Draft version March 11, 2015
Preprint typeset using LaTeX style emulateapj v. 04/17/13

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

Author order to be determined: Maryam Modjaz[1], Federica B. Bianco[1], Seung Man Oh[1,2], David Fierroz[1], Yuqian Liu[1], Lisa Kewley[3,4]

Draft version March 11, 2015

ABSTRACT

MODIFY & FINALIZE AT THE END: We present an open-source Python code for the determination of the strong-emission-line estimators of oxygen abundance in the standard scales, based on the original IDL-code in Kewley & Dopita (2002). The standard strong line Metallicity scales and diagnostics IMPROVE have been used to estimate metal abundance through emission line ratios. Here we introduce a Monte Carlo resampling of these methods in order to better characterize an oxygen abundance confidence region. We output median values, 16th and 84th percentile confidence regions, for various standard metallicity diagnostics, and, when possible, for reddening E(B-V). We produce Monte Carlo parameter distributions for the oxygen abundance and when possible 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. We show that our metallicity estimates are consistent with previous methods but yields smaller uncertainties. The code is open source and can be found at `www.github.com/nyusngroup/` **(add repo and DOI)**.

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 our 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 galaxies can tightly constrain models of galaxy formation and evolution (e.g., Kewley & Ellison 2008 and references therein), tas 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; Lunnan et al. 2014; Leloudas et al. 2014), to just name a few examples.

[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).

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 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 represent the number of Oxygen and Hydrogen atoms, 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, many different diagnostic techniques, relying on different lines, have been developed (e.g., Kewley & Dopita 2002; Pettini & Pagel 2004; Kobulnicky & Kewley 2004; Kewley & Ellison 2008), which are discussed in the next section.

1.1. *The different oxygen abundance diagnostics*

Here we present a brief overview of the various observational methods for measuring the gas-phase oxygen abundance - however, for a full discussion with all the caveats we encourage the reader to see reviews by e.g. Stasińska (2002); Kewley & Ellison (2008); Moustakas et al. (2010); Stasińska (2010); Dopita et al. (2013); Blanc et al. (2015). The so-called "classical" way to estimate the oxygen abundance is the electron temperature (T_e) method, which estimates the electron temperature and density of the nebula using a number of oxygen lines with different ionization states, including the auroral [OIII] λ4363 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] λ4363 line is very weak, even

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

in low-metallicity environments, and saturates at higher metallicity (since at higher metallicities the cooling is dominated by the Oxygen NIR fine structure lines) – 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 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 population synthesis and photoionization models) and empirical ones that calibrate various observed strong line ratios using observed T_e-based metallicities. While historically there have been large systematic offsets between the T_e method and the strong line methods, Dopita et al. (2013) demonstrated that the T_e method gives the same results as the strong line methods, if the energy distribution of the electrons in the HII regions is assumed to not be a simple Maxwell-Boltzmann distribution (as assumed in prior works), but a more realistic κ distribution, as observed in solar system astrophysical plasma. They also find that the effect of the κ distribution on the strong-line methods is minor.

For the theoretical strong-line method, one ratio that is commonly used to determine the metallicity of galaxies is ([OII] λ3727 + [OIII] $\lambda 4959, \lambda 5007$)/$H_\beta$ (Pagel et al. 1979) and is referred to as R23. The drawback of this 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. This ionization parameter needs to be taken into account in the various strong-line methods, as HII regions at the same metallicity but with different ionization parameters produce different line strengths. Calibrations of R23 by McGaugh (1991) (hereafter M91) and by Kewley & Dopita (2002) (hereafter KD02) use different theoretical photoionization models and take the 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. M91 and KD02 use an iterative process to break the R23 degeneracy (KD02 uses different ratios [NII]/[OII] and [NII]/H_α) and to also constrain the ionization parameter q in order to arrive at the metallicity estimate.

As to empirical strong-line methods, the most commonly used one is that by Pettini & Pagel (2004) (hereafter PP04). PP04 used HII regions with T_e-based metallicities to derive empirical fits to strong-line ratios, and introduce the line ratios of ([NII]/H_α (N2) and ([OIII]/H_β)/([NII]/H_α (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 the ionization parameter, it was initially derived based on only 137 extragalactic HII regions, and the nitrogen emission line employed saturates at high metallicity (**CHECK!**), and thus this method may not be well-suited for high-metallicity galaxies. An updated calibration by Marino et al. (2013) 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 it can be seen, each scale has different advantages and disadvantages and should 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, one should note that there is a debate about the value of the solar oxygen abundance (Asplund et al. 2009; Caffau et al. 2011), such that the absolute oxygen calibration is still uncertain.

