The Astrophysical Journal, 798:99 (21pp), 2015 January 10 doi:10.1088/0004-637X/798/2/99

IZI: INFERRING THE GAS PHASE METALLICITY (Z) AND IONIZATION PARAMETER (q) OF IONIZED NEBULAE USING BAYESIAN STATISTICS

Guillermo A. Blanc[1], Lisa Kewley[2,3], Frédéric P. A. Vogt[2], and Michael A. Dopita[2,3,4]
[1] Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA
[2] Research School of Astronomy and Astrophysics, Australian National University, Cotter Road, Weston, ACT 2611, Australia
[3] Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[4] Astronomy Department, King Abdulaziz University, P.O. Box 80203, Jeddah, Saudi Arabia
Received 2014 July 31; accepted 2014 October 29; published 2014 December 31

ABSTRACT

We present a new method for inferring the metallicity (Z) and ionization parameter (q) of H II regions and star-forming galaxies using strong nebular emission lines (SELs). We use Bayesian inference to derive the joint and marginalized posterior probability density functions for Z and q given a set of observed line fluxes and an input photoionization model. Our approach allows the use of arbitrary sets of SELs and the inclusion of flux upper limits. The method provides a self-consistent way of determining the physical conditions of ionized nebulae that is not tied to the arbitrary choice of a particular SEL diagnostic and uses all the available information. Unlike theoretically calibrated SEL diagnostics, the method is flexible and not tied to a particular photoionization model. We describe our algorithm, validate it against other methods, and present a tool that implements it called IZI. Using a sample of nearby extragalactic H II regions, we assess the performance of commonly used SEL abundance diagnostics. We also use a sample of 22 local H II regions having both direct and recombination line (RL) oxygen abundance measurements in the literature to study discrepancies in the abundance scale between different methods. We find that oxygen abundances derived through Bayesian inference using currently available photoionization models in the literature can be in good (∼30%) agreement with RL abundances, although some models perform significantly better than others. We also confirm that abundances measured using the direct method are typically ∼0.2 dex lower than both RL and photoionization-model-based abundances.

Key words: astrochemistry – galaxies: abundances – galaxies: ISM – H II regions – ISM: abundances

1. INTRODUCTION

The ability to measure chemical abundances in the interstellar medium (ISM) of galaxies is of fundamental importance in astrophysics. We rely on these measurements to constrain models of stellar evolution, galaxy evolution, and the cosmological evolution of the universe as a whole. The abundance of different elements in the ISM is modulated by the processes driving stellar evolution (nucleosynthesis, convection, winds, stellar rotation, mass ejection, supernova explosions, etc.), as well as by the way in which star formation proceeds both spatially and temporally in the ISM of galaxies. Factors such as the accretion and recycling of gas from the intergalactic medium (IGM) and circumgalactic medium, the star formation efficiency and star formation history (SFH), the initial mass function (IMF), and the impact of feedback-driven gaseous outflows all shape the overall chemical structure of galaxies. The universe therefore has a dynamical chemical history in which the abundance of heavy elements increases as a function of cosmic time in a way that traces the efficiency with which gas is turned into stars within galaxies and the way in which these systems interact with the IGM by accreting, enriching, recycling, and ejecting gas (e.g., Larson 1974; Tremonti et al. 2004; Dalcanton et al. 2004; Dalcanton 2007; Brooks et al. 2007; Köppen et al. 2007; Finlator & Davé 2008; Mannucci et al. 2010; Lara-López et al. 2010, 2013; Lilly et al. 2013; Zahid et al. 2014).

This article focuses on the measurement of chemical abundances in nebulae ionized by recent star formation (i.e., H II regions). Ionized gas has been widely used in the literature to trace the chemical composition of the ISM. This is thanks to the fact that transitions from several common elements are observable in emission at UV, optical, and IR wavelengths and that the low densities (10–10^3 cm^{-3}) typical of H II regions translate into an optically thin medium for most of these transitions and a low level of collisional de-excitation. These two conditions greatly simplify the modeling of the emission spectra and allow a relatively straightforward measurement of chemical abundances. There are three main methods used to measure the metallicity or metal abundance[5] in ionized gas, which we describe here briefly. For a review see Stasińska (2004).

The Direct Method. The emissivity of collisionally excited lines (CELs) depends strongly on the electron temperature (T_e). In this method direct measurements of T_e and the electron density (n_e) are used to calculate the emissivity of CELs of particular ions (e.g., [O II] λλ3726, 3729 or [O III] λλ4959, 5007 in the case of singly or doubly ionized oxygen). Ionic abundances can then be inferred from comparing the intensity of these lines to that of hydrogen recombination lines (RLs, typically Balmer lines in the optical). Assuming an ionization correction to account for the abundance of unobserved ions allows one to estimate the total elemental abundance. Auroral to nebular CEL temperature-sensitive ratios are used to measure T_e (e.g., [O III] λ4363/[O III] λ5007) while n_e is typically computed from the relative component intensity of density-sensitive doublets like [O II] λλ3726, 3729 and [S II] λλ6717, 6731. The two main limitations of this method are first that temperature-sensitive auroral lines are too faint to be observed directly in very distant and very high metallicity (i.e., low T_e) sources (∼10^1–10^2 times

[5] While the ideas exposed in this paper are in principle applicable to the chemical abundances of different elements, in this work we focus on the oxygen abundance (12+log(O/H)), which we assume traces the metallicity (Z). Section 2 provides a discussion on the relative abundance patterns used in this work.

fainter than Hβ), and second that the method is sensitive to temperature fluctuations in nebulae that can translate into underestimates of the abundance if a correction is not applied (e.g., Aller 1954; Peimbert 1967; Peimbert & Costero 1969). Furthermore, the existence of T_e gradients in H II regions implies that the direct method saturates for abundances higher than solar (Stasińska 2002).

The Recombination Line (RL) Method. Unlike CELs, the emissivities of RLs of ions show a very mild dependence on T_e and n_e, and their brightnesses relative to hydrogen RLs scale almost directly with the ionic abundance. An ionization correction for unobserved ions is still necessary to obtain total elemental abundances using this method. The RL method does not suffer from the biases associated with temperature fluctuations that affect the direct method and is thought to be more robust (although it might suffer from its own set of systematic uncertainties; c.f. Stasińska 2004). On the other hand, this same direct scaling with abundance translates into extremely faint lines for elements heavier than He ($\sim 10^{-4}$ fainter than Hβ for O and C). Therefore, RL abundance measurements require extremely high S/N, high-resolution spectra and have only been made for He, C, O and Ne in a few dozen bright H II regions in the Milky Way (MW) and the Local Group (e.g., Peimbert 2003; Peimbert et al. 2005; Tsamis et al. 2003; Esteban et al. 2004, 2009; García-Rojas & Esteban 2007; López-Sánchez et al. 2007; Bresolin et al. 2009).

The Strong Emission Line (SEL) Method. In order to be able to measure abundances in faint, distant, and high-metallicity nebulae, several authors have calibrated line ratios combining bright CELs and Balmer lines (together called SELs) as abundance diagnostics. While the brightness of SELs of different elements is strongly affected by other parameters beyond the metal abundance, these calibrations take advantage of the fact that in real H II regions correlations exist between some of these parameters (e.g., T_e, the N/O abundance ratio, the ionization parameter) and the metallicity. By exploiting these "secondary" correlations, useful abundance diagnostic ratios can be constructed out of SELs.

Two different approaches have been taken in the literature to calibrate SEL abundance diagnostics. The first approach is to calibrate SEL ratios against direct method abundances using local samples of H II regions. These are typically referred to as "empirical calibrations." The second approach consists of calibrating SEL ratios as a function of abundance using theoretical photoionization models of H II regions. These are typically referred to as "theoretical calibrations."

Examples of empirical calibrations include those of the $R23$ = ([O II] $\lambda\lambda$3726, 3729 + [O III] $\lambda\lambda$4959, 5007)/Hβ ratio (Pagel et al. 1979; Pilyugin & Thuan 2005), the $O3N2$ = [O III] λ5007/[N II] λ6583 ratio (Alloin et al. 1979; Pettini & Pagel 2004), the $N2$ = [N II] λ6583/Hα ratio (Denicoló et al. 2002; Pettini & Pagel 2004), the $S23$ = ([S II] $\lambda\lambda$6717, 6731 + [S III] $\lambda\lambda$9069, 9532)/Hβ ratio (Vilchez & Esteban 1996; Díaz & Pérez-Montero 2000), and more sophisticated calibrations using combinations of multiple line ratios like those presented in Pilyugin & Mattsson (2011) and Pilyugin et al. (2012).

Theoretical calibrations use predictions from full radiative transfer and excitation/ionization calculations for idealized nebulae to calibrate the abundance dependence of different diagnostics. Early works attempting this include those by Shields & Searle (1978), Pagel et al. (1979), Dufour et al. (1980), Edmunds & Pagel (1984), McCall et al. (1985) and Dopita & Evans (1986). More recent theoretical calibrations for individual diagnostics are given by McGaugh (1991), Kewley & Dopita (2002) and Kobulnicky & Kewley (2004) for the $R23$ ratio and by Kewley & Dopita (2002) for the $O3N2$, $N2$, $S23$, and $N2O2$ = [N II] $\lambda\lambda$6548, 6583/[O II] $\lambda\lambda$3726, 3729 ratios. Even more recently, Dopita et al. (2013) has calibrated several diagnostics based on pairs of abundance- and ionization-sensitive ratios that can be used to simultaneously constrain the metallicity and ionization parameter of ionized nebulae (see also Kobulnicky & Kewley 2004).

The reliability of SEL methods to derive chemical abundances is challenged by the alarming systematic differences seen between different calibrations. These systematic biases include offsets of 0.2–0.6 dex in the absolute abundance scale derived using different calibrations, nonlinearities (i.e., curvature) in the correlations between results from different diagnostics, and dispersions of up to 0.3 dex in abundance around these correlations. These issues have been pointed out by several authors (Peimbert et al. 2007; Bresolin et al. 2009; López-Sánchez & Esteban 2010; Moustakas et al. 2010), and recently Kewley & Ellison (2008) and López-Sánchez et al. (2012) have studied these biases in detail using a large sample of galaxies from the Sloan Digital Sky Survey (SDSS; York et al. 2000) and theoretical H II region photoionization models respectively. In these works a dominant trend arises in which empirically calibrated SEL diagnostics yield abundances that are systematically lower than those derived from theoretically calibrated diagnostics.

Possible reasons behind these discrepancies have been thoroughly discussed in the literature, and López-Sánchez et al. (2012) provide an excellent account. A problem with some SEL diagnostics is that they ignore the dependence of the emission-line ratios on the ionization state of the gas. This is done by simply marginalizing over the ionization parameter q, when calibrating a diagnostic using photoionization models, inducing a large scatter and nontrivial nonlinearities in the method, or in the case of methods calibrated against samples of H II regions with direct T_e abundances by not considering the dispersion and range in ionization parameter of the calibration sample. A good example of the latter case is the $N2$ and $O3N2$ calibrations of Pettini & Pagel (2004), which are widely used for high-redshift galaxies where the density, temperature, and ionization conditions are most likely different from those in the H II regions used to calibrate the method (e.g., Brinchmann et al. 2008; Kewley et al. 2013a, 2013b; Steidel et al. 2014).

Some authors have proposed SEL diagnostics that constrain the ionization state of the gas at the same time as the chemical abundance. These methods typically use recursive techniques or other line ratios to compute either the ionization parameter q (McGaugh 1991; Kewley & Dopita 2002; Kobulnicky & Kewley 2004; Dopita et al. 2013) or the empirically derived excitation parameter P (Pilyugin & Thuan 2005; Pilyugin & Mattsson 2011). Even when the ionization and excitation states of the gas are considered, large systematic differences are still seen between theoretically and empirically calibrated methods.

Another possible source of discrepancy is the potential underestimation of the chemical abundance by the direct method in the presence of density and temperature fluctuations and temperature gradients in ionized nebulae (Peimbert 1967; Stasińska 1978, 2002). H II regions do not have homogeneous density distributions; on the contrary, pockets, filaments, and shells

showing enhanced electron temperatures are ubiquitously observed in these objects (Stasińska 2004 and references within). In regions where T_e is enhanced the emissivity of the temperature-sensitive auroral lines is disproportionally enhanced over that of the stronger nebular lines. This translates into an overestimation of the average T_e and an underestimation of the abundance when integrating the line ratios over the whole nebula. While photoionization models assume a uniform density distribution, they do include gradients in the temperature distribution that somewhat alleviate this problem. A related bias in the measurement of T_e can also arise in the presence of non-Maxwellian electron energy distributions (e.g., the κ distribution introduced by Nicholls et al. 2012, 2013).

Photoionization models, on the other hand, suffer from a series of systematic uncertainties that are not well understood. For starters, they typically use libraries of synthetic stellar population models to generate the input ionizing spectra for stellar populations of particular metallicities, ages, and SFHs (see Section 2). These models are plagued by large uncertainties associated with the effects of stellar rotation, stellar winds, and binarity, especially at the high-mass end of the IMF, which is the most relevant for the ionization of H II regions (e.g., Meynet & Maeder 2007; Eldridge & Stanway 2009; Levesque et al. 2010; Steidel et al. 2014). Furthermore, photoionization models like the ones used in this work typically assume a single-valued function describing the nitrogen-to-oxygen abundance (N/O) as a function of 12+ log O/H. This is a reasonable assumption attempting to capture the primary+secondary nature of nitrogen production, but the observed relation shows a large scatter, and there is still disagreement regarding its actual shape. Furthermore, the relation might change systematically depending on the star formation, accretion, and gas ejection history of galaxies. These assumptions regarding the N/O ratio can have significant effects on diagnostics involving nitrogen lines (Pérez-Montero & Contini 2009; Pérez-Montero 2014; Steidel et al. 2014; Belfiore et al. 2014). Finally, assumptions regarding the ionization-bounded nature of the nebulae and their geometry can also influence the predicted line fluxes.

Given all these systematic uncertainties associated with current SEL diagnostics, a more robust method that does not suffer from these biases is pressingly needed. In this work we propose a technique to estimate the chemical abundance and ionization parameter of H II regions that circumvents some, but not all, of the problems discussed above by removing the need of "calibrating" a particular SEL diagnostic, while instead doing a direct statistical comparison between a theoretical photoionization model and all the emission-line data available to the observer. Our method uses the formalism of Bayesian inference to derive the joint and marginalized probability density functions (PDFs) of the parameters Z and q given the observed line fluxes and errors (the "data"), a grid of emission-line ratios predicted by a theoretical photoionization model (the "model"), and any "prior" information on Z and q. We also provide and describe a publicly available user-friendly IDL implementation of our method called IZI[6] (inferring metallicities (Z) and ionization parameters).

A similar approach to the one presented here was adopted by Brinchmann et al. (2004) and Tremonti et al. (2004) to measure the physical properties of star-forming galaxies in the SDSS. More recently, Pérez-Montero (2014) also presented a method based on χ^2 minimization that uses information from multiple emission lines to constrain the metallicity and ionization parameter. Also worth mentioning is the pioneering work of Garnett & Shields (1987) on a sample of H II regions in M81. This work was among the first to explore the simultaneous use of multiple emission-line ratios from different elements to constrain the properties of H II regions by comparing to photoionization model grids.

After presenting our new method (Section 2) and the way in which it is implemented in IZI (Section 3), we apply it to a sample of 186 extragalactic H II regions from van Zee et al. (1998b) (Section 4). We use this sample of H II regions to validate our method against the results from the interpolation method based on pairs of emission-line diagnostics recently presented in Dopita et al. (2013).

