MONTE CARLO METHOD FOR CALCULATING UNCERTAINTY IN OXYGEN ABUNDANCE FROM STRONG-LINE FLUX MEASUREMENTS

Seung Man Oh[1,2], Maryam Modjaz[1], David Fierroz[1], Federica Bianco[1], Yuqian Liu[1], Lisa Kewley[3,4]

Draft version January 17, 2015

ABSTRACT

We present here a Python implementation for the determination of the strong-emission-line estimators of oxygen abundance by Kewley & Dopita (2002). The standard strong line Metallicity scales and diagnostics IMPROVE have been used to estimate metal abundance by using emission line ratios. Here we introduce Monte Carlo resampling to these methods in order to better characterize an oxygen abundance confidence region. We output probability distributions, measured values for metallicity and, when desired for reddening E(B-V). We test our code on emission lines measurements from a sample of galaxies ($z < 0.15$) and compare our metallicity results with those from previous methods. **MODIFY AT THE END,**. We show that our metallicity estimate is consistent with previous methods but yields smaller uncertainties. The code is open source and can be found at www.github.com/snyugroup.

Subject headings:

1. INTRODUCTION

The low quantity of carbon, oxygen, nitrogen, sulfur and iron among other elements provide a splash of color to the otherwise dominating greyscape of hydrogen and helium in the stars and gas of galaxies. Nevertheless, even the minute presence of heavy elements (all elements heavier than H and He, also called metals or collectively metallicity) is important for many areas of astrophysics. For example, Johnson & Li (2012) suggest that if it was not for the relatively high metallicity level in our solar system, planet formation may not have been possible. With Z representing the mass fraction of metals, for own sun the value is measured to be Z=0.0153 (Caffau et al. 2011), though there are others who suggest a lower solar metallicity of Z=0.0134 in particular because of oxygen (Asplund et al. 2009; Grevesse et al. 2010)[5]. Furthermore, when properly observed and estimated, metallicity measurements of the galactic gas can tightly constrain models of galaxy formation and evolution (e.g., Kewley & Ellison 2008 and references therein), as well as shed light on the metallicity dependence and production conditions for different types of SNe and long-duration GRBs (e.g., Modjaz et al. 2008; Levesque et al. 2010; Anderson et al. 2010; Modjaz et al. 2011; Kelly & Kirshner 2012; Sanders et al. 2012).

However, for almost all astronomical objects, metallicity cannot be measured directly. The oxygen abundance in the gas-phase is the canonical choice of metallicity indicator for interstellar medium (ISM) studies, since oxygen is the most abundant metal and only weakly depleted

[1] Center for Cosmology and Particle Physics, New York University, 4 Washington Place, New York, NY 10003, USA
[2] NYU Abu Dhabi PO Box 129188 Abu Dhabi, UAE
[3] Australian National University, Research School for Astronomy & Astrophysics, Mount Stromlo Observatory, Cotter Road, Weston, ACT 2611, Australia
[4] Institute of Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

[5] Note that these abundances refer to the current abundances in the sun, which are lower than the value with which the sun was formed 4.56 Gyr ago, since diffusion at the bottom of the convection zone has decreased metallicity over time (Grevesse et al. 2010).

onto dust grains (in contrast to refractory elements such as Mg, Si, Fe, with Fe being depleted by more than a factor of 10 in Orion; see Simón-Díaz & Stasińska 2011). The oxygen abundance[6] is expressed as $12 + \log_{10}(\frac{O}{H})$ where O and H are the number fractions of Oxygen and Hydrogen, respectively. Importantly, oxygen exhibits very strong nebular lines in the optical wavelength range of HII regions (e.g., Pagel et al. 1979; Osterbrock 1989; Tremonti et al. 2004), and thus, well-established diagnostic techniques have been developed (e.g., Kewley & Dopita 2002; Pettini & Pagel 2004; Kobulnicky & Kewley 2004; Kewley & Ellison 2008).