Here we introduce the open-source python 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. DESCRIPTION OF METALLICITY CODE

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 the code into python, added new features, most importantly the capability of obtaining uncertainties on the metallicity outputs via Monte Carlo resampling, and made it open source on *github*, as we explain below.

We assume that the observed emission lines to be used to indicate metallicity originate in 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 should be executed using the recommended line ratios by e.g., Baldwin et al. 1981; Kauffmann et al. 2003; Kewley et al. 2006 prior to running this code. Furthermore these lines should have all the correct calibration (at least correct relative calibration) and should have a signal-to-noise ratio (S/N) of at least 3. The latter is important for the

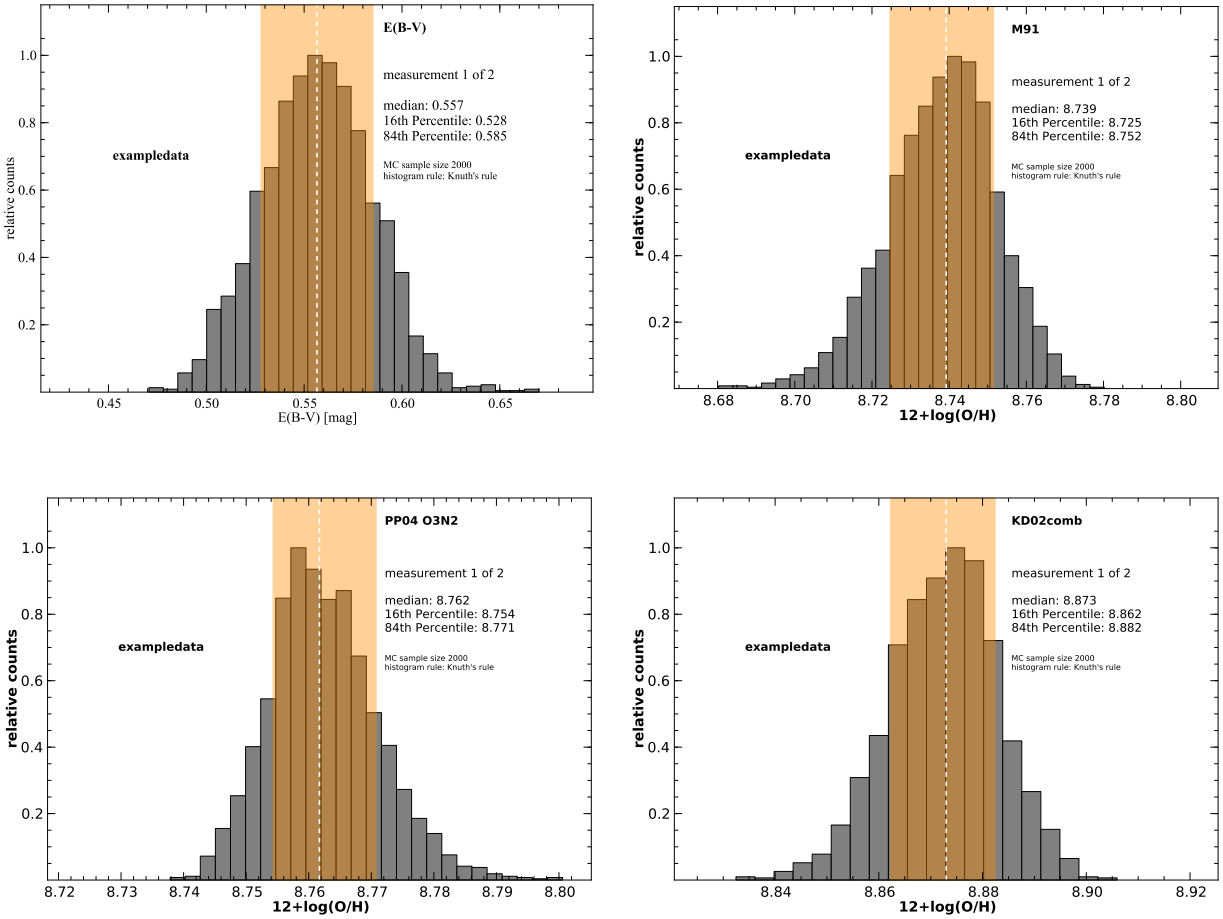

FIG. 1.— WILL UPDATE WITH NEW FIGURES BASED ON A PUBLISHED SN WITH SIGNFICANT E(B-V) AND WRITE MORE TEXT. Metallicity and reddening ($E(B-V)$) parameter distributions based on example data (namely, emission line data of the HII regions at the position of SN 2008D, published in Modjaz et al. (2011), based on N=20,0000 samples. The median value is shown with the dashed lines, while the 16th till 84th confidence region is shaded in tan color. The shown metallicity scales are by Kewley & Dopita (2002), updated by Kewley & Ellison (2008) (KD02combupdated), Pettini & Pagel (2004), using OIII and NII (PP04-O3N2) and McGaugh (1991) (M91). These are the same plots that are outputted by the code, which produces such plots for all scales (namely those of s**LIST THEM ALL**). The legend also includes the ID number of

success of the Monte Carlo resampling technique as described below.

2.1. *Input and Output of code*

Emission line flux values are fed into our Python implementation as in the original IDL code by Kewley & Dopita (2002), hereafter referred to as IDLKD02. 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 **CHECK AT THE END WITH CODE!** (consult the README.md in the *github* repository for details about the input format). If the fluxes for the specified lines are not available, the entry is left to 'NaN' and the outputted oxygen abundance will be calculated only for metallicity scales that use the provided line fluxes. As part of the code, the inputted line fluxes are corrected for reddening by using the observed Balmer decrement, for which Hα and Hβ flux values need 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 can 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 [NII]/[OII] ratio, using the [NII]/Hα 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), **?** (both the diagnostic based on R23, C01_R23, and based on [NII]/[SII], C01_N2S2), Denicoló et al. (2002) (D02), Pettini & Pagel (2004) (2 computations: PP04_N2, PP04_O3N2) **CHECK WITH FINAL CODE - ALSO CHECK whether P01 is P01 or P05 by looking at the code**. In addition, if the user has installed the publicaly available `pyqz`[7] Python Module, based on **?