The Bayesian formalism provides us with a tool to evaluate how much information regarding the metallicity and the ionization parameter is carried by different subsets of the data (i.e., by different emission lines). This provides an opportunity to evaluate the intrinsic performance of different diagnostics in terms of constraining these two parameters. By intrinsic performance we mean the baseline amount of information carried by a particular line ratio, independent of any biases in its calibration. We evaluate the performance of a series of commonly used SEL diagnostics in the literature and present these results in Section 5.

We also apply IZI to a sample of local H II regions having both direct method and RL abundance measurements in order to explore the abundance scale discrepancies discussed above. Our method allows us to perform this comparison using several input photoionization models available in the literature. The RL abundances provide an important reference point from which to obtain insight regarding the discrepancy between direct method and empirical SEL abundances and those derived from photoionization models. We present the results of this comparison in Section 6. Finally, we summarize our work and provide our conclusions in Section 7.

2. THE METHOD: BAYESIAN INFERENCE OF Z AND q

In this section we describe the statistical method used to derive the parameters Z and q from a set of observed SEL fluxes. In the context of Bayesian statistics this is a "parameter estimation" problem in which we want to estimate the joint and marginalized PDF for the parameters $\boldsymbol{\theta} = \{\log(Z), \log(q)\}$ given the data $D = \{\boldsymbol{f}, \boldsymbol{e}\}$ (where $\boldsymbol{f} = \{f_1 \dots f_n\}$ are the observed extinction-corrected line fluxes and $\boldsymbol{e} = \{e_1 \dots e_n\}$ their associated measurement errors), the model M given by a photoionization model grid of line fluxes $\boldsymbol{f'}(\boldsymbol{\theta}) = \{f'_1(\boldsymbol{\theta}) \dots f'_n(\boldsymbol{\theta})\}$, and any prior information I regarding the parameter values. We follow the notation of Gregory (2005).

From Bayes's theorem it can be shown that the joint posterior PDF of $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{\theta} \,|D, M, I) = \frac{p(\boldsymbol{\theta} \,|M, I)\; p(D \,|M, \boldsymbol{\theta}, I)}{p(D \,|M, I)}. \quad (1)$$

For simplicity, following the notation of Gregory (2005) we omit I in the following equations, although it is understood that all conditional probabilities are calculated given the prior information on the parameter values. We can then rewrite Equation (1) as

$$p(\boldsymbol{\theta} \,|D, M) = \frac{p(\boldsymbol{\theta} \,|M)\; p(D \,|M, \boldsymbol{\theta})}{p(D \,|M)}. \quad (2)$$

[6] http://users.obs.carnegiescience.edu/gblancm/izi

The first term in the numerator of the right-hand side reflects the prior probability on the parameters $\boldsymbol{\theta}$ given the model M and all available prior information on the parameters. The second term in the numerator of the right-hand side of Equation (2) is the "likelihood" of the data given the model evaluated for parameters $\boldsymbol{\theta}$. Assuming Gaussian errors in the flux, this term is given by

$$p(D\,|M,\boldsymbol{\theta}) = 2\pi \prod_{1}^{n} \frac{\exp\left(-\frac{(f_i - f_i'(\boldsymbol{\theta}))^2}{2\left(e_i^2+\epsilon^2 f_i'(\boldsymbol{\theta})^2\right)}\right)}{\sqrt{\left(e_i^2+\epsilon^2 f_i'(\boldsymbol{\theta})^2\right)}}, \quad (3)$$

where n is the number of observed emission lines and ϵ is a term that takes into account the systematic fractional error in the flux predicted by the photoionization model. As a zeroth-order approximation we assume that this systematic error is Gaussian and is the same for all modeled emission lines. For the results presented in this paper we adopt a fractional error of 0.1 dex (Kewley & Ellison 2008; Dopita et al. 2013). In principle, ϵ could be turned into an extra "nuisance" parameter in the calculation. For the sake of computing speed and simplicity, we do not adopt this approach, although we expect to experiment with this in the future. The denominator in the right-hand side of Equation (2) is the "global likelihood" of the data for model M. This is equivalent to a normalization factor and can be calculated by demanding the integral of $p(\boldsymbol{\theta}\,|D, M)$ to be unity.

Once the joint posterior PDF for $\boldsymbol{\theta}$ is known, we can compute the marginalized posterior PDF for $\log(Z)$ and $\log(q)$ by integrating the joint posterior PDF over the other parameter, respectively, so

$$p(\log(Z)\,|D, M) = \int_{-\infty}^{\infty} p(\boldsymbol{\theta}\,|D, M)\, d\log(q), \quad (4)$$

$$p(\log(q)\,|D, M) = \int_{-\infty}^{\infty} p(\boldsymbol{\theta}\,|D, M)\, d\log(Z). \quad (5)$$

The joint and marginalized posterior PDFs can be used to derive "best-fit" values and confidence limits for Z and q by either searching for the maximum in the distributions (equivalent to a maximum likelihood approach in the case of uniform priors) or computing the moments of the distributions. In some cases the shape of the PDF will be such that a best-fit value will not be a well-defined quantity. This can happen when the data only provide upper or lower limits on a parameter, in which case the PDF is only bounded on one side, or in cases in which the PDF shows a complex topology with multiple probability peaks reflecting multiple solutions.[7] In the next section we describe how the method outlined above is implemented in IZI and how we deal with these sorts of complications.

This Bayesian approach has six major advantages over classical SEL methods to derive chemical abundances and ionization parameters in ionized nebulae:

1. This method removes the arbitrary choice of a particular SEL diagnostic and therefore does not suffer from many of the systematic uncertainties associated with the calibration of these diagnostics.
2. The method can make use of all the information available to the observer, as an arbitrary set of emission lines can be used as long as they have been modeled as part of the photoionization grid.
3. The method allows for the inclusion of upper limits in emission-line fluxes as the Bayesian approach allows us to include censored data in the likelihood calculation in a straightforward manner. This is particularly useful when studying the properties of high-redshift galaxies where upper limits in the line fluxes are common.
4. IZI is implemented in a flexible fashion that allows the use of user-provided photoionization model grids. Therefore, unlike SEL theoretical calibrations in the literature, the method is not "married" to a particular set of models. This implies that IZI can be run using different input photoionization models in order to test how the results change as a function of the underlying assumptions in the models.
5. In certain cases the PDFs of Z and q can have complicated topologies, showing multiple local maxima and asymmetric probability tails that reflect non-Gaussian uncertainties in the derived values. Since the method does a full calculation of the joint and marginalized PDFs, it naturally yields information about the asymmetry in the error bars of the derived parameters, about the existence of multiple degenerate solutions, and on whether a parameter is being constrained or only a limit can be set on it.
6. IZI is fast, is easy to use and to incorporate as a subroutine into larger pieces of code, and is publicly available for use by the astronomical community.

In the following sections we describe how this method is implemented and applied to different samples of H II regions.

3. IZI: A USER-FRIENDLY IDL IMPLEMENTATION

We have implemented the method described in Section 2 as an IDL function called IZI. Here we provide a detailed description of how this function computes the chemical abundance and ionization parameter from a set of input emission-line fluxes and a photoionization model grid.

3.1. Input SEL Fluxes

Input fluxes and errors are assumed to be corrected for dust extinction. All fluxes are normalized to the flux of the Hβ line or, if the flux of this line is not provided, to the flux of the brightest line for which a measurement is available. Upper limits on the fluxes of particular lines can be provided, and the algorithm deals with them by integrating the argument of the product in Equation (3) below these limits. Besides the measured line fluxes and errors, the user must also provide an array identifying the name of each line. These IDs are also stored in the input photoionization model grids and are used to properly match the data and the models. IZI allows the user to define sets of lines for which the flux must be co-added before fitting to allow for low spectral resolution measurements that do not separate the flux of neighboring transitions.

3.2. Input Model Grids

Our method is flexible in the sense that it can use an arbitrary model grid as long as the observed emission lines have been modeled. This is an advantage of IZI as, unlike theoretically calibrated SEL diagnostics, the method is not married to a particular set of photoionization models. We stress that any

[7] A common example of this situation occurs when using the lines that compose the $R23$ diagnostic, which is known to be double valued in metallicity.

arbitrary model can be provided as input to IZI as long as it is stored in the appropriate format. In principle, this is not restricted to photoionization models but could also include empirically based grids of emission-line fluxes as a function of abundance and ionization/excitation parameter, similar to the approach used in the C-method of Pilyugin et al. (2012). Although we have not experimented with this possibility, this would be an interesting application for IZI. Here we briefly describe the photoionization models used throughout this paper. We refer the reader to the original references for a more detailed description.

3.2.1. Kewley et al. (2001) Models

Kewley et al. (2001) present a suite of photoionization models aimed at reproducing the emission-line properties of a sample of starburst galaxies in the local universe. The authors use the MAPPINGS-III photoionization code (Sutherland & Dopita 1993) to compute the nebular emission spectra of idealized H II regions ionized by synthetic stellar populations. Kewley et al. (2001) use both the PEGASE v2.0 (Fioc & Rocca-Volmerange 1997) and STARBURST 99 (Leitherer et al. 1999) stellar population synthesis codes to compute the input ionizing spectra. Here we only adopt the STARBURST 99 models, which use the plane-parallel stellar atmosphere models of Lejeune et al. (1997) plus the extended model atmospheres of Schmutz et al. (1992) for stars with strong winds in combination with the Geneva stellar evolution tracks (Schaller et al. 1992). The input spectra are computed assuming a constant SFH, a Salpeter IMF (Salpeter 1955) with a lower-mass cutoff of 0.1 $M_{\odot}$ and an upper-mass cutoff of 120 $M_{\odot}$, and a stellar population age of 8 Myr, which allows for the stellar birth and death rate to balance for stellar masses $>20\,M_{\odot}$ and for Wolf–Rayet star emission to develop and contribute to the ionizing spectrum.

The synthetic spectra ionize MAPPINGS-III isobaric plane-parallel photoionization models with electron densities of $n_e = 10$ cm^{-3} and 350 cm^{-3}. In these models the ionization parameter[8] is defined at the inner boundary of the nebular models by the following expression:

$$q = \frac{Q_{H^0}}{n_H}, \tag{6}$$

where Q_{H^0} is the flux of ionizing photons above the Lyman limit and n_H is the particle density. For all elements except He and N the MAPPINGS-III calculations adopt an undepleted solar abundance pattern from Anders & Grevesse (1989) with $12 + \log\,\mathrm{O/H}_{\odot} = 8.93$ and assume primary production. For He a primordial nucleosynthesis component is added and the N abundance follows a broken power law as a function of metallicity in an attempt to account for secondary N production above 0.23 $Z_{\odot}$. The MAPPINGS-III code includes the effects of dust absorption, dust photoelectric heating, and element depletion as described in Dopita et al. (2000).

3.2.2. Levesque et al. (2010) and Richardson et al. (2013) Models

Levesque et al. (2010) presents new MAPPINGS-III photoionization models aimed at reproducing the nebular spectra of large samples of galaxies in the local universe. This new version of the MAPPINGS-III models is similar in nature to the Kewley et al. (2001) models described above but includes a series of upgrades with respect to the older models. Levesque et al. (2010) adopt an updated version of the STARBURST 99 population synthesis code (Vázquez & Leitherer 2005) that uses the Pauldrach/Hillier stellar atmosphere models (Pauldrach et al. 2001; Hillier & Miller 1998). These models, unlike the Lejeune/Schmutz models used in Kewley et al. (2001), include a detailed non-LTE modeling of metal line blanketing, which significantly affects the shape of the ionizing spectrum, making it harder in the 1–4 Ry region. Levesque et al. (2010) used two updated versions of the Geneva tracks with "standard" and "high" mass-loss rates (Meynet et al. 1994). Here we adopt the models calculated using the "high" mass-loss tracks, a constant SFH, a Salpeter IMF with a 100 $M_{\odot}$ upper cutoff, and an age of 6 Myr.

These synthetic spectra are used to ionize an updated version of the MAPPINGS-III photoionization models that includes a more rigorous treatment of dust that includes the effects of radiation pressure on grains (Groves et al. 2004). Levesque et al. (2010) also adopt isobaric plane-parallel models but with electron densities of $n_e = 10$ cm^{-3} and 100 cm^{-3}. The ionization parameter is defined in the same way as in the Kewley et al. (2001) models. The solar abundance pattern is updated to the revised abundance set of Grevesse et al. (2010) with $12 + \log\,\mathrm{O/H}_{\odot} = 8.69$. Richardson et al. (2013) presents an extension of these models that samples very high values for the ionization parameter. These models, which are aimed at reproducing the emission-line properties of high-redshift Lyα emitters, are computed following the same prescriptions as in Levesque et al. (2010) and therefore can be merged with these grids in a straightforward manner.

3.2.3. Dopita et al. (2013) Models

Recently, Dopita et al. (2013) presented photoionization model grids computed with a significantly upgraded version of MAPPINGS. The MAPPINGS-IV code has been updated to include the latest atomic data, an increased number of ionic species treated as full non-LTE multilevel atoms, and the ability to use electron energy distributions that differ from a simple Maxwell–Boltzmann (M-B) distribution. Dopita et al. (2013) adopt a κ electron energy distribution that shows a power-law-shaped high-energy tail. This distribution is inspired by observations of plasmas in the solar system, and its inclusion is found to solve many discrepancies in the measurement of the electron temperatures in planetary nebulae and H II regions (Nicholls et al. 2012, 2013). The grids used here are calculated by adopting a κ electron energy distribution with values of $\kappa = 20$ and $\kappa = \infty$ (the latter being equivalent to an M-B distribution). The new version of the code also uses the revised solar abundance set from Grevesse et al. (2010) and new smooth functions to parameterize the abundance of both N and C as a function of O abundance to account for secondary production of these two elements.

The input ionizing spectrum is computed using the STARBURST 99 population synthesis code adopting the Lejeune/Schmutz extended stellar atmosphere models (as in Kewley et al. 2001), a constant SFH, a Salpeter IMF with a lower-mass cutoff of 0.1 $M_{\odot}$ and an upper-mass cutoff of 120 $M_{\odot}$, and an age of 4 Myr. Unlike the Kewley et al. (2001) and Levesque et al. (2010) models, Dopita et al. (2013) computes isobaric photoionization models with an electron density $n_e \simeq 10$ cm^{-3} and a spherical geometry for which the radial divergence of the radiation field must be taken into account when computing the ionization parameter. For these spherical models the ionization parameter is defined at the inner boundary of the

[8] Sometimes reported in the literature in its dimensionless form $\mathcal{U} = q/c$, with c the speed of light.

nebulae by

$$q = \frac{Q_{H^0}}{4\pi R_s^2 n_H} \propto (Q_{H^0}\, n_H)^{1/3}, \quad (7)$$

where R_s is the Strömgren radius of the spherical H II region (Strömgren 1939).

3.3. Priors and Limits on log(Z) and log(q)

By default we assume a maximum ignorance situation and use a uniform PDF to compute the first term in the numerator of the right-hand side of Equation (2). Notice that since we use the logarithm of Z and q as the model parameters, this is equivalent to using a Jeffreys prior on both parameters, implying that we assume uniform probability per decade in the parameters Z and q (Gregory 2005). We do this instead of adopting a uniform prior in linear space since the latter assigns an unreasonably high probability to high abundance and high ionization parameter solutions. IZI also allows the user to provide an arbitrary prior PDF for each parameter independently. For example, one could introduce a prior for Z based on the stellar mass of a star-forming galaxy and the observed scatter in the $M-Z$ relation. We assume that the parameters live within a range $Z_{\rm min}$, $Z_{\rm max}$ and $q_{\rm min}$, $q_{\rm max}$, which by default are taken as the minimum and maximum values modeled in the input photoionization grid, but that can also be set by the user. The prior probability is set to zero outside this range. The implications of this last assumption for the calculation of upper and lower limits in Z and q are discussed in Section 3.5.

