

RECALIBRATING STRONG-LINE METALLICITY DIAGNOSTICS WITH
ELECTRON TEMPERATURE MEASUREMENTS FROM COMPOSITE
SPECTRA OF STAR-FORMING GALAXIES

BY
REAGEN LEIMBACH

A Thesis Submitted to The Honors College
In Partial Fulfillment of the Bachelors degree
With Honors in

Astronomy

THE UNIVERSITY OF ARIZONA

MAY 2020

Approved by:

Dr. Chun Ly

Recalibrating Strong-Line Metallicity Diagnostics with Electron Temperature Measurements from Composite Spectra of Star-Forming Galaxies

Reagen Leimbach

Abstract

Galaxy evolution is driven by the accretion of gas, the birth of stars, and the outflows of enriched gas from dying stars. We can understand how these processes behave with heavy element abundance measurements. Current studies of distant galaxies utilize “strong-line diagnostics” to determine these abundances. These measurements are not accurate as they are calibrated with nearby galaxies, which differ from galaxies seen in the early universe. One method for solving this problem is to calibrate these commonly used strong-line diagnostics for distant galaxies with a “direct” metallicity indicator that is sensitive to the electron temperature, $[\text{O III}]\lambda 4363$. In my Honors thesis project, I analyzed data from the DEEP2 Survey with the goal of detecting this emission line. I explored different binning approaches to find the best technique to obtain optimal signal-to-noise for stacked spectra. Ultimately, I yielded seven detections and eight reliable limiting cases of $[\text{O III}]\lambda 4363$ out of twenty-seven composite spectra. Using these measurements, I evaluated two recently published calibrations for high-redshift galaxies. I find that one calibration developed by [Jiang et al. \[2019\]](#) yielded more reliable estimates with a difference on the “ R_{23} ” flux ratio of 0.03 dex with a dispersion of 0.07 dex. However, the calibration was not robust for low and high metallicities. Given these limitations, I plan to develop my own calibration that encompasses the broad range of metallicities seen for high-redshift galaxies.

1 INTRODUCTION

Galaxy evolution is driven by physical processes such as the accretion of gas, the birth of stars, and the outflows of gas from the deaths of stars, i.e., supernovae (e.g., [Lilly et al. 2013](#)). These processes affect the interstellar gas that resides between the stars in galaxies. One way to understand these processes is through measurements of the gas-phase heavy element abundances (or “metallicity”).

Metallicity is quantified as the mass ratio of heavy elements¹ to hydrogen. As stars progress through their lives, they form heavy elements through nuclear fusion in their cores. When the more massive stars die, producing Type II supernovae, they release heavy elements in their explosions. Less massive stars have the potential to form heavier elements and produce Type Ia supernovae if they are in a binary system, and are able to accrete enough mass to surpass the Chandrasekhar Mass limit. Both processes deposit metals into the interstellar medium. For the purpose of this study, I will be utilizing observations that are, for the most part, directly impacted by heavy elements from Type II supernova.

One method to test theoretical predictions of these processes (e.g., [Davé et al. 2011](#)), as described above, is to measure how the gas metallicity compares against other observable properties such as the stellar mass² (e.g., [Tremonti et al. 2004](#), [Erb et al. 2006](#), [Andrews & Martini 2013](#)). Oxygen is used as a proxy for heavier elements, such as iron, because it is the most abundant heavy element in the universe. Additionally, the oxygen emission lines are easily observed and detected in rest-frame optical spectra.

In order to determine the oxygen abundances, astronomers commonly use strong-line flux ratio diagnostics of nebular gas (e.g., [McGaugh 1991](#), [Pettini, & Pagel 2004](#)). One such diagnostics set is “ R_{23} ” (ratio of O^+ and O^{++} transitions to hydrogen transition) and “ O_{32} ” (ratio of O^{++} to O^+ transition) ([McGaugh 1991](#)). R_{23} is a metallicity proxy (O/H) as it uses emission from both ionic species, O^+ and O^{++} , to compare against hydrogen emission, while O_{32} examines the ionization of the gas. The latter is important as metallicity estimates are dependent on the degree of ionization.

Calibration of these strong-line metallicity diagnostics has been accepted by the astronomy com-

¹Atomic mass greater than He^4 .