**, [NII], [SII], [OIII], Hα, and Hβ lines are fed to the `pyqz` module, which produces up to 8 emission line ratio diagnostics for 12+log(O/H), each using two of the line ratios [NII]/[SII], [NII]/Hα, [OIII]/[SII], and [OIII]/Hβ. The `pyqz` module assumes a κ-distribution of energy for the electrons in HII regions, as suggested by **?**. Our code sets the κ parameter to 20, which is the value that **?** found to best resolve inconsistencies between oxygen aboundances derived with theoretical methods based on temperature (i.e. from temperature sensitive line ratios such as [OIII] λ4363/[OIII] λ5007) and the R23 based methods[8].

If the line fluxes necessary for specific scales are not provided, the output metallicities will default to 'NaN', if the errors in he measurements are not provided, the code will specify that it cannot create a measurement distribution and determine a confidence interval, but calculate the noinal metallicity.

2.2. *Computing Uncertainties*

The novel aspect of our work is that for every set of input line measurements we introduce a Monte Carlo (MC) resampling method to obtain iterations via random sampling within the measurement errors, and thus we obtain a robust result for error estimation (e.g., Efron 1979; Hastie et al. 2009; Andrae 2010).

Given a data set with error bars from which certain parameters are estimated, Monte Carlo resampling generates synthetic data samples drawing from a given distribution. Here we draw synthetic data from a Gaussian distribution centered on each measured line flux value, with standard deviation corresponding to the measurement error. The implicit assumption is made, of course, that the line flux error is Gaussian distributed in nature[9].

For each metallicity scale, for each of N values chosen randomely within the relevant emission line distributions we run the calculations that derives the metallicity. This effectively simulates conducting multiple experiments when repeating observations is impractical or impossible, as in the case of the emission line flux data, and thus generates alternative data sets. The sample size N is set by the user, and one should expect an appropriate value of N to be a few 1000s, depending on the metallicity scale chosen and measurement errors (for example $N = 2,000$ is determined to be sufficient for our example data, as shown below, and we provide tools to assure the sample size is sufficiently large). A parameter estimate distribution of oxygen abundance is generated for each scale, the metallicity and its confidence region are calculated, and the results are bunned abd visualized in a histogram (see below)[10]. This is done for each scale the user chooses to calculate. The fiftieth (50%) percentile, i.e. the median, is reported as the measuredtrue metallicity value, and the 16th and 84th percentiles of the distribution as its confidence region.

