofis. Conf. Ser.*, 29, 72
- Pettini M., Pagel B. E. J., 2004, *MNRAS*, 348, L59
- Pilyugin L. S., 2005, *A&A*, 436, L1
- Pilyugin L. S., 2006, *MNRAS*, 375, 685
- Pilyugin L. S., Grebel E. K., 2016, *MNRAS*, 457, 3678
- Pilyugin L. S., Thuan T. X., 2005, *ApJ*, 631, 231
- Pilyugin L. S., Thuan T. X., Vílchez J. M., 2006a, *MNRAS*, 367, 1139
- Pilyugin L. S., Vílchez J. M., Thuan T. X., 2006b, *MNRAS*, 370, 1928
- Pilyugin L. S., Mattsson L., Vílchez J. M., Cedrés B., 2009, *MNRAS*, 398, 485
- Pilyugin L. S., Vílchez J. M., Cedrés B., Thuan T. X., 2010a, *MNRAS*, 403, 896
- Pilyugin L. S., Vílchez J. M., Thuan T. X., 2010b, *ApJ*, 720, 1738
- Pilyugin L. S., Grebel E. K., Mattsson L., 2012, *MNRAS*, 424, 2316
- Ricciardelli E., Vazdekis A., Cenarro A. J., Falcón-Barroso J., 2012, *MNRAS*, 424, 172
- Salim S. et al., 2007, *ApJS*, 173, 267
- Sánchez S. F. et al., 2012, *A&A*, 538, A8
- Sánchez-Blázquez P. et al., 2006, *MNRAS*, 371, 703
- Stasińska G., 2002, in Esteban C., García López R. J., Herrero A., Sánchez F., eds, *Cosmochemistry. The melting pot of the elements*. Cambridge Univ. Press, Cambridge, p. 115
- Stasińska G., 2005, *A&A*, 434, 507
- Tremonti C. A. et al., 2004, *ApJ*, 613, 898
- Troncoso P. et al., 2014, *A&A*, 563, A58
- Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, *MNRAS*, 404, 1639
- Vazdekis A., Ricciardelli E., Cenarro A. J., Rivero-González J. G., Díaz-García L. A., Falcón-Barroso J., 2012, *MNRAS*, 424, 157
- Zahid H. J., Bresolin F., Kewley L. J., Coil A. L., Davé R., 2012, *ApJ*, 750, 120
- Zaritsky D., Kennicutt R. C., Jr, Huchra J. P., 1994, *ApJ*, 420, 87

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