3.4. Grid Interpolation and Calculation of the PDFs

We conduct a full calculation of the joint posterior PDF over a finely spaced grid in both log(Z) and log(q). Since the input photoionization grids typically sample a coarse set of values for these parameters, before computing the PDF we use either a bi-linear or a quintic surface interpolation to resample the line flux values in the input model onto a finely spaced grid. By default the grid has $2500 = 50 \times 50$ equally spaced elements in the log(Z)–log(q) plane, within the $Z_{\rm min}$, $Z_{\rm max}$ and $q_{\rm min}$, $q_{\rm max}$ ranges. IZI allows for the number of elements to be set by the user, but computing time is strongly affected by this choice.

Once the input model has been interpolated, we use Equations (2) and (3) to calculate the joint posterior PDF at each element of the parameter grid. The marginalized posterior PDF for each parameter is calculated then using Equations (4) and (5).

3.5. Calculations of Best-fit Values and Confidence Intervals

Choosing how to define a "best-fit" value for a parameter is nontrivial when its PDF is non-Gaussian. Strong asymmetries and the presence of multiple peaks (i.e., multiple likely solutions) in the PDF can deviate the mean of the PDF from the most likely value (mode), making the mean not representative of a likely solution for the parameter in question. This is, of course, alleviated by computing the full PDF, which in principle removes the need to calculate a best-fit value in the first place. In practice, however, for many applications it is not straightforward to propagate the full PDF of each parameter through all subsequent calculations in the analysis, and having a best-fit value complemented by a confidence interval is desirable. Here we describe how we calculate the best-fit values for Z and q, although we remind the reader that a proper analysis should make use of the full PDFs (e.g., Brinchmann et al. 2004).

We calculate three different best-fit values for log(Z) and log(q), respectively. The first is the "marginalized mean," given by

$$\langle \log(Z) \rangle = \int_{Z_{\rm min}}^{Z_{\rm max}} p(\log(Z)\,|D, M) \log(Z)\, d\log(Z), \quad (8)$$

$$\langle \log(q) \rangle = \int_{q_{\rm min}}^{q_{\rm max}} p(\log(q)\,|D, M) \log(q)\, d\log(q). \quad (9)$$

The other two are the modes of the joint posterior PDF and of the marginalized posterior PDFs, that is, the values of log(Z) and log(q) that maximize either the joint PDF or the marginalized PDF of each parameter independently. We refer to these from here on as the "joint mode" and "marginalized mode."

We calculate a single confidence interval for each parameter by defining the high-probability region in the log(Z)–log(q) plane that encloses 68.3% of the integrated joint posterior probability density. The parameter values that bound this region are taken as our 1σ uncertainty on the best-fit values. Using this definition, we find that the confidence interval always bounds the three "best-fit" values defined above.

3.6. Identifying Problematic PDFs

IZI automatically analyzes the shape of the marginalized PDFs in order to flag PDFs that are not "well behaved" in the sense that the best-fit values discussed above might not be representative of the actual parameter values. We use the first and second derivatives of the marginalized PDFs to search for maxima, and we characterize the number of peaks in the distributions, flagging objects that show multiple possible solutions in Z and q. In these cases the marginalized mean is not ensured to fall close to a region of high probability, and the joint and marginalized modes might not sample the highest probability solution as a wider, lower-amplitude peak in the PDF might be actually more likely than a narrower, higher-amplitude peak (i.e., the mode) once the probability density is integrated over them.

We also study the PDFs to check whether the solutions are bounded or only provide upper or lower limits or no limits at all on Z and q. A solution is considered bound if the probability density at the ends of the parameter range ($Z_{\rm min}$, $Z_{\rm max}$, and $q_{\rm min}$, $q_{\rm max}$) is lower than 50% of the marginalized mode. Otherwise, the object is flagged as showing an upper limit, lower limit, or no limit at all on the corresponding parameter.

4. APPLICATION TO REAL SOURCES AND VALIDATION OF THE METHOD

In this section we apply the method described in the last two sections to a sample of 186 extragalactic H II regions observed and catalogued by (van Zee et al. 1998b, hereafter V98). We also compare the results of IZI against measurements of the metallicity and ionization parameter for these same objects using the methods described in Dopita et al. (2013). The goal of this section is to test the performance of IZI on real data and validate our method.

4.1. Application to Observed Local H II Regions

We use the data of V98, who uniformly obtained low-resolution ($\Delta\lambda = 7.8$ Å) spectra for 186 extragalactic H II regions in a sample of 13 nearby spiral galaxies using the Double Spectrograph on the 5 m Palomar telescope. Their data cover

the 3500–7600 Å range and therefore include the [O II] λλ3726, 3729, Hβ, [O III] λ5007, Hα, [N II] λ6583, and [S II] λλ6717, 6731 lines that we use here.[9] The publicly available catalog reports reddening-corrected emission-line fluxes (normalized to Hβ) and errors for all the above transitions. This sample was also used by Dopita et al. (2013) to evaluate the performance of their newly proposed metallicity and excitation diagnostics, so it provides a good reference for comparing the results of IZI to that work. In order to allow for this comparison in this section, we use the $\kappa = 20$ photoionization models of Dopita et al. (2013) as input to IZI.

The top panels of Figure 1 present the joint and marginalized PDFs calculated by IZI using all of the above SELs for a single H II region in the V98 sample (NGC 925 +087−031). For this particular object both Z and q are well constrained and show positive covariance. The marginalized PDF for $\log(Z)$ is roughly Gaussian, while the $\log(q)$ PDF shows some asymmetry toward large values. The three "best-fit" values for the two parameters in the model are shown by the red (joint mode), blue (marginalized mode), and green (marginalized mean) solid lines in the two rightmost panels. The confidence intervals described in Section 3.5 are shown by the cyan curve in the joint PDF and the dotted red lines in the marginalized PDFs. For the metallicity it is seen that all three estimations agree well within the 1σ confidence limits. The joint mode is $12+\log \mathrm{O/H} = 8.70^{+0.22}_{-0.15}$, close to the solar value. For the ionization parameter the joint mode is $\log q = 7.19^{+0.37}_{-0.16}$, and the marginalized mean is offset to a slightly higher value because of the asymmetry in the PDF, although all three estimations fall well within the 1σ confidence limits.

To exemplify the behavior of the PDF as a function of the available input information, the second row of panels in Figure 1 shows the joint and marginalized PDFs now calculated using only the [O II] λλ3726, 3729, Hβ, and [O III] λ5007 transitions. This is similar to using the $R23$ diagnostic. The double-valued nature of $R23$ as a metallicity diagnostic is caused by the strengthening of the oxygen lines with respect to Hβ as the abundance increases, followed by a weakening of these lines toward higher abundances ($12+\log \mathrm{O/H} \gtrsim 8.5$) as a higher metal cooling rate causes the temperature, and therefore the rate of collisional excitations to the upper levels of these oxygen transitions, to decrease. This degeneracy has been extensively documented in the literature (e.g., Pagel et al. 1979; Kobulnicky & Kewley 2004; Kewley & Ellison 2008) and arises naturally when using our method as two high-probability peaks seen in the joint PDF.

In this case, since IZI uses the individual line fluxes for the input transitions and not the sum of the singly and doubly ionized oxygen lines (as in the $R23$ diagnostic), it also constrains the ionization parameter although it shows a broader and more asymmetric PDF than when all transitions are used. In this sense, using IZI with these transitions as input is somewhat similar to the $R23$ method of Kobulnicky & Kewley (2004), in which the ratio of [O III] to [O II] is used in combination with $R23$ to constrain the ionization parameter and the metallicity simultaneously using a recursive method.

The third row of panels in Figure 1 shows the PDFs calculated by IZI when using only the [O II] λλ3726, 3729 and [N II] λ6583 lines (i.e., equivalent to the $N2O2$ diagnostic). The similar ionization potentials of N^+ and O^+ translate into this line ratio being very insensitive to the ionization parameter, as can be seen in Figure 1, where q is absolutely unconstrained. On the other hand, both the secondary nature of N production and the fact that the energy of the upper level of the [N II] doublet is lower than that of the upper level for the [O II] doublet (making its line flux less sensitive to a drop in T_e toward higher abundances) translate into this line ratio being very sensitive to metallicity (Dopita et al. 2000; Kewley & Dopita 2002). This can be seen in the PDFs shown in Figure 1, where the metallicity is well constrained while the ionization parameter is completely unconstrained. It is worth mentioning that $N2O2$ is very sensitive to the N/O abundance ratio (Pérez-Montero & Contini 2009; Pérez-Montero 2014) and that its good performance as a metallicity diagnostic relies heavily on the assumption of a correlation between N/O and O/H in the photoionization models being used here. We discuss this further in Section 5.

Finally, the bottom row of panels in Figure 1 shows the resultant PDFs if only the [N II] λ6583 and Hα lines are fed to IZI as input (i.e., equivalent to the $N2$ diagnostic). In this case the ionization parameter is completely unconstrained, and the abundance shows solutions for two branches (see Figure 2 and discussion in Section 5.2). The topology of the joint PDF translates into three peaks in the marginalized PDF for the abundance. In each of these branches the best-fit abundance is highly correlated with the ionization parameter. It is interesting to note that if the ionization parameter is constrained to the value derived when using all available lines, the correct value for the abundance is recovered in the low-abundance branch. This reflects the fact that the performance of $N2$ as an abundance diagnostic relies heavily on the fact that the ionization parameter of local H II regions is limited to a fairly narrow range and that a large majority of them are in the low-abundance branch of $N2$. We discuss this further in Section 5.2.

In order to visualize where the constraining information on the parameters Z and q is coming from, Figure 2 presents eight line ratios commonly used in the literature as metallicity and excitation diagnostics, as a function of oxygen abundance and ionization parameter. The solid gray lines show the input photoionization grids for different values of q in the top eight panels and different values of Z in the bottom eight panels. The best-fit solution (joint mode) for NGC 0925 +087−031 (using all available lines) is shown by the solid red line and the black error bars, and the families of allowed photoionization models within the 1σ confidence limits are bounded by the dashed red lines. All panels showing dependence against Z (q) sample a common range of 3 dex (4 dex) in the line ratios and therefore provide a fair picture of how sensitive these ratios are to these parameters.

Many insightful conclusions can be drawn from Figure 2. First, the $R23$, $N2O2$, $N2$, and $N2S2$ diagnostics are better metallicity indicators than the $O3N2$, $O3O2$, $R3$, and $S2$ ratios. The former show less dependence on the ionization parameter, which translates into a small scatter as a function of metallicity. The $R23$, $N2$, and $N2S2$ diagnostics show mild dependences with q in different regions of the abundance range, which translates into larger scatter for these diagnostics in these regions. The $N2$ diagnostic also shows a flattening toward the high-abundance end that is caused by the same process that flattens and inverts the behavior of $R23$ (i.e., a drop in T_e toward high abundances), so this diagnostic is also double valued (see discussion in Section 5.2). The $N2O2$ diagnostic, on the

[9] While V98 reports the summed intensity of the [O III] and [N II] doublets, here we assume the theoretical line ratio to compute the flux of [O III] λ5007 and [N II] λ6583. This is done for consistency with Dopita et al. (2013) in order to allow the comparison presented in Section 4.2.

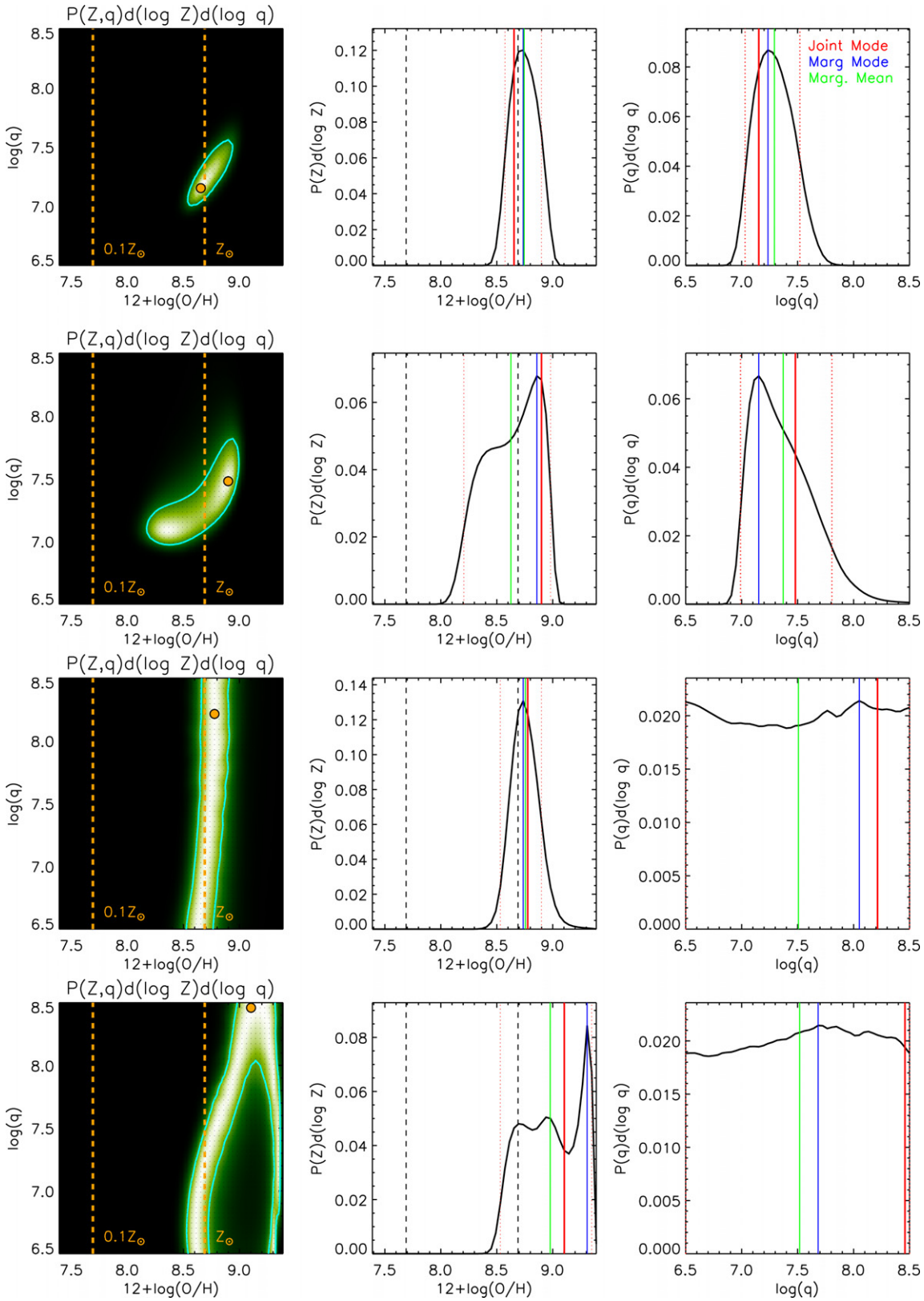

Figure 1. First row: joint (left) and marginalized (center, right) PDFs for $\log Z$ and $\log q$ calculated by `IZI` for the H II region NGC 0925 +087−031 using all emission-line measurements reported in V98 (i.e., [O II] λλ3726, 3729, Hβ, [O III] λ5007, Hα, [N II] λ6583, and [S II] λλ6717, 6731). The orange circle and cyan line in the left panel show the mode of the joint PDF and the 1σ confidence level. The dashed orange (black) lines in the left (middle) panel mark the adopted solar abundance and 1/10 of its value. On the middle and right panels the joint mode, marginalized mode, marginalized mean, and 1σ confidence interval are shown by solid red, blue, green, and dashed red vertical lines, respectively. Second row: same as above, but only using the [O II] λλ3726, 3729, Hβ, and [O III] λ5007 lines (i.e., equivalent to $R23$). Third row: same as above, but only using the [O II] λλ3726, 3729 and [N II] λ6583 lines (i.e., equivalent to $N2O2$). Bottom row: same as above, but only using the Hα and [N II] λ6583 lines (i.e., equivalent to $N2$).