This MC resampling 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 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 output, a probability distribution plot for E(B-V) is provided (see last plot in figure 1, along with confidence intervals derived using the same method as for the metallicity measurements. 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 and the E(B-V) output will be set to zero.

[7] `https://datacommons.anu.edu.au:8443/DataCommons/item/anudc:5037`

[8] The user may wish to modify the value of κ in the code.

[9] The user may wish to provides their own probability distribution for the emission line uncertainties, and easily modify the code to suite their needs.

[10] However, note this method is a conservative approach, since it overestimates the intrinsic metallicity uncertainty, as we are centering this error distribution on the measured values instead of the (unknown) true values (Andrae 2010).

Figure 1 shows the metallicity estimate distribution for 3 representative scales, and for the reddening parameter E(B-V) - similar plots that are outputted by our code for all scales as listed above (not all shown here). Although the input distributions are Gaussian, the metallicity distributions are not, for two reasons: first, since the metallicities are computed based on log values of line flux ratios, symmetric error bars in linear space will translate into asymmetric error bars in log space; and second, some metallicity computations are non-linear, and (especially those that include R_{23}) sometimes bimodal since they choose upper vs lower branch to break degeneracy.

Since the metallicity distributions are rarely Gaussian, the percentiles we report cannot be expressed in terms of σ values. In determining the confidence region intervals for asymmetric and multi-modal distributions, there are broadly three approaches (e.g., Andrae 2010): choosing a symmetric interval, the shortest interval, or a "central" interval. With the "central" method we determined the confidence interval by choosing the left and right boundaries such that the region outside the confidence interval each equally contains 16% of the total distribution - in analogy to the one-sigma-interval of the Gaussian distribution. This ensures that the algorithm finds the proper boundaries even for asymmetric, non-Gaussian, distributions, (i.e., multimodal likelihood distributions) and in the case of multiple peaks.

In summary, the output for the measured value corresponds to the fiftieth (50%) percentile, while the lower error bar corresponds to the 50^{th}-16^{th} percentile and the upper error bar corresponds to 84^{th}-50^{th} of the metallicity parameter estimate distribution. However, we urge that the reader always inspect the appropriate metallicity distribution plots, which are also outputted, to check for themselves whether the outputted median and confidence regions properly represent the full metallicity distribution (see Section 2.2.1).

We note that our code does not output the *systematic* uncertainty of each scale, which are e.g., ~ 0.07 dex (or 0.14 dex **CHECK**) for PP04-O2N2. Thus systematic errors can be as large, if not larger, than the statistical errors, however, if all metallicity measurements are in the *same* scale and only *relative* comparisons are made, then the systematic error does not have any impact (by definition!). The distributions for the D02 scale include the uncertainty in the fit parameters: the oxygen aboundance in this scale is generated as $12 + \log(O/H) = 9.12 \pm 0.05 + (0.73 \pm 0.10)$ NII, as published in Denicoló et al. (2002). The parameter fit are generated as the sum of the nominal parameter (9.12 and 0.73) and a Gaussianly distributed random value centered on zero, and within a standard deviation of 0.05 and 0.10, respectively.

2.2.1. *Visual diagnostics*

In order for the user to check the validity of a measurement, and to better understand the distribution, we provide two visualizations: for each metallicity scale calculated, for each input set, we generate a histogram of the output distribution (Figure 1 and 2), and for each set of input line we generate a *box-and-whiskers* plot (herehance *boxplot*, for short) summarizing the result of all scales calculated (Figure 3).

Choosing the binning size for a histogram is not a trivial task. 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 miss features of the distribution. By default, we use *Knuth's Method* to choose the number of bins $N_{\rm bins}$ for each histogram. Knuth's method optimizes a Bayesian fitness function across fixed-width bins (Knuth 2006). Additionally, however, we enable a number of binning options from which the user can choose, including: the square root of the number of bins, *Rice rule* ($N_{\rm bins} = 2\sqrt[3]{N}$, e.g., Hastie et al. 2009), *Doane's formula* ($N_{\rm bins} = 1 + \log_2 N + \log_2\left(1 + {\rm Kurt}\sqrt{(N/6)}\right)$, where Kurt is the third standardized moment of the distribution Doane 1976[11]), and the full Bayesian solution, known as Bayesian Blocks, which optimizes a fitness function across an arbitrary configuration of bins, such that the bins are of variable size (Scargle et al. 2013). The implementation of the latter method requires the `astroML` python package to be installed on the user's system (Vanderplas et al. 2012[12]). If the `astroML` package is not found, the code will default to Knuth's Rule.