1.1. *Oxygen Abundance diagnostics*

The "classical" way to estimate the oxygen abundance is the electron temperature (T_e) method, which estimates the electron temperature and density of the nebular using a number of oxygen lines with different ionization states, including the auroral [OIII] $\lambda4363$ line, to then directly estimate the OII and OIII abundances to obtain the total oxygen abundance, after correcting for the unseen stages of ionization. However, the auroral [OIII] $\lambda4363$ line is very weak, even in low-metallicity environments, and thus, other methods had to be developed that use other, stronger lines, in the spectra of HII regions. These are called strong-line methods and are the subject of this manuscript. Strong-line methods can be basically categorized into two types: theoretical methods, that rely on calibrating various observed line ratios using photoionization methods (basically theoretically simulating HII regions, using stellar model atmospheres, stellar populations and photoionization models) and empirical ones that calibrate various observed line ratios using observed T_e-based metallicities.

For the theoretical strong-line method, one ratio that is commonly used to determine the metallicity of galaxies is ([OII] $\lambda3727$ + [OIII] $\lambda4959, \lambda5007$)/H$_\beta$ (Pagel et al. 1979) and is referred to as R23. The drawback of this

[6] We note that in many cases in the literature, including here, the terms metallicity and oxygen abundance are used interchangeably.

method is that it is double-valued with metallicity, and thus other line ratios need to be used to break the degeneracy between the high values ("upper branch") and the low values ("lower branch") of the R23 metallicities. Furthermore, Kewley & Dopita (2002) showed the importance of ionization parameter, which can be physically understood to correspond to the maximum velocity of an ionized front that can be driven by the local radiation field of hot massive stars that are ionizing the ISM gas, that needs to be taken into account in the various strong-line methods, as HII regions with different ionization parameters produce different line strengths for the same metallicity. Calibrations of R23 by McGaugh (1991) (hereafter M91) and by Kewley & Dopita (2002) (hereafter KD02) use different theoretical photoionization models and take ionization parameter into account, while other calibrations such as of Zaritsky et al. (1994) (hereafter Z94) do not. Thus, Z94 is mostly valid for only metal-rich galaxies. KD02 use an iterative process and two different ratios [NII]/[OII] and [NII]/H$_\alpha$ to break the R23 degeneracy in order to arrive at the metallicity estimate.

As to empirical strong-line methods, the most commonly used one is by Pettini & Pagel (2004) (hereafter PP04). PP04 used HII regions with Te-based metallicities to derive empirical fits to strong-line ratios, and introduce the line ratios of ([NII]/H$_\alpha$ (N2) and ([OIII]/H$_\beta$)/([NII]/H$_\alpha$ (O2N2) as metallicity diagnostics. Since PP04-N2 employs two closely spaced lines (Halpha and NII), which are not affected by stellar absorption, nor uncertain reddening, and are easily observed in one simple spectroscopic setup, it has become an often-used scale for at least low-z SN host galaxy studies (e.g. see metal-analysis by e.g,. Sanders et al. (2012); Modjaz (2012); Leloudas et al. (2014). However, it is important to remember that this scale has a number of short-comings: it does not take into account the impact of ionization parameter, was initially derived based on only 137 extragalactic HII regions, and the used nitrogen emission line saturates at high metallicity, and thus this method may not be well-suited for high-metallicity galaxies. An updated calibration by **?** based on many more Te-based metallicities (almost three time larger that that of PP04) derives a significantly shallower slope between O3N2 index and oxygen abundance than the PP04 calibration.

As can be seen, each scale has different advantages and disadvantages and may be used in different metallicity regimes (see detailed discussion in e.g., Kewley & Dopita 2002; Stasińska 2002; Kewley & Ellison 2008; Moustakas et al. 2010; Dopita et al. 2013; Blanc et al. 2015). Thus, this open-source code outputs the oxygen abundance in the main 6 metallicity scales (for which the KD02 diagnostic has four outputs and the PP04 diagnostic has two outputs). While there is a long-standing debate about which diagnostic to use, as there are systematic metallicity offsets between different methods (recombination lines vs. strong-line method vs. "direct" T_e method, see the above sources), **the relative metallicity trends can be considered robust, if the analysis is performed self-consistently in the same scale, and trends are seen across different scales (Kewley & Ellison 2008; Moustakas et al. 2010)**. Note however, that while there are conversion values between different

scales (Kewley & Ellison 2008), they apply for large data sets, since those conversion values were derived based on ten thousands of SDSS galaxies, and thus should be used with caution (or not at all) for smaller samples. In addition, there the debate about the value of the solar oxygen abundance value (Asplund et al. 2009; Caffau et al. 2011), such that the absolute oxygen calibration is still uncertain.