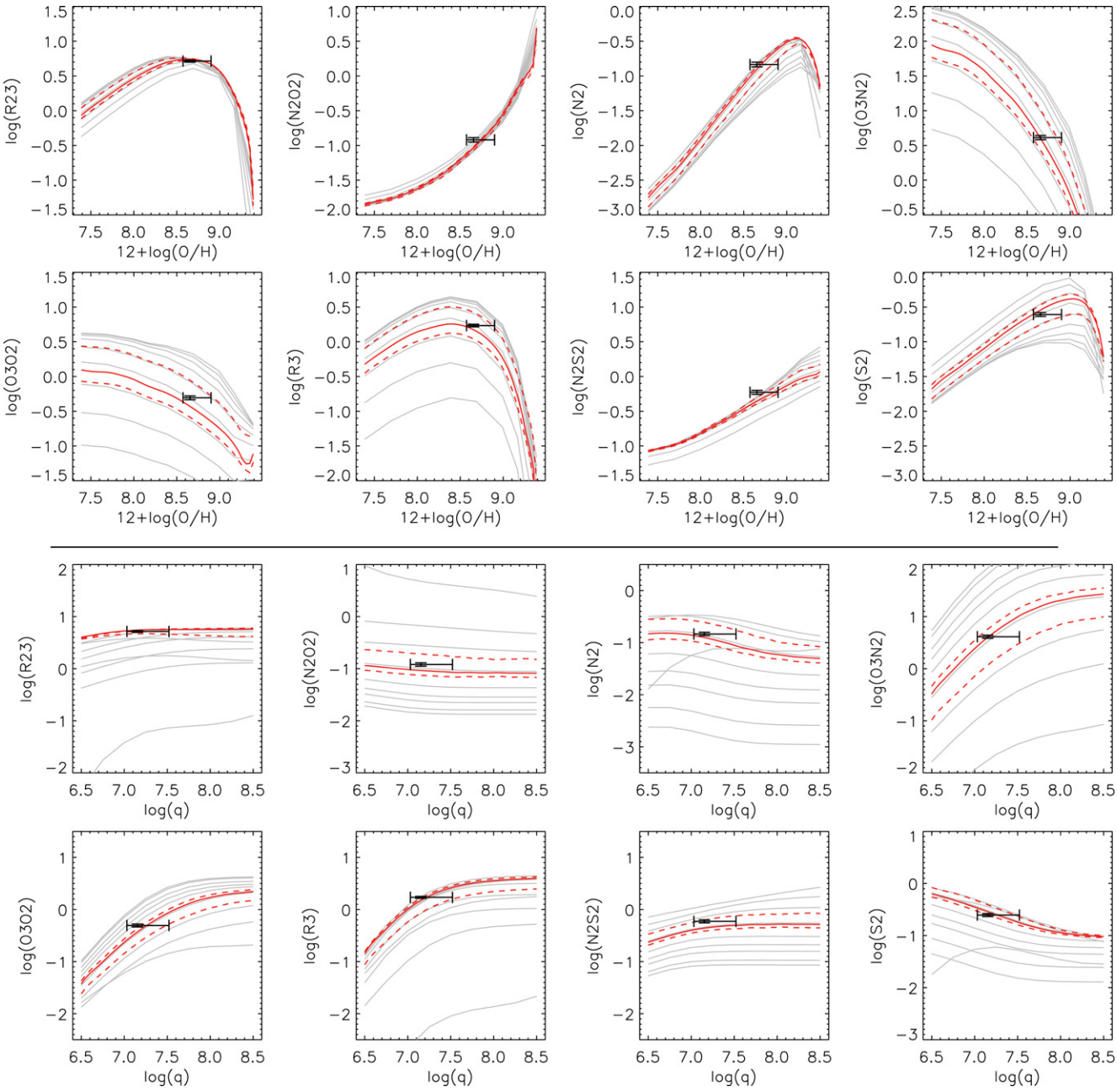

Figure 2. Eight line ratios commonly used in the literature as metallicity and excitation diagnostics ($R23$ = ([O II] λλ3726, 3729+[O III] λ5007)/Hβ; $N2O2$ = [N II] λ6583/[O II] λλ3726, 3729; $N2$ = [N II] λ6583/Hα; $O3N2$ = [O III] λ5007/[N II] λ6583; $O3O2$ = [O III] λ5007/[O II] λλ3726, 3729; $R3$ = [O III] λ5007/Hβ; $N2S2$ = [N II] λ6583/[S II] λλ6717, 6731; and $S2$ = [S II] λλ6717, 6731/Hα), as a function of oxygen abundance (top eight panels) and ionization parameter (bottom eight panels). The solid gray lines show the input photoionization model grids (Dopita et al. 2013) for different values of q in the top eight panels and different values of Z in the bottom six panels. The best-fit solution (joint mode) for NGC 0925 +087−031 is shown by the solid red line and the black error bars, and the families of allowed photoionization models within the 1σ confidence limits are bounded by the dashed red lines.

other hand, is monotonic over the whole abundance range and shows the least dependence on q and therefore the smallest scatter as a function of metallicity. These two factors make it a good abundance diagnostic, modulo variations in the N/O ratio, which can be significant.

The $O3N2$, $O3O2$, $R3$, and $S2$ line ratios show a strong dependence on the ionization parameter. In fact, $O3O2$ has long been recognized as a good excitation diagnostic (e.g., McGaugh 1991; Lilly et al. 2003; Nakajima et al. 2013), and out of all the line ratios shown in Figure 2, it shows the smallest dependence on the metallicity. Nevertheless, these line ratios are sometimes used in the literature as metallicity diagnostics, empirically calibrated against local samples of H II regions (e.g., Nagao et al. 2006; Maiolino et al. 2008; Mannucci et al. 2009; Cullen et al. 2014). The reason why these empirical calibrations appear to trace metallicity with a reasonably moderate scatter is that the local samples against which they are calibrated only span a restricted range in ionization parameters. Therefore, extrapolating the use of these calibrations to other objects that might have different excitation conditions (e.g., high-redshift galaxies) can introduce significant biases in the derived abundances (Kewley et al. 2013a; Steidel et al. 2014). This is a possible reason behind the discrepancies seen among different authors studying the behavior of the mass–metallicity–SFR

relation as a function of redshift (e.g., Mannucci et al. 2010; Troncoso et al. 2014; Cullen et al. 2014).

Figure 2 also exemplifies how IZI allows the information from all available emission lines (and upper limits if available) to be used in order to constrain the values of Z and q. This is a major improvement over classical metallicity and excitation diagnostics, which are typically limited to either a single SEL ratio or a pair of SEL ratios.

4.2. Comparison to pyqz

In order to validate the output of IZI, we compare our results for the full V98 sample against those produced by the publicly available Python module pyqz[10] presented in Dopita et al. (2013). The pyqz code uses pairs of abundance- and excitation-sensitive line ratios to define a plane in which the metallicity and ionization parameter can be determined by interpolating the photoionization model grid to match the observed line ratios.

Dopita et al. (2013) shows that parameter estimations using their new updated photoionization model grids and pyqz on four different pairs of line ratios based on combinations of the [N II]/[O II] and [N II/S II] abundance-sensitive ratios and the [O III]/[O II], [O III]/[S II], and [O III]/Hβ excitation-sensitive ratios yield self-consistent results. Figures 3 and 4 present a comparison between the results of IZI for the full V98 H II region sample and the average of eight pyqz metallicity and ionization parameter estimations based on the following pairs of line ratios:

1. [N II]/[O II] versus [O III]/[O II]
2. [N II]/[O II] versus [O III]/[S II]
3. [N II]/[S II] versus [O III]/Hβ
4. [N II]/[S II] versus [O III]/[S II]
5. [N II]/[O II] versus [O III]/Hβ
6. [N II]/[S II] versus [O III]/[O II]
7. [N II]/Hα versus [O III]/Hβ
8. [N II]/Hα versus [O III]/[O II]

The IZI values are calculated using all the emission-line fluxes reported in V98 and the Dopita et al. (2013) model grids for $\kappa = 20$. The exact same information is used in the pyqz calculations, so any differences between the two methods arise from the different approach taken to estimate the Z and q parameters, that is, the difference between the average of eight different interpolations using different pairs of abundance- and excitation-sensitive ratios, and a full posterior PDF calculation followed by a best-fit parameter estimation.

Figure 3 shows good agreement between the "joint mode" best-fit abundance (top panel) as calculated by IZI and the average of the eight pyqz estimations. The measurements agree well within the IZI 1σ confidence limits, with a median offset of -0.03 and a scatter of 0.08 dex. Note that the median error bar reported in Figures 3 and 4 for the pyqz measurements corresponds to the standard deviation between the eight measurements and does not include either measurement errors or systematic uncertainties, so they are surely underestimated.

The marginalized mode (second panel of Figure 3) shows similar agreement with the average pyqz values, but significant systematic deviations at the 0.2 dex level can be seen when using the marginalized mean (bottom panel), particularly at the low-abundance end ($12+\log \mathrm{O/H} < 8.7$). This is caused by

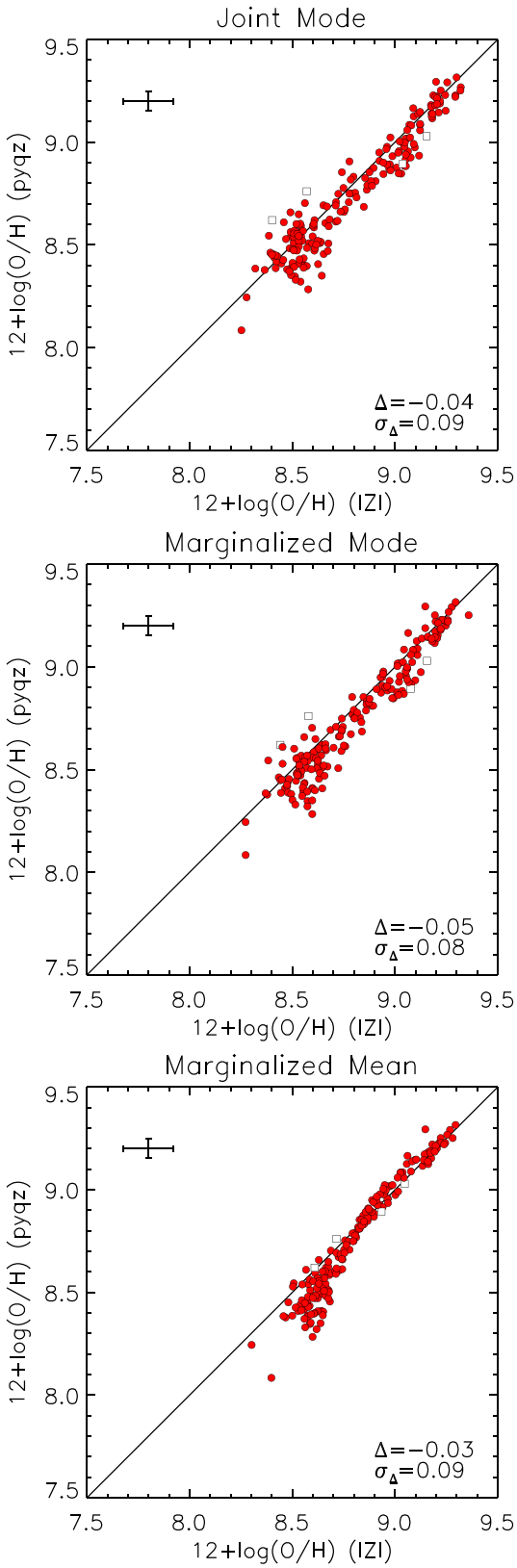

Figure 3. Comparison between the joint mode (top), marginalized mode (middle), and marginalized mean (bottom) best-fit values for the metallicity estimated with IZI for the V98 sample using all available emission-line fluxes, and the average of eight pyqz estimations using different diagnostic pairs of SELs (see text). Objects shown in red have "well-behaved" posterior PDFs for the metallicity, meaning that they are not flagged by IZI as having either upper or lower limits in Z, nor multiple peaks in the PDF. Squares show objects with multiple peaks in the marginalized metallicity PDF, while circles show single-peak objects. Median error bars for the full sample are shown.

[10] pyqz was developed at ANU by Frederic Vogt and is publicly available at http://dx.doi.org/10.4225/13/516366F6F24ED.

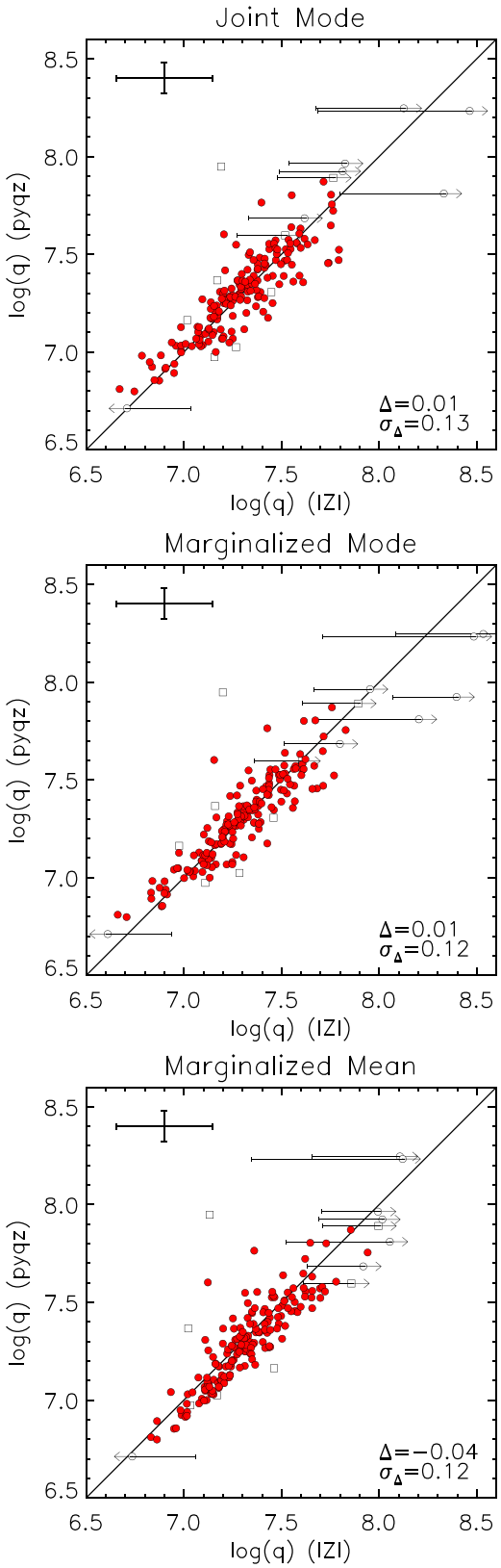

Figure 4. Same as Figure 3, but for the ionization parameter q.

asymmetries in the marginalized PDF that shift the mean away from the highest probability value.

The ionization parameter also shows good agreement between the two methods within the 1σ confidence limits, as can be seen in Figure 4. In this case we do not see significant differences between our three best-fit estimates of q. Overall, the consistency with the Dopita et al. (2013) methods as implemented in pyqz is excellent. This is reassuring and implies that our Bayesian inference method is able to recover the metallicity and ionization parameter of local H II regions as well as, if not better than, the state-of-the-art methods currently available in the literature, while also providing all the advantages listed in Section 2.