As mentioned, Knuth's method implies an optimization. In cases in which the convergence of this minimization takes too long (or if the number of bins after the minimization is $N_{\rm bins}/\sqrt{N} > 5$ or $N_{\rm bins}/\sqrt{N} < 1/3$) the code will revert to Rice rule.

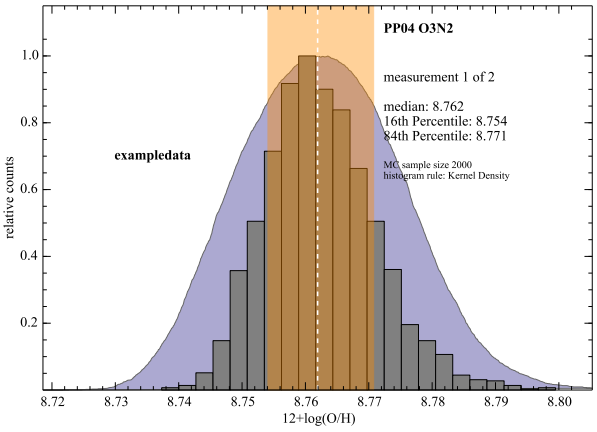

FIG. 2.— **FINISH**

Lastly, we enable the user to obtain a distribution Kernel Density if the `sklearn` is available. The Kernel Density of the distribution is then calculated via *KD Tree* with a top-hat function, as explained in the sklearn package documentation [13]. The results will then show both a histogram, with $N_{\rm bins}$ chosen via Knuth's method, as well as the distribution Kernel Density, as shown in Figure 2.

The boxplot summarizes the resut from each scale the user chooses to calculate. For each scale a box is shown,

[11] Doane (1976) attempted to address the issue of finding the proper number of bins for the histogram of a skewed distribution. Several version of the so-called Doane's formula can be found in the literature. Our formula can, for example, be found in **?**