We assume that the observed emission lines to be used to indicate metallicity are from HII regions and are not due to non-thermal excitation by e.g., AGN or interstellar shocks from SNe or stellar winds. Tests to exclude data contaminated by such non-thermal sources have to be executed by using the recommended line ratios by e.g., **???**.

Here we introduce the open-source code "". pro .. In § 2 we describe our method, the input and output values of the code. In § 4, we compare our method of obtaining abundance uncertainties to previous methods in the literature.

2. METHOD

For computing oxygen abundances, we use the iterative code by Kewley & Dopita (2002), which has been updated in Kewley & Ellison (2008) and reflects .. **LISA: YOUR INPUT HERE:what is the update??** which was initially written in IDL. We translated it into python, added the new feature of obtaining uncertainties on the metallicity outputs via bootstrap resampling, and made it open source via github, as we explain below.

Emission line flux values are fed into our Python implementation as in the original code by Kewley & Dopita, which is written in IDL and hereafter referred to as IDL02. The inputs are emission line flux values and their uncertainties for the following lines: Hα, Hβ, [OI] 6300, [OII] 3727, [OIII] 4959, [OIII] 5007, [NII] 6584, [SII] 6717, [SII] 6731, SIII 9532, and SIII 9096 and their uncertainties. If the fluxes for the specified lines are not available, the entry is left to 0 and the outputted oxygen abundance scales will be in only metallicity scales that use the line fluxes provided. The inputted line fluxes are corrected for reddening by using the observed Balmer decrement, for which Hα and Hβ flux values are needed to be provided. We assume case B recombination and thus, the standard value of 2.86 as the intrinsic Hα/Hβ ratio (Osterbrock 1989), and apply the standard Galactic reddening law with $R_V = 3.1$ (Cardelli et al. 1989). However, the user can choose other extinction laws and R_V values, if desired, given the code's open-source nature.

While other parameters, such as the ionization parameter q and the electron density (using the $SiII$ lines) are computed as long as the necessary lines are provided, they are not outputted in the current version of our code − however, the reader is welcome to easily modify the code to suite their needs, given it's an open-source code.

As output, we obtain metallicity values and their uncertainties in the following calibrations, as discussed in detail in Kewley & Dopita (2002); Kewley & Ellison (2008): Kewley & Dopita (2002)(KD02, for the 4 following computations: R23, using the NIIOII ratio, using the NII/Halpha ratio, and a combined method that uses the optimal method given the input line fluxes), McGaugh (1991) (M91), Zaritsky et al. (1994) (Z94), Pilyugin (2001) (P01), Denicoló et al. (2002) (D02), Pettini &

Pagel (2004) (2 computations: PP04-N2, PP04-O3N2). If the line fluxes necessary for specific scales are not provided, the output will be -1.

The novel aspect of our work is that we introduce a Monte Carlo (MC) algorithm to obtain numerous iterations and random sampling to obtain a more robust result with respect to the uncertainty in the input values. We apply the bootstrapping method which is a resampling technique for error estimation (e.g., Efron 1979; Hastie et al. 2009; Andrae 2010. Given a data set with error bars from which some parameters are estimated, bootstrapping consists of resampling the data to produce alternative data sets. Here we generate a distribution of possible inputs by generating a gaussian distribution centered on the line flux measurement value, with a standard deviation corresponding to the measurement error, where we have made the assumption that the line flux error is gaussian distributed in nature. Every iteration randomly takes an input value from this distribution (for each inputted line) to run the calculations. This effectively emulates conducting multiple experiments when repeated observation is impractical or impossible, as in the case of the emission line flux data, and thus generates alternative data sets. We generated N bootstrap samples of emission line fluxes (where an appropriate value of $N = 50,000$ is determined below) and calculated the metallicity for each set of line fluxes. At the end of the iterations, likelihood distributions of the oxygen abundance value are generated for each scale by binning the results to histograms and deriving a confidence region (see below).