Hereafter we adopt the joint mode as the best estimate of the Z and q parameters. This is because it incorporates information regarding the covariance between the two parameters (unlike the marginalized mode) and because it is not dependent on how asymmetric the PDF is (unlike the marginalized mean).

5. EVALUATING THE PERFORMANCE OF INDIVIDUAL SEL DIAGNOSTICS

In Figures 1 and 2 we exemplified how the resulting PDFs depend on the information carried by different subsets of emission lines for a single H II region in the V98 catalog. Here we extend this analysis to the full V98 sample in order to statistically evaluate the performance of different SEL diagnostics that have been proposed in the literature and discuss the reasons behind the systematic differences observed. Having access to the actual shape of the metallicity and ionization parameter PDFs proves to be revealing in terms of understanding the sources of systematic biases associated with different diagnostics.

By adopting the best-fit abundances derived using the full set of emission lines available for the V98 sample as fiducial "true" values and comparing them to estimates of the abundance done using subsets of these lines (i.e., emulating different SEL diagnostics), we can study the origin of the scatter associated with different methods, as well as the causes behind any systematic deviations from the "true" values.

In this section we use `IZI` to evaluate the performance of the four diagnostics recently proposed by Dopita et al. (2013), as well as a set of six classic diagnostics commonly used in the literature: $R23$, $N2O2$, $N2$, $O3N2$, $O3O2$, and $R3$. These line ratios are commonly used to estimate the metallicity of H II regions and star-forming galaxies in the local universe, as well as at high redshift. All these diagnostics have been calibrated against local samples of H II regions, theoretical photoionization models, or in some cases combinations of the two (e.g., Maiolino et al. 2008). Here we are not interested in studying the differences between these calibrations, but rather we want to understand how sensitive different diagnostics are to the ionized gas metallicity, and what are the intrinsic biases and expected scatter associated with them. To this effect we do not use the original calibrations proposed in the literature and instead run `IZI` using only the subset of SELs associated with each diagnostic and the Dopita et al. (2013) grids as the input model. The comparison between these results and the fiducial metallicities computed using all available SELs allows us to evaluate the intrinsic performance of these diagnostics in terms of how much information they carry regarding the metal abundance independently of the way in which they are calibrated. In Section 6 we revisit the subject of abundance discrepancies caused by different types of calibrations.

5.1. Dopita et al. (2013) Diagnostics

Dopita et al. (2013) propose a set of new diagnostics consisting of four pairs of SEL ratios able to simultaneously constrain the abundance and ionization parameter of nebulae. These

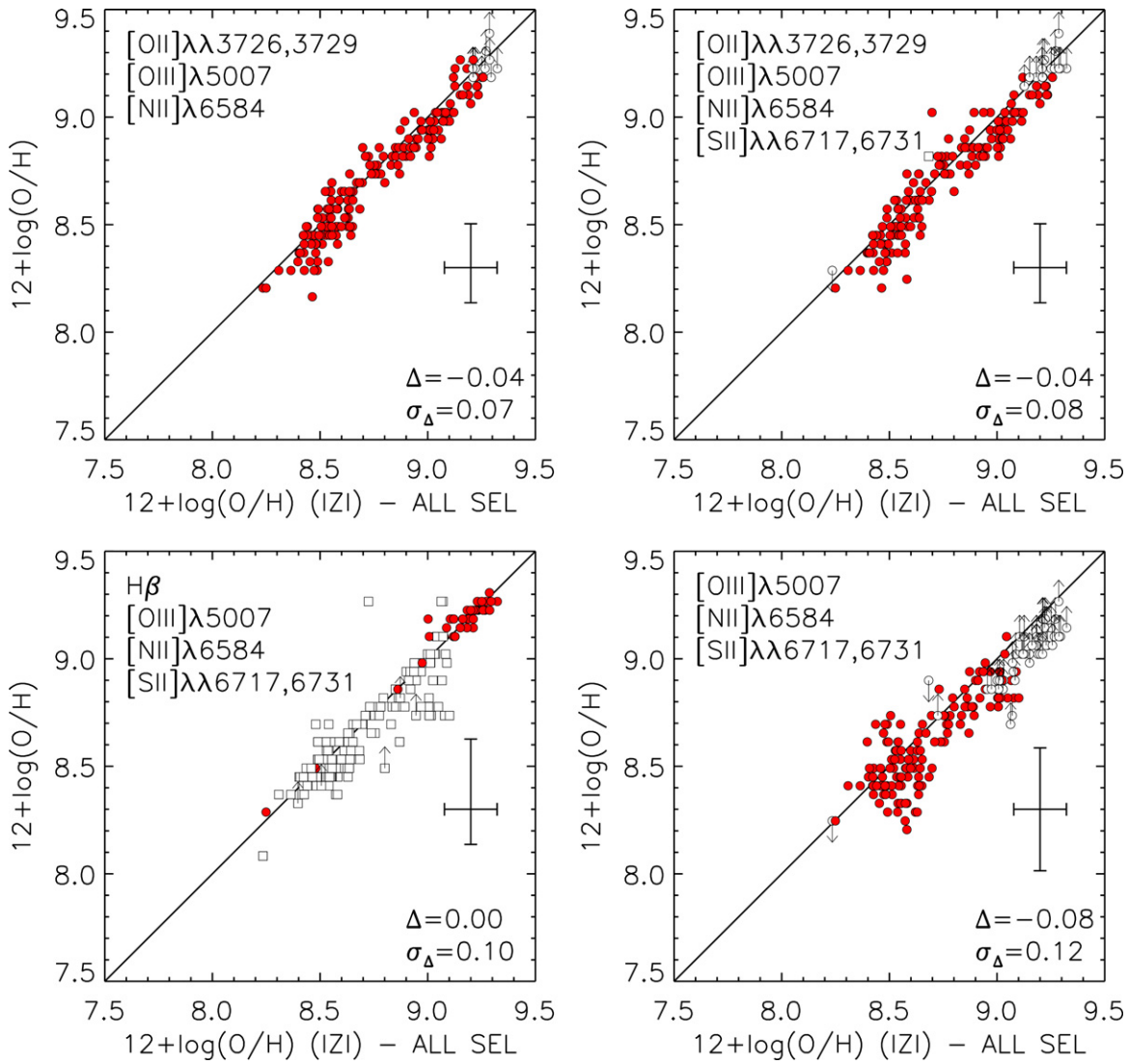

Figure 5. Comparison of IZI-derived abundances using all lines available in the V98 catalog (*x*-axis) vs. abundances calculated using subsets of these lines (*y*-axis). Each plot reports the subset of lines being used to compute the abundance in the *y*-axis, the median offset and the standard deviation of the datapoints, and the median error bars for all points. Circles show objects with single-peaked marginalized abundance PDFs, while squares show objects with multiple peaks in the PDF. Upper and lower limits are marked by arrows and indicate objects in which the PDF is not bounded on one side. Objects satisfying both criteria (i.e., single-peaked and bounded PDF) are shown in red.

correspond to the first four methods listed in Section 4.2. Figure 5 presents a comparison between the fiducial metallicities derived with IZI and the results of using only the lines associated with these four diagnostics. Below we comment on the performance of each:

[N II]/[O II] versus [O III]/[O II]. This diagnostic (top left panel of Figure 5) performs well over the whole range of sampled metallicities with a median offset of −0.04 dex and a scatter of 0.07 dex. At the high-abundance end ($12+\log \mathrm{O/H} > 9.2$) IZI only provides lower limits on the metallicity when using this restricted set of lines. This is because the posterior PDF becomes wider when only using a subset of the lines and its high-abundance end starts to overlap with the $Z_{\rm max}$ value (i.e., the highest abundance modeled in the input photoionization model grids). The limits are therefore associated with the limits of the input model grids and do not reflect any physical process taking place at these metallicities.

[N II]/[O II] versus [O III]/[S II]. This diagnostic (top right panel of Figure 5) also performs well with a median offset of −0.04 dex and a scatter of 0.08 dex. It shows the same issue regarding the lower limits in the metallicity at the high-abundance end as the previous diagnostic. Besides objects having lower limits in *Z*, one H II region (NGC 4395 +099−029) at $12+\log \mathrm{O/H} \simeq 8.7$ is flagged as having multiple peaks in the marginalized abundance PDF. Close inspection shows that this object presents a particularly elevated [S II] flux, which is in tension with the best-fit model preferred by the [O II], [O III], and [N II] lines. This gives rise to a secondary low-abundance (i.e., low [N II]/[S II] ratio) peak in the PDF, although IZI still chooses the correct higher probability peak as the best-fit solution for this object. Similarly, the object at the low-abundance end (IC2458 −028−007) showing an upper limit in *Z* also shows tension between its [S II] flux and the other emission lines. In this case the low- and high-abundance solutions in the PDF are merged into a single broad peak whose low-metallicity end hits the $Z_{\rm min}$ value modeled in the input grids, resulting in an upper limit that is still consistent with the "true" metallicity.

These conflicts raise an important point that we further discuss in Section 7. The results obtained when using this method will

only be reliable as long as the objects being studied are properly described by the input model grids. That is, if tension exists between the data and the models, then the best-fit values and particularly the confidence limits on them can be biased. This is of great importance to keep in mind when trying to measure the properties of ionized gas in high-redshift galaxies using photoionization model grids aimed at reproducing the observed spectrum of local galaxies and H II regions where the physical conditions in the gas and the shape of the ionizing spectrum are not ensured to be similar (e.g., Kewley et al. 2013b; Steidel et al. 2014).

[N II]/[S II] versus [O III]/Hβ. The bottom left panel of Figure 5 shows the comparison between this diagnostic and the fiducial IZI metallicities. The agreement is good overall with zero median offset and a scatter of 0.1 dex. When using this subset of lines most H II regions show a secondary low-probability peak in the joint PDF toward high metallicities and low ionization parameters. Sometimes the high-abundance end of this peak overlaps with the $Z_{\rm max}$ value, giving rise to the few reported lower limits. This secondary peak also introduces a high level of asymmetry in the 1σ confidence limits, as can be seen in the median error bars in Figure 5. Only in one case (NGC 2903 +171+236) is this high-abundance solution chosen instead of the "true" value, and this object stands out as an obvious outlier. Otherwise, the deviations observed are consistent with the 1σ uncertainty. The origin of the secondary high-abundance solutions is related to the absence of the [O II] doublet in this diagnostic. Leaving out this line translates into a significantly looser constraint on the ionization parameter, therefore opening a region of parameter space that would have been largely penalized had the [O II] doublet been included. Since the secondary peak typically has a much lower probability, we consider [N II]/[S II] versus [O III]/Hβ to be a good abundance diagnostic.

[N II]/[S II] versus [O III]/[S II]. Of the four diagnostics proposed by Dopita et al. (2013), this pair of line ratios (bottom right panel of Figure 5) show the highest systematic offset (−0.08 dex) with respect to the fiducial IZI metallicities, although this deviation is still within the 1σ scatter of 0.12 dex and well within the 1σ confidence limits in the metallicity PDF. At high abundances many objects have PDFs that overlap with the $Z_{\rm max}$ value in the models resulting in several upper limits. Similarly, two objects have PDFs that overlap with $Z_{\rm min}$ at the low-abundance limit. Overall, the behavior of this diagnostic is satisfactory.

5.2. Classic Literature Diagnostics

In this section we evaluate the performance of the six SEL diagnostics most widely used in the literature: $R23$, $N2O2$, $N2$, $O3N2$, $O3O2$, and $R3$. The top six panels of Figure 6 present these comparisons. We will find that the performance of some of these diagnostics is extremely poor unless prior information is given regarding the excitation and ionization state of the gas. Figure 7 presents a histogram of the ionization parameter as derived by IZI for the V98 sample using the full set of available SELs. Only objects for which $\log q$ is well constrained by the PDF (93% of the sample) are included. The distribution is well fit by a Gaussian distribution with mean $\mu = \langle \log q \rangle = 7.29$ and $\sigma = 0.23$ dex. This implies that 71% (93%) of the objects in this sample of local H II regions have ionization parameters that populate a fairly restricted 0.5 dex (1.0 dex) range around the mean value.

Many of the diagnostics discussed below have been calibrated against local samples of H II regions that, like the V98 sample, only span a very limited dynamic range in ionization and excitation. We are interested in studying how the performance of different methods depends on this fact, since this will have important consequences regarding the use of these diagnostics on objects that are not ensured to have the same ionization/excitation properties as the calibrators (e.g., very low metallicity and high redshift galaxies). To obtain insight on this problem, we repeat the comparisons described above using the best-fit Gaussian distribution shown in Figure 7 as a prior probability distribution on $\log q$ when computing the joint and marginalized posterior PDFs with IZI. The fiducial values are kept unmodified during this analysis (i.e., calculated without any priors on $\log q$). The result of this experiment is shown in the bottom six panels of Figure 6. We now proceed to discuss the performance of each diagnostic individually:

R23. The double-valued nature of the $R23$ diagnostic already discussed in Section 4.1 typically results in a PDF for the abundance with two probability peaks corresponding to the so-called lower branch and upper branch solutions. Most of the H II regions in the V98 sample are in the upper branch metallicity range of the $R23$ diagnostic. Figure 6 shows that when only the [O II], [O III], and Hβ fluxes are known and no prior information is available regarding the expected metallicity and ionization state of the gas in these objects, it is not possible to discriminate between the correct upper branch solution and the incorrect lower branch solution for about half of the objects at $12+\log {\rm O/H} > 8.7$. Below this value the $R23$ dependence with metallicity flattens out (see Figure 2), the two probability peaks in the PDF merge into a single broad peak, and the scatter increases substantially ($\simeq$0.2 dex at $8.3 < 12+\log {\rm O/H} < 8.7$). A few objects at the low-abundance end of the sample suffer from the opposite problem with the upper branch solution being chosen as the best-fit value instead of the correct lower branch solution. Furthermore, even when the two peaks are merged together, sometimes the PDF is skewed in a way that the mode is shifted toward higher values, translating into a 0.07 dex offset from the fiducial values in the $8.3 < 12+\log {\rm O/H} < 8.7$ range. We do not attempt to calculate an offset and a scatter at $12+\log {\rm O/H} > 8.7$.