²The stellar mass is defined as the total mass of stars in a galaxy.

munity for galaxies in the local universe. It has been implicitly assumed that, as we look further into space, at higher redshifts, these calibrations would hold. However, previous studies (e.g., Kewley et al. 2013, Masters et al. 2014, Shapley et al. 2015) have shown that the properties of the gas in these distant galaxies differ from those seen in galaxies in the local universe. Thus, it is unclear whether or not metallicity studies have demonstrated an evolution of the metal abundances rather than other gas properties of high redshift galaxies.

The gas metallicity is preferably calculated by measuring the flux of the $[\text{O III}]\lambda 4363$, a classically forbidden transition. This method is often referred to as the electron temperature (T_e) method (Aller 1984). This transition is populated by the collisional excitation of doubly-ionized oxygen ions by free electrons followed by the radiative emission of these ions. As illustrated in Figure 1, $[\text{O III}]\lambda 4363$ has a higher excitation energy than other $[\text{O III}]$ transitions (e.g., $[\text{O III}]\lambda 5007$). Thus, the flux ratio of $[\text{O III}]\lambda 4363$ to $[\text{O III}]\lambda 5007$ measures the excitation energy driven by the free electron temperature. T_e is set by radiative cooling, with an inverse dependence. Cooling is driven by radiative emission from heavier elements, such as oxygen, in the gas such that higher T_e corresponds to less effective cooling. Since metal-rich gas are able to cool more effectively, the T_e is lower and the $[\text{O III}]\lambda 4363$ transition line is relatively weaker. For reference, solar metallicity would yield a ratio of $<1\%$ between $[\text{O III}]\lambda 4363$ and $[\text{O III}]\lambda 5007$.

In my Honor’s thesis project, I will use spectroscopic measurements of these oxygen and hydrogen emission lines to examine existing strong-line metallicity diagnostic calibrations for high-redshift galaxies. To increase the sensitivity toward weak emission lines used for calculating metallicity, I will utilize a spectral stacking technique. This paper is outlined as follows. First, the dataset used for this study is described in Section 2. Section 3 then discusses the methodology for stacking spectra to determine the average electron temperature and metallicity for groups of galaxies in the $R_{23}\text{--}O_{32}$ plane. The organization and stacking of spectra helps increase detections of $[\text{O III}]\lambda 4363$. In this section, I also describe how I calculate electron temperature and metallicity measurements. Finally, I will present a comparison to other calibrations made in past studies in Section 4. Finally, I conclude with my results in Section 5.

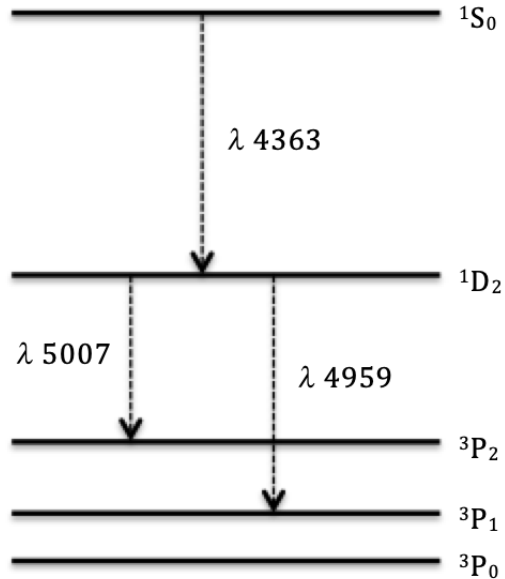


Figure 1: Energy diagram for doubly ionized oxygen, O^{++} . Illustrated here are optical emission lines detectable in H II regions: $[\text{O II}]\lambda 5007$, $[\text{O III}]\lambda 4959$, and $[\text{O III}]\lambda 4363$. Since $[\text{O III}]\lambda 4363$ is a higher excitation transition, an intensity comparison of $[\text{O III}]\lambda 4363$ against $[\text{O III}]\lambda 5007$ and $[\text{O III}]\lambda 4959$ determines the free electron temperature, T_e .