This MC bootstrap approach takes into account the impact of the uncertain reddening (due to the uncertainties in the measurement of the Hα and Hβ fluxes), when the option for de-reddened metallicities is chosen. Since for each bootstrap iteration, a new reddening value is calculated based on the resampled Hα and Hβ fluxes, which is used to compute the de-reddened metallicity value, the derived distribution of metallicity values takes into account the uncertain redding. As part of the optional output, a distribution plot for E(B-V) will be provided, along with confidence intervals **add the feature of getting error bars on E(B-V)???**. If either Hα or Hβ flux is not provided, then no reddening correction can be applied and the computed metallicity will not be reddening-corrected.

However, we note that our code naturally does not include the *systematic* uncertainty of each scale, which are e.g., ~ 0.07 dex for PP04-O2N2. Thus systematic errors can be as large, if not larger than the statistical errors.

2.1. *Histogram Bin Size and Confidence Region*

Choosing the binning for a histogram is not trivial and Hogg (2008) describes various data analysis recipes for selecting a histogram bin size. Too many bins will result in many empty bins and an "over-fit" histogram, while too few bins may lose features of a probability distribution. Using a bin size of $\sqrt{N}$ was recommended, but this proved to be slightly over-fitting, and after a number of tests, we found $2 * \sqrt[3]{N}$ to be appropriate for most cases.

In determining the confidence region interval, there are broadly three approaches (e.g., Andrae 2010): choosing a symmetric interval, the shortest interval , or a central interval. We determined the confidence interval using the "central" method, because it ensures that the algorithm finds the proper boundaries even in the case of multiple peaks (i.e., multimodal likelihood distributions) and for asymmetric, non-Gaussian, distributions. With the "central" method we determined the confidence interval by choosing the left and right boundaries so that the region outside the confidence interval each equally contains 16% of the total distribution for one standard deviation confidence interval - in analogy to the one-sigma-interval of the Gaussian distribution. While the histograms at sufficiently high N, where N is the total number of iterations, yielded single peaked results, at lower N there were occasional multiple peaks resulting from an non-smooth gaussian being sampled. In selecting the value of N, we found that around $N = 50,000$ provides reliably smooth histograms. The fiftieth (50%) percentile, i.e. the median, is reported as the measured value.

3. COMPARISON

(will edit more) A previous method for determining the uncertainty in the oxygen abundance (as used in Modjaz et al. 2008; **?**; **?**, 2011 was an analytic approach of propagating the emission-line flux uncertainties by finding the difference in maximum and minimum possible abundances via maximizing and minimizing, respectively, the various line ratios. As can be seen from Fig.**??**, the MC method has a smaller uncertainty than the analytic method (mean of xx dex). This could be attributed to it reflecting on a more realistic confidence region. The method used the worst case scenario

Rupke: Errors were propagated primarily using analytic expressions, but for abundance and gradient errors we employed Monte Carlo methods.

Here we discuss comparisons with the literature.

IZI:

Blanc et al. (2015) employ Bayesian inference for deriving using photoionization codes given a set of observed line fluxes and an input photoionization model

Using the MC method for finding uncertainty in metallicity has been done by IZI -¿ FILL IN only for one scale and calibration method

Sanders: by propagation of the uncertainties in the line flux measurements and the 0.07 dex diagnostic systematic error. (line flux errors from MCMC fitting of a gaussian to emission lines)

4. CONCLUSIONS

REFERENCES

Anderson, J. P., Covarrubias, R. A., James, P. A., Hamuy, M., & Habergham, S. M. 2010, MNRAS, 407, 2660