Introducing a Gaussian prior on $\log(q)$ somewhat alleviates the above problems, as can be seen in Figure 6. Typically when inspecting the joint posterior PDF, one sees that the upper branch solution lies at $\log q$ values that are higher than those corresponding to the lower branch solution. At the high-abundance end, these values are closer to the mean of the $\log q$ distribution presented in Figure 7. The second row of panels in Figure 1 provides a good example of this behavior. Therefore, using the Gaussian prior on $\log q$ tends to decrease the probability of the lower branch solution for objects in the high-metallicity end of the sample, bringing the $R23$ abundance of many objects into agreement with the "true" abundance. The opposite effect happens at the low-abundance end, where the ionization parameters associated with the lower branch solution are typically closer to the mean of the $\log q$ distribution, so objects for which the $R23$ diagnostic was choosing the wrong upper branch solution when no prior on $\log q$ was used also come into better agreement with the fiducial values. This lowers the systematic offset in the $8.3 < 12+\log {\rm O/H} < 8.7$ range from 0.07 dex to −0.03 dex but is accompanied by an increase in the scatter in this region from 0.18 dex to 0.24 dex. Regardless

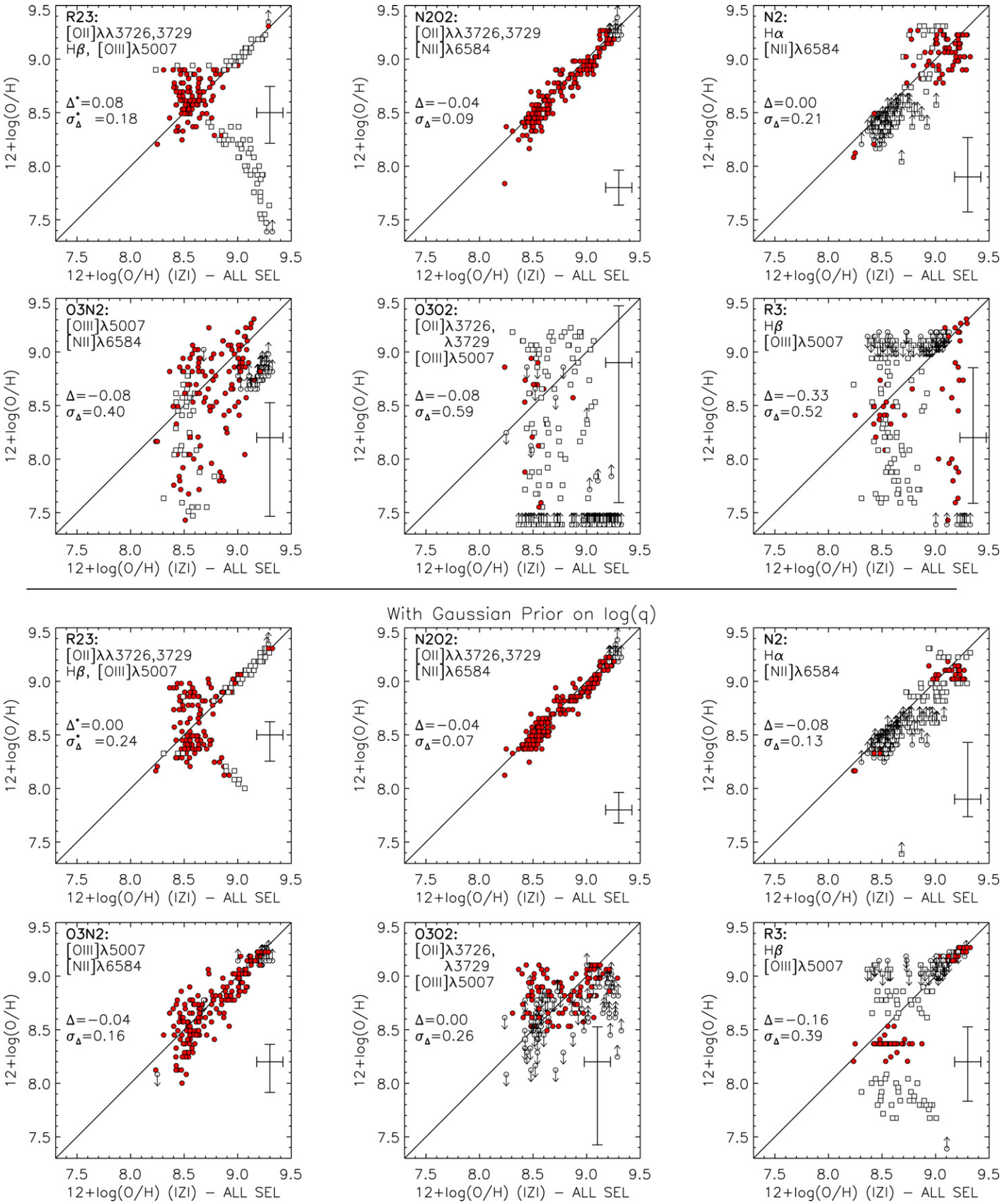

Figure 6. Top six panels: same as in Figure 5, but for subsets of lines emulating the $R23$, $N2O2$, $N2$, $O3N2$, $O3O2$, and $R3$ diagnostics. Maximum ignorance priors on Z and q are assumed in these calculations. Bottom six panels: same as above, but this time using the Gaussian prior on the ionization parameter shown in Figure 7. For $R23$ the median offset and scatter are calculated only in the $8.3 < 12+\log \mathrm{O/H} < 8.7$ range.

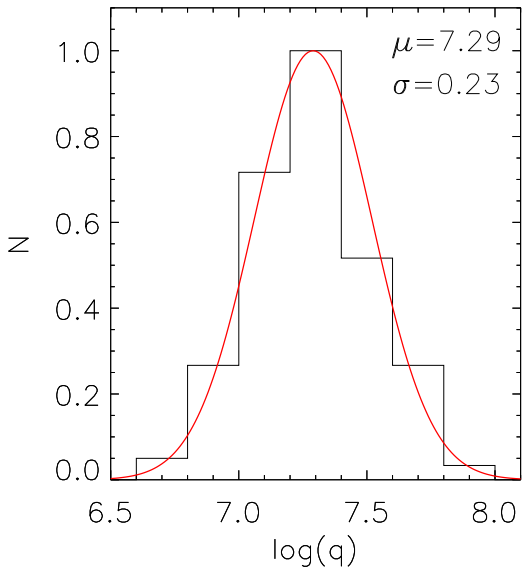

Figure 7. Black histogram presents the normalized distribution of best-fit ionization parameters for the V98 sample using all available SELs. Only sources where the marginalized PDF is single peaked and bounded on both sides are included (93% of the sample). The best-fit Gaussian PDF adopted as a prior on $\log(q)$ in the bottom six panels of Figure 6 and the bottom three panels of Figure 8 is shown in red. Also reported are the best-fit parameters of this Gaussian PDF.

of these improvements, the wrong solution is still preferred for a significant fraction of the objects.

In line with what is common usage in the literature (e.g., Pilyugin & Thuan 2005; Kewley & Ellison 2008; Moustakas et al. 2010), we conclude that $R23$ can only be used as a metallicity diagnostic if prior information regarding the expected metallicity range (i.e., the branch) is available. In order to provide a fair evaluation of the performance of this diagnostic, we include an extra prior on the metallicity that assigns all objects in the V98 sample to the upper branch. This prior is a step function with uniform probability per decade in Z above $12+\log \mathrm{O/H} = 8.3$ and zero probability below this value. The results of using this branch prior and of using both the branch and Gaussian ionization parameter priors together are shown in the top left and bottom left panels of Figure 8, respectively. After including prior information on the expected metallicity range, the performance of the $R23$ diagnostic is excellent at the high-abundance end ($12+\log \mathrm{O/H} > 8.8$) but degrades dramatically below this value, where the ratio becomes highly insensitive to the metallicity. The scatter in the $8.3 < 12+\log \mathrm{O/H} < 8.7$ range remains high at 0.24 dex, and we measure a global scatter of 0.2 dex over the full metallicity range of the V98 sample. Introduction of the prior information on the branch removes any systematic offset with respect to the fiducial abundances.

N2O2. As can be seen in Figure 6, this diagnostic performs extremely well over the full metallicity range analyzed here. It shows a median offset of −0.04 dex and a scatter of 0.09 dex with respect to the fiducial values. For a few objects at the high-abundance end the PDF overlaps with the $Z_{\max}$ limit of the input grids, resulting in a few lower limits. In Section 4.1 we discussed how part of the power of $N2O2$ as a metallicity diagnostic comes from the fact that this line ratio is very insensitive to the ionization parameter. This is evident from the fact that the results remain almost unchanged when using a Gaussian prior on the $\log(q)$ distribution.

At first sight our results seem to imply that reliable abundances can be obtained from this line ratio without any prior knowledge of the expected metallicity range or the ionization and excitation conditions in the gas. This statement is subject to a very important caveat already mentioned in Section 4.1. As shown in Pérez-Montero (2014), the $N2O2$ diagnostic traces the N/O abundance almost linearly. The photoionization models used here assume a monotonic single-valued function between N/O and O/H with no scatter, and the ability of $N2O2$ to trace the oxygen abundance relies heavily on this assumption. In reality, the observed correlation between these two quantities shows a scatter of 0.15–0.25 dex, and the N/O ratio appears to increase systematically as a function of galaxy total stellar mass (e.g., van Zee et al. 1998a; Andrews & Martini 2013; Pérez-Montero 2014, Belfiore et al. 2014). In the absence of systematic trends like the latter one, we could simply incorporate the scatter in the N/O versus O/H relation to the error bars on the abundance calculated by `IZI`. Since both are of a similar magnitude, we do not expect the issue of the scatter to be of great significance. The most worrisome problem is the presence of systematic trends in the N/O-versus-O/H relation as a function of global and local galactic properties and the increased scatter in N/O toward the low-abundance regime. The proper way to deal with these systematics is to leave N/O as a free parameter in the fits, as done by Pérez-Montero (2014). In the future we plan to extend the capabilities of `IZI` to deal with extra parameters beyond Z and q in order to deal with these types of problems.

N2. Figure 6 shows a large number of objects with lower limits and double-peaked PDFs in metallicity when using the $N2$ diagnostic. The reason behind this behavior is the double-valued nature of this diagnostic (see Figure 2 and the bottom panels of Figure 1), which is rarely appreciated in the literature (although see Kewley & Dopita 2002) because the flattening and inversion of the metallicity dependence of $N2$ take place at relatively high abundances ($12+\log \mathrm{O/H} \simeq 9.2$). In the sample used by Pettini & Pagel (2004) to calibrate this method there are only a couple of objects at $12+\log \mathrm{O/H} \simeq 9.2$, so the "upper branch" of $N2$ has not been exposed observationally. A closer inspection of Figure 1 in Pettini & Pagel (2004) actually shows evidence for this plateau in $N2$ at the high-abundance end. The Dopita et al. (2013) photoionization models sample the oxygen abundance past the region where this inversion happens (upper right panel of our Figure 2).

With this in mind we can better understand the results presented in Figure 6. For objects showing single-valued PDFs (red circles) the entirety of the high-abundance peak (i.e., the upper branch solution) is past the $Z_{\max}$ value in the models. Most objects at lower abundances show two peaks in the PDF corresponding to the lower and upper branch solutions for $N2$. For the V98 sample the low-abundance peak is typically the "correct" solution, but without any prior information available on either the abundance or the ionization parameter, many times the wrong high-abundance solution is preferred. This happens most often in the $8.7 < 12+\log \mathrm{O/H} < 9.1$ range. Toward lower abundances upper limits become common as the two peaks become more widely separated and the upper branch solution overlaps with the $Z_{\max}$ value. Overall, ignoring lower limits this diagnostic shows zero systematic deviation from the fiducial values and a large scatter of 0.21 dex.

Including a Gaussian prior on the ionization parameter censors regions of parameter space where the ionization parameter is high, and this reduces the number of objects for which

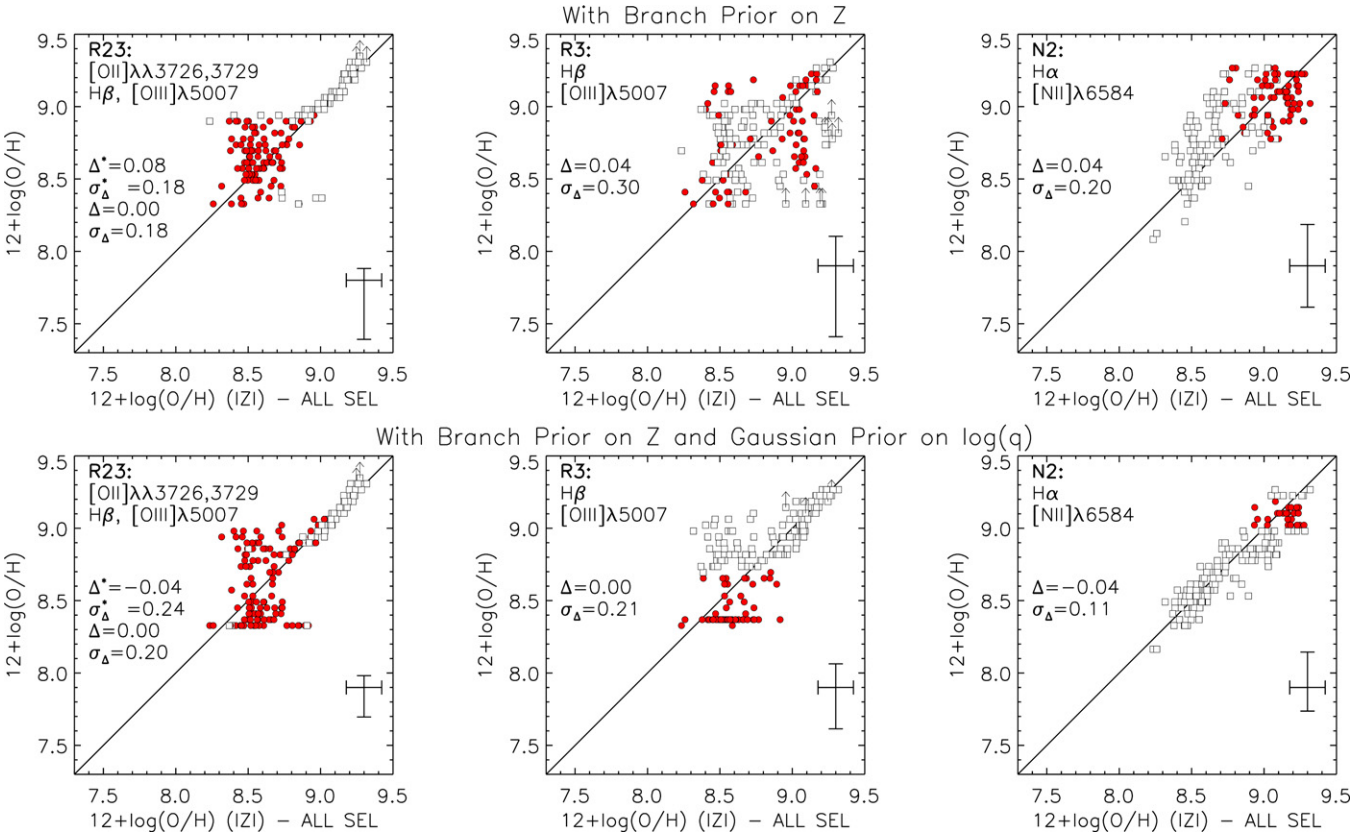

Figure 8. Same as in Figure 6 for $R23$, $R3$, and $N2$, but now using a branch prior in the top three panels and both a branch prior and a Gaussian ionization parameter prior in the bottom three panels. For $R23$ we now report the offset and scatter in both the $8.3 < 12+\log \mathrm{O/H} < 8.7$ range (Δ^*, σ_Δ^*) and the full range of the data (Δ, σ_Δ).

the high-abundance solution is preferred. This is because the two parameters are correlated with high-abundance solutions typically showing higher values of q. The scatter in this case is significantly reduced to 0.13 dex, but the median offset increases to −0.08 dex. In any case, the number of objects for which we can only provide lower limits in the metallicity remains high.

The double-valued nature of $N2$ implies the need to choose a branch, just as is typically done when using the $R23$ diagnostic (see discussion above). In the case of the V98 sample all objects populate the lower branch. In the top right panel of Figure 8 we show the effects of imposing a prior on the metallicity in order to emulate this preference for the lower branch solutions. We use a flat prior with equal probability per decade in Z at $12+\log \mathrm{O/H} < 9.3$ and zero probability above this value. This decreases the median offset, but the scatter remains large (0.2 dex) because of the covariance between the abundance and the ionization parameter. We find that only after introducing prior information on both the metallicity range (i.e., the branch) and the ionization parameter of the gas does $N2$ become a well-behaved abundance diagnostic with a median offset of −0.04 dex and scatter of 0.11 dex. Finally, just as $N2O2$, $N2$ is also affected by the assumed behavior of the N abundance discussed above (Pérez-Montero 2014).