[12] `https://github.com/astroML/astroML`

[13] `http://scikit-learn.org/stable/modules/density.html`

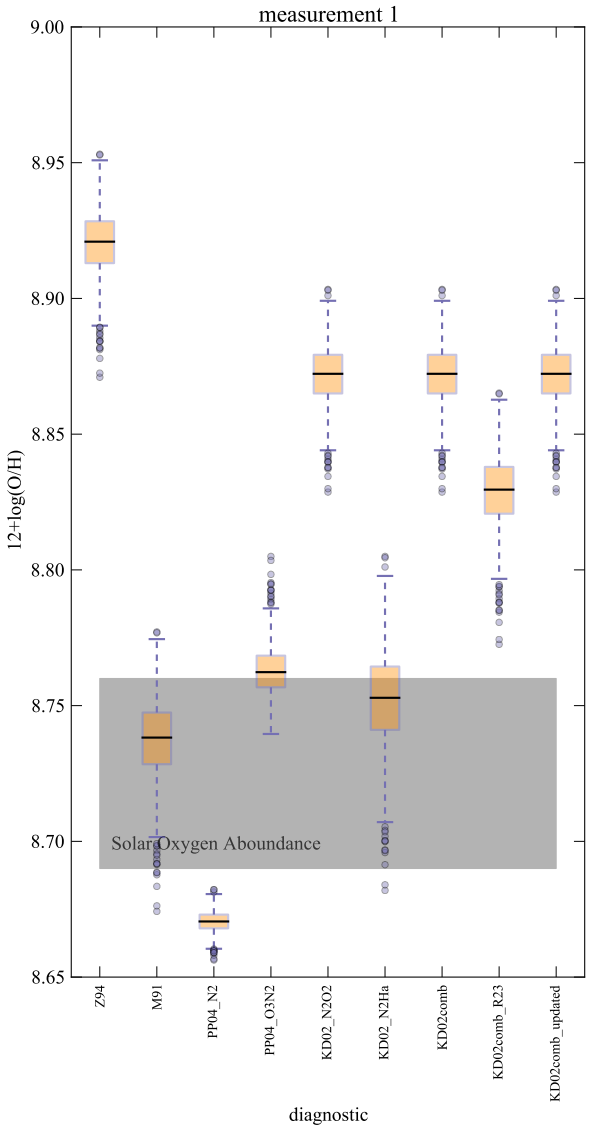

FIG. 3.— **FINISH**

and the box height represents the 25% percentile of the $12 + \log(O/H)$ distribution. Inside the box a black line represents the distribution median, the bars represent the maximum and minimum value of the distribution, excluding outliers, and the outliers are plotted as circles, and are defined as all datapoints more than one and a half times the length of the box from either end of the box (or 1.5xIQR, where IQR is the the *interquartile range*). The solar oxigen aboundance is indicated in this plot for comparison: a gray box shows a range of estimated values for solar oxigen aboundances, from $12 + \log(O/H)$=8.69 (**?**) to $12 + \log(O/H)$=8.76 (**?**). Notice that only the diagnostics requested by the user have a slot in the plot (in the example in Figure 2 the computed scales are M91, the PP04 scales, and the KD02 scales. However these slot exists on the plot whether the diagnostic can be produced or not, i.e. if the input lines to not allow a requested scale to be calculated an empty coulmn will be generated in this plot in correspondence of said metallicity scale.

3. COMPARISON TO PRIOR UNCERTAINTY COMPUTATION AND OTHER WORKS

A previous method for determining the uncertainty in the oxygen abundance (as used in Modjaz et al. 2008; Kewley et al. 2010; Rupke et al. 2010; Modjaz et al. 2011) was an *analytic* approach of propagating the emission-line flux uncertainties: it found the maximum and minimum abundances via maximizing and minimizing, respectively, the various line ratios by adding/subtracting to the measured line values their uncertainties. For comparison we computed the metallicities and their errors in both ways (both analytic and using our current MC resampling method) for 3 representative scales. We plot our results and the residuals in Fig.4, which shows a number of important points: i) The metallicity reported as the 50th percentile of the metallicity parameter distribution from the MC resampling method is completely consistent with the analytically derived metallicity - well within the respective error bars - and thus, the prior published results still stand (unsurprisingly, since our code, aside for the calculation of the confidence interval, uses the same algorithms developed for IDLKD02). ii) The MC resampling method has smaller error bars than the analytic method, especially for the scales of M91 and KD02. This is easily understandable, since the analytic method assumes the worst-case-scenario, as it basically yields 2 metallicity parameter draws (the "minimum" and "maximum") which are in the tail of the full metallicity probability distribution. However, the MC resampling method is the more appropriate method as it empirically characterizes the full parameter estimation distribution.

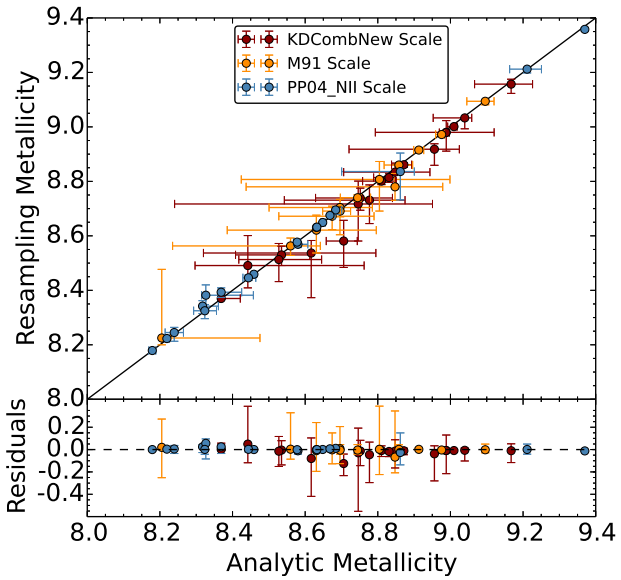

FIG. 4.— **FINISH**.Comparison of metallicity estimation between the analytic method and our Monte Carlo resampling method (top) and their residuals (bottom) for 3 different metallicity scales. Flux measurements come from 19 galaxies previously measured in Modjaz et al. (2011). To add asymmetric errors in quadrature we use $residual_{min} = \sqrt{x_{max}^2 + y_{min}^2}$ and $residual_{max} = \sqrt{x_{min}^2 + y_{max}^2}$

3.1. *Comparison with other works*

The field of SN host metallicity studies has been rapidly developing as these kinds of studies may be crucial avenues for constraining the progenitor systems of different kinds of explosions - however, a few of the works do not even compute errors and others not show how they compute their statistical errors or (e.g., Anderson et al. (2010); Leloudas et al. (2011); Sanders et al. (2012); Leloudas et al. (2014)). While Sanders et al. (2012) compute the line flux errors from a Markov-Chain Monte Carlo fitting of a gaussian to the emission lines, they only mention in passing that they propagate the line flux uncertainties into the metallicity measurements, but do not describe how **this is a little harsh, right? i think mentioning it without explaining is generally the standard for iid errors -¿ Maryam: " How would you rephrase it then? "** .