Andrae, R. 2010, ArXiv e-prints, arXiv:1009.2755
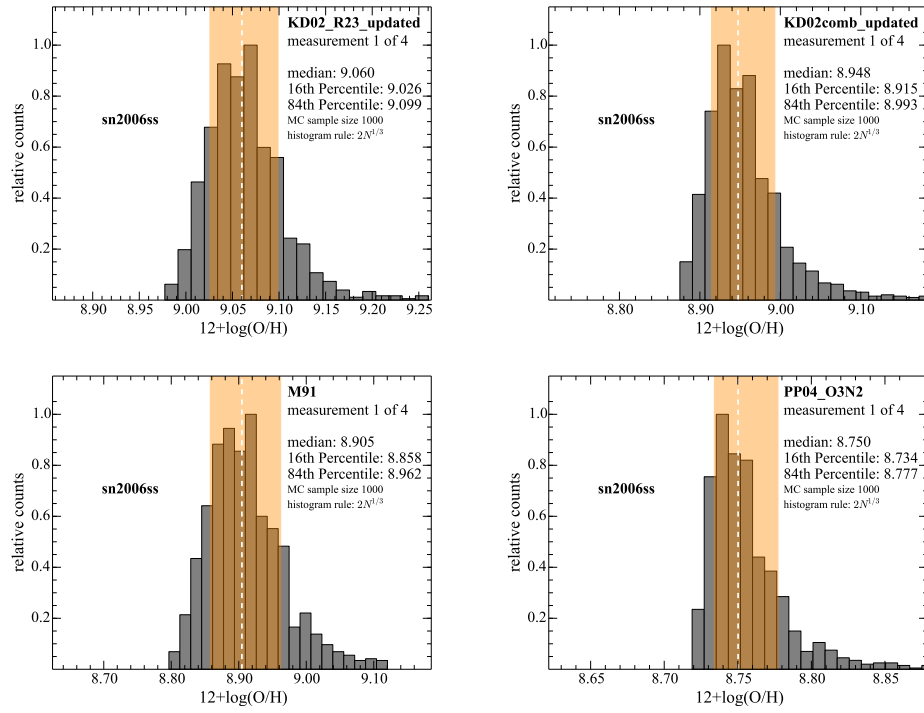
FIG. 1.— Examples of confidence region determined by the algorithm, shaded in tan color.

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481

Blanc, G. A., Kewley, L., Vogt, F. P. A., & Dopita, M. A. 2015, ApJ, 798, 99

Caffau, E., Ludwig, H.-G., Steffen, M., Freytag, B., & Bonifacio, P. 2011, Sol. Phys., 268, 255

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Denicoló, G., Terlevich, R., & Terlevich, E. 2002, MNRAS, 330, 69

Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, ApJS, 208, 10

Efron, R. 1979, Ann. Stat., 7, 1

Grevesse, N., Asplund, M., Sauval, A. J., & Scott, P. 2010, Ap&SS, 328, 179

Hastie, T., Tibshirani, R., & Friedman, J. 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer Science+Business Media, New York)

Hogg, D. W. 2008, ArXiv e-prints, arXiv:0807.4820

Johnson, J. L., & Li, H. 2012, ApJ, 751, 81

Kelly, P. L., & Kirshner, R. P. 2012, ApJ, 759, 107

Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35

Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183

Kobulnicky, H. A., & Kewley, L. J. 2004, ApJ, 617, 240

Leloudas, G., Schulze, S., Kruehler, T., et al. 2014, ArXiv e-prints, arXiv:1409.8331

Levesque, E. M., Berger, E., Kewley, L. J., & Bagley, M. M. 2010, AJ, 139, 694

McGaugh, S. S. 1991, ApJ, 380, 140

Modjaz, M. 2012, in IAU Symposium, Vol. 279, IAU Symposium, 207–211

Modjaz, M., Kewley, L., Bloom, J. S., et al. 2011, ApJ, 731, L4

Modjaz, M., Kewley, L., Kirshner, R. P., et al. 2008, AJ, 135, 1136

Moustakas, J., Kennicutt, Jr., R. C., Tremonti, C. A., et al. 2010, ApJS, 190, 233

Osterbrock, D. E. 1989, Astrophysics of Gaseous Nebulae and Active Galaxies (Mill Vallery: University Science Books)

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

Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59

Pilyugin, L. S. 2001, A&A, 369, 594

Sanders, N. E., Soderberg, A. M., Levesque, E. M., et al. 2012, ApJ, 758, 132

Simón-Díaz, S., & Stasińska, G. 2011, A&A, 526, A48+

Stasińska, G. 2002, ArXiv Astrophysics e-prints, astro-ph/0207500

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Zaritsky, D., Kennicutt, Jr., R. C., & Huchra, J. P. 1994, ApJ, 420, 87