O3N2. This diagnostic, first proposed by Alloin et al. (1979) and later recalibrated by Pettini & Pagel (2004), shows a very poor performance with a median offset of −0.08 dex with respect to the fiducial metallicities and a large scatter of 0.4 dex. This is not surprising after inspection of Figure 2, which shows that $O3N2$ is as sensitive to the ionization parameter as it is to the metallicity.

A scatter of 0.4 dex is significantly higher than the 0.25 dex scatter reported by Pettini & Pagel (2004) in their calibration. The cause behind this discrepancy is the limited dynamic range in ionization parameter spanned by the calibrators. To show this, Figure 6 also presents the results of using the Gaussian prior in $\log q$ discussed above in our calculations. This dramatically improves the performance of the $O3N2$ diagnostic by lowering the median offset to −0.04 dex and the scatter to 0.16 dex, in much better agreement with Pettini & Pagel (2004). This implies that $O3N2$ can only be used as a reliable abundance estimator if prior information regarding the ionization and excitation state of the gas is available. This diagnostic is also strongly affected by the N abundance (Pérez-Montero 2014).

O3O2 and R3. Figure 6 shows that these two ratios perform very poorly as abundance diagnostics with median offsets of −0.08 dex and −0.33 dex and scatters of 0.59 dex and 0.52 dex (i.e., factors of three) for $O3O2$ and $R3$, respectively. This is not surprising in light of the strong dependence on the ionization parameter seen in Figure 2 for these diagnostics. In fact, both line ratios have been typically used as ionization parameter diagnostics in the literature (e.g., McGaugh 1991; Kewley & Dopita 2002; Lilly et al. 2003; Nakajima et al. 2013). Maiolino et al. (2008) combined a sample of low-metallicity galaxies with direct T_e abundances with a sample of high-metallicity star-forming galaxies from SDSS-DR4 with metallicities measured using the Kewley & Dopita (2002) photoionization model prescriptions, to calibrate a series of SEL diagnostics over a

broad dynamic range in metallicity. In particular, they find relatively tight correlations with metallicity for $O3O2$ and $R3$ with typical scatter of ∼0.1–0.3 dex, so the authors adopt them as part of their suite of abundance diagnostics.

The apparent contradiction between the observed correlation of these line ratios with metallicity in Maiolino et al. (2008) and the poor performance seen in Figure 6 can again be explained by the limited dynamic range in ionization and excitation conditions in the calibration sample. Similar to what we observed for $O3N2$, the performance of these two diagnostics improves significantly when using our Gaussian prior on $\log(q)$. The bottom panels of Figure 6 show that this reduces the median offsets to zero and −0.16 dex and the scatter to 0.26 and 0.39 dex for $O3O2$ and $R3$, respectively, closer to the scatters seen in the Maiolino et al. (2008) calibrations. Even after these improvements, the performance of $O3O2$ and $R3$ as abundance diagnostics is still poor.

In the case of $R3$, Figure 2 shows that it follows a double-valued behavior similar to that of $R23$ (although with a much stronger dependence on the ionization parameter), so proper usage of this diagnostic also requires choosing between a lower and upper branch. The middle panels of Figure 8 present the results of using a branch prior on the metallicity with uniform probability per decade in Z above $12+\log \mathrm{O/H} = 8.3$ and zero probability below this value, as well as a combination of a branch prior and a Gaussian prior on $\log(q)$. The branch prior marginally improves the performance of this diagnostic, and it is only after including priors on both q and Z that its performance becomes acceptable at the high-abundance end but degrades rapidly at $12+\log \mathrm{O/H} < 8.9$ as the ratio becomes insensitive to the metallicity.

Overall, we do not consider $O3O2$ to be a reliable abundance indicator, and we recommend against using $R3$ without previous knowledge of the ionization conditions and expected metallicity range of the objects under study.

5.2.1. Classic SEL Diagnostics Performance Summary

In summary, out of the six SEL diagnostics examined above, only $N2O2$ seems to provide reasonably good results without requiring prior knowledge of the ionization and excitation conditions in the gas and/or the expected range in metallicity of the objects. This diagnostic also presents the lowest scatter among the six methods, but it is subject to systematic uncertainties associated with the assumed behavior of the N/O abundance as a function of metallicity. Furthermore, these two lines are well separated in wavelength, and using them requires an accurate correction for dust attenuation and enough spectral resolution to separate [N II] λ6583 from Hα.

Our analysis shows that reliable results, although with an increased level of scatter, can also be obtained from $N2$ and $O3N2$ if prior information is available regarding the ionization conditions in the gas. Furthermore, the performance of the $N2$ diagnostic can become comparable to that of $N2O2$ if, on top of having prior information on q, we know in advance in which branch objects will fall. This is typically the case with most samples of galaxies and H II regions since the separation between the lower and upper branches of $N2$ happens at a very high abundance. The separation between the lower and upper branches of the $R23$ diagnostic, on the other hand, occurs at a much lower abundance, so $R23$ can only be used reliably if the branch is known a priori, and even in that case the scatter becomes uncomfortably large at intermediate values where the ratios are insensitive to the metallicity (see also López-Sánchez et al. 2012).

Finally, we find that the $O3O2$ and $R3$ ratios are always fairly poor abundance diagnostics. The results go from catastrophic to poor when prior information on the ionization parameter is included. For $R3$ the results become marginally acceptable if further information regarding the branch in which the objects fall is available.

The analysis presented above is conducted in the context of the photoionization models adopted as input for IZI. While these models surely do not provide perfect representations of real star-forming galaxies and H II regions, general trends of relative SEL intensity as a function of metallicity and ionization parameter like those presented in Figure 2 are thought to be robust, at least in a qualitative sense. Therefore, while the quantitative details of the analysis might change as a function of the input model adopted, the conclusions presented in this section should not be significantly affected, at least in a qualitative way.

We emphasize here that the above analysis is aimed at evaluating the intrinsic performance of these diagnostics in terms of how much information useful to constrain the metallicity is encoded in these line ratios. So far we have not evaluated any systematics associated with the calibration of these diagnostics against either direct method abundances or photoionization models. This is the subject of the next section.

6. COMPARING PHOTOIONIZATION MODEL, DIRECT METHOD, AND RECOMBINATION LINE ABUNDANCES

In this section we explore the discrepancies in the abundance scale known to exist between the different families of methods used to measure chemical abundances in ionized gas. In particular, we compare oxygen abundances measured with IZI using different photoionization models in the literature with those obtained by applying the direct method, a set of empirically calibrated SEL methods (i.e., calibrated against direct method abundances), and abundances derived from the direct observation of oxygen RLs.

Unlike CELs and temperature-sensitive auroral lines, the emissivity of RLs has a very mild dependence on T_e and therefore on the presence of temperature fluctuations within nebulae. This method does not suffer from the systematic uncertainties that affect the direct and SEL methods, although it is subject to its own set of systematic uncertainties chiefly having to do with the accurate calculation of recombination coefficients to high quantum level states (Liu et al. 2000; Stasińska 2004). Furthermore, RL abundances have been shown to agree very well with the stellar abundances of OB stars in star-forming regions (e.g., Peimbert et al. 2005; Simón-Díaz & Stasińska 2011). The above makes them a good reference point to evaluate the performance of other methods.

Comparisons of this type have been performed before by García-Rojas & Esteban (2007) and López-Sánchez et al. (2012) for a handful of local H II regions with RL measurements. These studies find that in H II regions the direct method consistently yields abundances that are ∼0.2 dex lower than the RL method. García-Rojas & Esteban (2007) conclude that this offset is best explained by the presence of temperature fluctuations inside H II regions, which tend to bias the direct T_e measurements toward higher values and the abundances toward lower values, although other authors disagree (c.f. Simón-Díaz & Stasińska 2011). López-Sánchez et al. (2012) compare the average abundances from several empirically

calibrated and theoretically calibrated SEL diagnostics to RL abundances and conclude that SEL methods calibrated against photoionization models yield abundances that are consistently ∼0.3 dex higher than the direct method and ∼0.1 dex higher than the RL method. This is a similar difference to the one seen between empirical and theoretical SEL diagnostics by Kewley & Ellison (2008) using a large sample of emission-line galaxies from SDSS.

IZI allows us to calculate the oxygen abundance of H II regions using an arbitrary photoionization model and without having to restrict the analysis to the choice of a particular set of SEL diagnostics. This offers the opportunity to extend the above comparisons to test how the results from different photoionization models compare to direct and RL abundances, independently from the systematics associated with a particular calibration of a particular diagnostic.

As mentioned in Section 1, the RLs of elements heavier than He are typically 10^3–10^4 times fainter than SELs, so measuring them in local H II regions typically requires several hours of exposure time using high-resolution spectrographs in 10 m class telescopes. Because of this, RL abundances have only been measured for He, C, O, and Ne in a few dozen bright H II regions in the MW and the Local Group (e.g., Peimbert 2003; Peimbert et al. 2005; Tsamis et al. 2003; Esteban et al. 2004, 2009; García-Rojas & Esteban 2007; López-Sánchez et al. 2007; Bresolin et al. 2009). For this comparison we use the compilation of 22 local H II regions with measured RL oxygen abundances presented in Table 6[11] of López-Sánchez et al. (2012). Although hundreds of emission lines are typically detected in the high-quality spectra used to measure RL abundances in these regions, we run IZI using only the [O II] λλ3726, 3729, Hβ, [O III] λ5007, Hα, [N II] λ6583, and [S II] λλ6717, 6731 line fluxes in order to emulate the information that would typically be available from a low-S/N, low-resolution spectrum.

Figure 9 presents the results of the comparison. The first row of panels compares the RL abundances of the 22 regions against the direct method abundances (left) computed using the same high-S/N, high-resolution spectra used to derive the RL abundances, and against the average of three empirically calibrated SEL methods from Pilyugin (2001), Pilyugin & Thuan (2005), and Pilyugin et al. (2010) (dubbed the *P*-method). The data presented in these two panels are taken from López-Sánchez et al. (2012). As already stated in that work, both direct method abundances and empirically calibrated SEL abundances consistently underpredict the RL abundances by 0.24 dex and 0.23 dex, respectively. After removing the offset direct method abundances show a scatter of 0.06 dex with respect to RL abundances. This is consistent with the errors in the direct and RL abundance determinations, which have a median value of 0.04 dex and 0.06 dex, respectively. On the other hand, the empirically calibrated SEL abundances show a scatter of 0.12 dex against the RL abundances. This is significantly larger than the reported scatter of 0.075 dex for the Pilyugin et al. (2010) calibrations.

The rest of the panels in Figure 9 compare abundances derived with IZI using different photoionization models against RL abundances. We present results for a subset of the photoionization models presented in Kewley et al. (2001), Levesque et al. (2010), Richardson et al. (2013), and Dopita et al. (2013). The

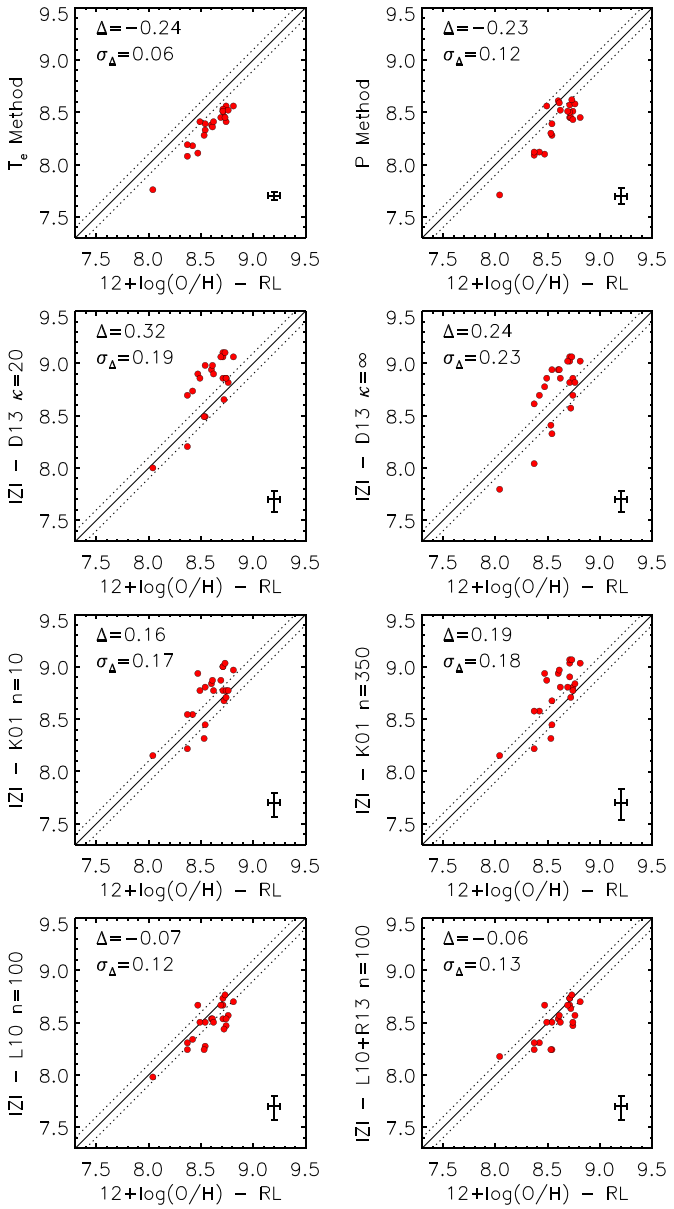

Figure 9. Oxygen RL abundances vs. those derived using different methods. RL abundances are plotted in the *x*-axis and, together with T_e method and *P* method abundances (first row), are taken from Table 6 of López-Sánchez et al. (2012). The *y*-axis in the rest of the panels shows abundances computed with IZI using the [O II] λλ3726, 3729, Hβ, [O III] λ5007, Hα, [N II] λ6583, and [S II] λλ6717, 6731 line fluxes from the original references in Table 6 of López-Sánchez et al. (2012) and different photoionization model grids. The second row corresponds to the Dopita et al. (2013) models for $\kappa = 20$ and $\kappa = \infty$ (i.e., Maxwell–Boltzmann T_e distribution). The third row corresponds to the Kewley et al. (2001) models with $n = 10$ and $n = 350$. The bottom row shows results for the Levesque et al. (2010) models on the left and the combined Levesque et al. (2010) and Richardson et al. (2013) models with extended sampling toward high values of q on the right. Dotted lines at ±0.1 dex offsets are shown for reference.

models have been described in Section 3.2. The Dopita et al. (2013) models systematically overpredict the RL abundances by 0.32 dex and 0.24 dex for electron energy kappa probability distributions with $\kappa = 20$ and $\kappa = \infty$ (i.e., a M-B distribution), respectively. The scatter seen for these two models is 0.19 dex and 0.23 dex, respectively. Similarly to the case of the *P*-method above, the observed scatter is significantly larger than what is expected from the median of the uncertainty calculated by IZI,

[11] Because of an error during the creation of this table, the line ratios reported in López-Sánchez et al. (2012) are incorrect (Angel R. Lopez-Sanchez 2014, private communication). We used the line fluxes and errors as reported in the original reference for each object.

which corresponds to 0.1 dex.[12] Although the dynamic range in abundance of the H I region sample with RL line measurements is limited and the number of regions is small, there appears to be a trend of an increasing overestimation toward higher abundances.