In contrast, the general metallicity field has considered in detail how to estimate the uncertainties in measured metallicities- however, none of those codes are open-source and many of them are for specific scales which were chosen by the authors: Moustakas et al. (2010) also use MC resampling to estimate the metallicity uncertainties (in their case using 500 trials which seems to lead to a Gaussian distribution) but only do this for two scale, KK04 and Pilyugin & Thuan (2005). For computing the metallicities of the SDSS star forming galaxies, Tremonti et al. (2004) fit a combination of stellar population synthesis and CLOUDY (**citation**) photoionization models to the observed strong emission lines [OII], $H\beta$, [OIII], $H\alpha$, NII and SII and report the median of the metallicity likelihood distribution as the metallicity estimate, with the width of the distribution giving the 1 σ (Gaussian) error. However, this is constitutes their own scale (the T04 scale).

In the last stages of preparing this manuscript Blanc et al. (2015) was published. Blanc et al. (2015) employ Bayesian inference for doing something similar to Tremonti et al. (2004) - they use Bayesian inference to derive the joint and marginalized posterior probability density functions for metallicity Z and ionization parameter q given a set of observed line fluxes and an input photoionization model. They provide a publicly available IDL implementation of their method named IZI (inferring metallicities (Z) and ionization parameters) on the author's website.

4. CONCLUSIONS

FINISH. We hope that this open-access code will be used in many different fields where gas-phase metallicities are important, including in the emerging field of SN and GRB host galaxies, where either it is not described how they got or no error bars are computed (e.g., Lunnan et al. 2014). GIven its public-access nature, the users are free to include any new metallicity diagnostics, e.g., that suggested by Dopita et al. (2013).

M. Modjaz is supported in parts by the NSF CAREER award AST-1352405 and by NSF award AST-1413260. FBB is supported by a *James Arthur Fellowship* at the NYU-Center for Cosmology and Particle Physics. Some plots are produced with public code DOI:10.5281/zenodo.15419 available at `https://github.com/fedhere/residuals_pylab`. This research made use of NASA 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.

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
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
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
Doane, D. P. 1976, The American Statistician, 30, 181
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
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
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
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961
Kewley, L. J., Rupke, D., Zahid, H. J., Geller, M. J., & Barton, E. J. 2010, ApJ, 721, L48
Knuth, K. H. 2006, ArXiv Physics e-prints, physics/0605197
Kobulnicky, H. A., & Kewley, L. J. 2004, ApJ, 617, 240
Leloudas, G., Gallazzi, A., Sollerman, J., et al. 2011, A&A, 530, A95
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
Lunnan, R., Chornock, R., Berger, E., et al. 2014, ApJ, 787, 138
Marino, R. A., Rosales-Ortega, F. F., Sánchez, S. F., et al. 2013, A&A, 559, A114
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
Pilyugin, L. S., & Thuan, T. X. 2005, ApJ, 631, 231

Rupke, D. S. N., Kewley, L. J., & Chien, L.-H. 2010, ApJ, 723, 1255
Sanders, N. E., Soderberg, A. M., Levesque, E. M., et al. 2012, ApJ, 758, 132
Scargle, J. D., Norris, J. P., Jackson, B., & Chiang, J. 2013, ApJ, 764, 167
Simón-Díaz, S., & Stasińska, G. 2011, A&A, 526, A48+
Stasińska, G. 2002, ArXiv Astrophysics e-prints, astro-ph/0207500
Stasińska, G. 2010, in IAU Symposium, Vol. 262, IAU Symposium, ed. G. R. Bruzual & S. Charlot, 93–96
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47 –54
Zaritsky, D., Kennicutt, Jr., R. C., & Huchra, J. P. 1994, ApJ, 420, 87

APPENDIX

MINIMUM CODE VALIDATION

WORK WITH FED - and any other code specific things