Results for the Kewley et al. (2001) models are presented in the third row of Figure 9 assuming two different values for the electron density ($n_e = 10$ cm^{-3} and $n_e = 350$ cm^{-3}). Abundances derived using these models are higher than RL abundances with offsets of 0.16 dex and 0.19 dex and scatters of 0.17 dex and 0.18 dex for the $n_e = 10\,\mathrm{cm}^{-3}$ and $n_e = 350\,\mathrm{cm}^{-3}$ models, respectively. No significant difference is seen between the two electron density regimes considered. The observed scatter is slightly larger than the 0.13 median uncertainty reported by IZI. This offset is similar to the offset measured by López-Sánchez et al. (2012) when comparing the average of a few theoretically calibrated SEL diagnostics based on the Kewley et al. (2001) models with RL measurements. The observed offset is similar in magnitude to the offset measured for direct-abundance-based methods but in the opposite direction. A similar trend to that observed for the Dopita et al. (2013) models is seen in which the difference in the values becomes larger toward the high-abundance end.

Results from the Levesque et al. (2010) models both with and without including their extension toward high values of q by Richardson et al. (2013) are presented in the two bottom panels of Figure 9. The inclusion of the Richardson et al. (2013) models leaves results practically unchanged since none of the H II regions in this sample present extremely large ionization parameters. Abundances derived using the Levesque et al. (2010) models present the best agreement with RL abundances among all the photoionization models considered here. This model shows a systematic offset of −0.07 dex and a scatter of 0.12, which is in excellent agreement with the median error of 0.11 dex reported by IZI. This is the best performance in terms of both the median offsets and the magnitude of intrinsic scatter beyond the measurement errors among all the SEL methods studied here.

This comparison highlights an interesting fact. Abundances derived using IZI in combination with different photoionization models show different offsets with respect to abundances measured using the direct and RL methods. A large range in median offsets is seen among different photoionization models, although this is driven mainly by the Dopita et al. (2013) models, which predict abundances typically ∼0.1 dex higher than the Kewley et al. (2001) models and ∼0.3 dex higher than the Levesque et al. (2010) models. The MAPPINGS photoionization code used to compute all these models suffered a major revision between the latter two works and the former. These upgrades are thoroughly discussed in Dopita et al. (2013) and include, among other things, the inclusion of a kappa function to describe the electron energy distribution instead of the classically used M-B distribution (Nicholls et al. 2012). Our results suggest that this is not the main cause behind the observed discrepancies, as adopting a M-B distribution ($\kappa = \infty$) only decreases the offset by less than 0.1 dex. Other factors such as the assumed nebular geometry, the updated atomic data and abundance set, and the different parameterization of the N/O-versus-O/H relation must contribute to the observed offsets. This also implies that our results do not contradict the fact that a κ electron energy distribution can help reconcile direct and RL abundances as proposed by Nicholls et al. (2013), as the systematic offsets observed in this work are most likely not associated with the adoption of the κ distribution. In the future we expect to explore in more detail the effects on the derived abundances caused by different assumptions in photoionization models.

From this analysis we can conclude that Bayesian inference using photoionization models can be successful at reproducing the results of the RL method. We do not find a significant discrepancy between RL abundances and the abundances computed using IZI with the Levesque et al. (2010) models. The large offset seen against the Dopita et al. (2013) models highlights the presence of poorly understood systematic uncertainties affecting photoionization models.

7. SUMMARY AND CONCLUSIONS

We have presented a new method to measure the metal abundance and ionization parameter of H II regions and star-forming galaxies using SELs. The method is based on the application of Bayesian inference to calculate the joint and marginalized PDFs for these two parameters given an arbitrary set of emission-line flux measurements or upper limits and a model for how the brightness of these lines depends on Z and q. We also present an implementation of the method called IZI that computes these PDFs using theoretical photoionization models, and we make the code public.

Using IZI, we have tested the performance of a series of popular SEL diagnostics in the literature to evaluate how much information regarding the metal abundance is carried by these line ratios. For this we have used a sample of 186 extragalactic H II regions from V98 and have run IZI using subsets of emission lines that emulate the information contained in a particular diagnostic. We have evaluated the $R23$, $N2O2$, $N2$, $O3N2$, $O3O2$, and $R3$ line ratios, as well as the four pairs of line ratio diagnostics proposed by Dopita et al. (2013).

We have also used a sample of bright local H II regions with direct method and RL abundance measurements to study the discrepancies in the abundance scale between these two methods and those based on the use of theoretical photoionization models. For this comparison we include results based on using IZI in combination with several photoionization models in the literature, including those presented in Kewley et al. (2001), Levesque et al. (2010), Richardson et al. (2013), and Dopita et al. (2013).

From the experiments presented above we conclude the following:

1. Bayesian inference provides an optimal tool to measure the physical conditions of ionized gas in nebulae. Our method has many advantages over the classic approach of calibrating a single line ratio or a pair of line ratios as an abundance diagnostic. These advantages are listed in Section 2. Here we highlight the fact that IZI uses all the available spectroscopic information simultaneously, circumvents the need to calibrate a particular diagnostic against a particular set of models, and allows one to naturally identify multiple peaks, asymmetries, and upper or lower limits in the PDF, which translates into realistic error estimates.
2. When evaluating the performance of different SEL abundance diagnostics in the literature, we find a large diversity in terms of how accurately the diagnostics trace the

[12] This median uncertainty corresponds to the median between the upper and lower error bars for both versions of the models presented.

abundance of H II regions. Some diagnostics only perform well when prior information is available regarding the abundance range (i.e., the branch) in which the sources are expected to fall, or the ionization and excitation conditions in the gas. The $N2O2$ diagnostic seems to be robust against the lack of prior information, but its dependence on the N/O ratio makes it subject to significant systematic uncertainties. In light of their poor performance, we recommend strongly against the use of $R3$ and $O3O2$ as abundance diagnostics. Other diagnostics such as $R23$ and $N2$ can only perform well if prior information regarding the correct branch is available. Finally, $O3N2$ only performs reasonably well if prior information regarding q is available. The four pairs of SEL ratios proposed by Dopita et al. (2013) carry enough information to constrain Z and q simultaneously in most cases.

3. The different limitations of individual diagnostics highlight the advantage of using all spectroscopic information available from transitions of different elements and different ions when computing abundances, since this minimizes systematics and helps constrain both the abundance and the ionization parameter in a much better way than individual diagnostics.
4. Abundances derived through Bayesian inference using photoionization models can agree very well with RL abundances, but the agreement depends on which models are used. Out of all the models tested here, the Levesque et al. (2010) grids produce abundances that agree with RL measurements remarkably well (to within 30%). The models of Kewley et al. (2001) show a 0.15–0.2 dex systematic offset toward higher abundances, but this offset is still within the observed scatter. The Dopita et al. (2013) models significantly overpredict RL abundances by 0.2–0.3 dex for this small sample of local H II regions, although the discrepancy is not associated with the use of a κ electron energy distribution. The scatter seen between different photoionization models highlights the systematic uncertainties that affect them, including the assumptions regarding the input ionizing flux from synthetic stellar population models, the assumed behavior of relative elemental abundances like N/O versus O/H, and the assumed physical structure of the nebulae.
5. Direct method abundances are found to be systematically lower than both RL and photoionization model abundances. This is in agreement with previous claims in the literature and is most likely caused by the unaccounted effect of temperature fluctuations within ionized nebulae.

Finally, we would like to stress two important points regarding the use of `IZI`. The first is the fact that the results produced by `IZI` will only be reliable if the input model being used corresponds to a proper characterization of the sources being analyzed. If the shape of the input ionizing spectrum and the assumed pattern of relative elemental abundances in the models are significantly different from what they really are in the objects under study, then even if a correct statistical analysis is used to compare the data to these models, results will be subject to systematic uncertainties. Second, as briefly mentioned in Section 3.2, this method is flexible and not limited to being used with theoretical photoionization models. In principle, the user can adopt any model that traces the behavior of the lines of interest as a function of Z and q, even one based on empirical grids of direct method abundances for real objects. In the future we expect to explore these possibilities and also to expand `IZI` to deal with extra parameters such as the N/O abundance, the electron density, and the age or effective temperature of the ionizing stellar population.

We would like to thank Gwen Rudie, Andrew Benson, Daniel Masters, and Juna Kollmeier for useful discussions regarding this paper. This research made use of NASA's Astrophysics Data System; the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration; and the VizieR catalogue access tool, CDS, Strasbourg, France. The original description of the VizieR service was published in A&AS, 143, 23.

REFERENCES

Aller, L. H. 1954, ApJ, 120, 401
Alloin, D., Collin-Souffrin, S., Joly, M., & Vigroux, L. 1979, A&A, 78, 200
Anders, E., & Grevesse, N. 1989, GeCoA, 53, 197
Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Belfiore, F., Maiolino, R., Bundy, K., et al. 2014, arXiv:1410.7781
Bresolin, F., Gieren, W., Kudritzki, R.-P., et al. 2009, ApJ, 700, 309
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Brinchmann, J., Pettini, M., & Charlot, S. 2008, MNRAS, 385, 769
Brooks, A. M., Governato, F., Booth, C. M., et al. 2007, ApJL, 655, L17
Cullen, F., Cirasuolo, M., McLure, R. J., & Dunlop, J. S. 2014, MNRAS, 440, 2300
Dalcanton, J. J. 2007, ApJ, 658, 941
Dalcanton, J. J., Yoachim, P., & Bernstein, R. A. 2004, ApJ, 608, 189
Denicoló, G., Terlevich, R., & Terlevich, E. 2002, MNRAS, 330, 69
Díaz, A. I., & Pérez-Montero, E. 2000, MNRAS, 312, 130
Dopita, M. A., & Evans, I. N. 1986, ApJ, 307, 431
Dopita, M. A., Kewley, L. J., Heisler, C. A., & Sutherland, R. S. 2000, ApJ, 542, 224
Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, ApJS, 208, 10
Dufour, R. J., Talbot, R. J., Jr., Jensen, E. B., & Shields, G. A. 1980, ApJ, 236, 119
Edmunds, M. G., & Pagel, B. E. J. 1984, MNRAS, 211, 507
Eldridge, J. J., & Stanway, E. R. 2009, MNRAS, 400, 1019
Esteban, C., Bresolin, F., Peimbert, M., et al. 2009, ApJ, 700, 654
Esteban, C., Peimbert, M., García-Rojas, J., et al. 2004, MNRAS, 355, 229
Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181
Fioc, M., & Rocca-Volmerange, B. 1997, A&A, 326, 950
García-Rojas, J., & Esteban, C. 2007, ApJ, 670, 457
Garnett, D. R., & Shields, G. A. 1987, ApJ, 317, 82
Gregory, P. C. (ed.) 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with "Mathematica" Support (Cambridge: Cambridge Univ. Press)
Grevesse, N., Asplund, M., Sauval, A. J., & Scott, P. 2010, Ap&SS, 328, 179
Groves, B. A., Dopita, M. A., & Sutherland, R. S. 2004, ApJS, 153, 9
Hillier, D. J., & Miller, D. L. 1998, ApJ, 496, 407
Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35
Kewley, L. J., Dopita, M. A., Leitherer, C., et al. 2013a, ApJ, 774, 100
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Kewley, L. J., Maier, C., Yabe, K., et al. 2013b, ApJL, 774, L10
Kobulnicky, H. A., & Kewley, L. J. 2004, ApJ, 617, 240
Köppen, J., Weidner, C., & Kroupa, P. 2007, MNRAS, 375, 673
Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53
Lara-López, M. A., Hopkins, A. M., López-Sánchez, A. R., et al. 2013, MNRAS, 433, L35
Larson, R. B. 1974, MNRAS, 169, 229
Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3
Lejeune, T., Cuisinier, F., & Buser, R. 1997, A&AS, 125, 229
Levesque, E. M., Kewley, L. J., & Larson, K. L. 2010, AJ, 139, 712
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Lilly, S. J., Carollo, C. M., & Stockton, A. N. 2003, ApJ, 597, 730

Liu, X.-W., Storey, P. J., Barlow, M. J., et al. 2000, MNRAS, 312, 585
López-Sánchez, Á. R., Dopita, M. A., Kewley, L. J., et al. 2012, MNRAS, 426, 2630
López-Sánchez, Á. R., & Esteban, C. 2010, A&A, 517, A85
López-Sánchez, Á. R., Esteban, C., García-Rojas, J., Peimbert, M., & Rodríguez, M. 2007, ApJ, 656, 168
Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Mannucci, F., Cresci, G., Maiolino, R., et al. 2009, MNRAS, 398, 1915
McCall, M. L., Rybski, P. M., & Shields, G. A. 1985, ApJS, 57, 1
McGaugh, S. S. 1991, ApJ, 380, 140
Meynet, G., & Maeder, A. 2007, A&A, 464, L11
Meynet, G., Maeder, A., Schaller, G., Schaerer, D., & Charbonnel, C. 1994, A&AS, 103, 97
Moustakas, J., Kennicutt, R. C., Jr., Tremonti, C. A., et al. 2010, ApJS, 190, 233
Nagao, T., Maiolino, R., & Marconi, A. 2006, A&A, 459, 85
Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2013, ApJ, 769, 3
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
Pagel, B. E. J., Edmunds, M. G., Blackwell, D. E., Chun, M. S., & Smith, G. 1979, MNRAS, 189, 95
Pauldrach, A. W. A., Hoffmann, T. L., & Lennon, M. 2001, A&A, 375, 161
Peimbert, A. 2003, ApJ, 584, 735
Peimbert, A., Peimbert, M., & Ruiz, M. T. 2005, ApJ, 634, 1056
Peimbert, M. 1967, ApJ, 150, 825
Peimbert, M., & Costero, R. 1969, BOTT, 5, 3
Peimbert, M., Peimbert, A., Esteban, C., et al. 2007, RMxAA, 29, 72
Pérez-Montero, E. 2014, MNRAS, 441, 2663
Pérez-Montero, E., & Contini, T. 2009, MNRAS, 398, 949
Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pilyugin, L. S. 2001, A&A, 369, 594
Pilyugin, L. S., Grebel, E. K., & Mattsson, L. 2012, MNRAS, 424, 2316
Pilyugin, L. S., & Mattsson, L. 2011, MNRAS, 412, 1145
Pilyugin, L. S., & Thuan, T. X. 2005, ApJ, 631, 231
Pilyugin, L. S., Vílchez, J. M., & Thuan, T. X. 2010, ApJ, 720, 1738
Richardson, M. L. A., Levesque, E. M., McLinden, E. M., et al. 2013, arXiv:1309.1169
Salpeter, E. E. 1955, ApJ, 121, 161
Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, A&AS, 96, 269
Schmutz, W., Leitherer, C., & Gruenwald, R. 1992, PASP, 104, 1164
Shields, G. A., & Searle, L. 1978, ApJ, 222, 821
Simón-Díaz, S., & Stasińska, G. 2011, A&A, 526, A48
Stasińska, G. 1978, A&A, 66, 257
Stasińska, G. 2002, RMxAA, 12, 62
Stasińska, G. 2004, in Cosmochemistry: The Melting Pot of the Elements, ed. C. Esteban, R. J. Garcí López, A. Herrero, & F. Sánchez (Cambridge: Cambridge Univ. Press), 115
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165
Strömgren, B. 1939, ApJ, 89, 526
Sutherland, R. S., & Dopita, M. A. 1993, ApJS, 88, 253
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Troncoso, P., Maiolino, R., Sommariva, V., et al. 2014, A&A, 563, A58
Tsamis, Y. G., Barlow, M. J., Liu, X.-W., Danziger, I. J., & Storey, P. J. 2003, MNRAS, 338, 687
van Zee, L., Salzer, J. J., & Haynes, M. P. 1998a, ApJL, 497, L1
van Zee, L., Salzer, J. J., Haynes, M. P., O'Donoghue, A. A., & Balonek, T. J. 1998b, AJ, 116, 2805
Vázquez, G. A., & Leitherer, C. 2005, ApJ, 621, 695
Vilchez, J. M., & Esteban, C. 1996, MNRAS, 280, 720
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zahid, J., Dima, G., Kudritzki, R., et al. 2014, ApJ, 791, 130