2 THE SAMPLE

This study will be using spectra from the DEEP2 Galaxy Evolution Survey (Newman et al. 2013), a spectroscopic survey of $\sim 50,000$ faint galaxies with redshifts of $z \sim 0.7\text{--}1.4$. The DEEP2 Survey was constructed using the Keck 10-m telescope with the DEep Imaging Multi-Object Spectrograph (DEIMOS; Faber et al. 2003) and surveyed four ~ 3 deg² fields. This collection of high redshift ($z \sim 1$) galaxy spectra was designed to be use in conjunction with local galaxy surveys to compare evolutionary properties of galaxies and clusters.

Following Ly et al. [2015], I first limit the initial data set to 4,140 galaxy spectra. These galaxies all have precise redshift measurements ranging from $z = 0.697\text{--}0.859$ and span the optical range of 3720–5010 Å needed to view required emission lines. To increase detections of the $[\text{O III}]\lambda 4363$ emission line, I require the initial fluxes of the $[\text{O II}]\lambda 3727$, $[\text{O III}]\lambda 4959$, $[\text{O III}]\lambda 5007$, and $\text{H}\beta$ to be greater than zero. I also required the signal-to-noise (S/N) of the $[\text{O II}]\lambda 3727$, $[\text{O III}]\lambda 4959$, $[\text{O III}]\lambda 5007$, and $\text{H}\beta$ to be greater than or equal to three. This ensures that each individual spectrum has the flux ratios used to

organize them, the R_{23} and O_{32} measurement (further explanation in Section 3.1). Finally, I require the $\log(R_{23})$ to be less than or equal to 1.4 to exclude spectra that most likely contain an active galactic nuclei (AGN). The final data set contains 2,799 galaxy spectra. Because I have accurate redshifts, all spectra are shifted to the rest frame for the stacking procedure described in Section 3.2.

Included in this set of 2,799 spectra are 28 individual spectra with detections of $[\text{O III}]\lambda 4363$, presented in Ly et al. [2015]. These individual detections are later used as another baseline for analyzing the results of this study.

3 METHODOLOGY

To perform a re-calibration of strong-line diagnostics, I use Python to detect and measure the flux of the $[\text{O III}]\lambda 4363$ emission line. The $[\text{O III}]\lambda 4363$ is a forbidden transition, which produces a weak emission line. One way to increase the signal-to-noise of this transition, and hence increase the number of detections, is to combine tens to hundreds of spectra based on a common property (a technique called “stacking”). By stacking spectra, average fluxes of the emission lines are determined based on the common property.

This study utilizes two strong-line diagnostics ratios that are commonly used in galaxy studies, R_{23} and O_{32} . The methodology, illustrated in Figure 2, is as follows. The spectroscopic sample is organized or “bin” into R_{23} – O_{32} groups as described in Section 3.1. Several different binning methods were considered to optimize the most number of detections. Next, the spectra in each bin are stacked or overlaid (see Section 3.2). Then, as discussed in Section 3.3, the flux of all emission lines can be measured by fitting the emission lines of the composite spectra with a Gaussian profile or the sum of Gaussian profiles. Finally, the electron temperature and gas metallicity are calculated from emission-line flux ratios (see Section 3.4).

3.1 BINNING IN STRONG LINE DIAGNOSTICS

The R_{23} and O_{32} diagnostics are defined as:

$$R_{23} \equiv \frac{[\text{O II}]\lambda 3727 + [\text{O III}]\lambda\lambda 4959, 5007}{\text{H}\beta}, \text{ and} \quad (1)$$

$$O_{32} \equiv \frac{[\text{O III}]\lambda\lambda 4959, 5007}{[\text{O II}]\lambda 3727}. \quad (2)$$

It is important that I organize these galaxies to find the maximum number of detections with a large

signal-to-noise confidence ($S/N \geq 3$). Thus, many different methods or algorithms were considered to optimally bin spectra in the R_{23} – O_{32} plane. For each method the average, median, maximum, and minimum R_{23} and O_{32} values are calculated for each bin, and are reported in Table 1 and 2 for two of three methods described below. These methods include: (1) a grid approach, (2) adaptive binning using an automated Voronoi tessellation algorithm, and (3) user-defined adaptive grid binning.

3.1.1 GRID METHOD

First, I considered a standard grid with a set dimensions so that each bin was geometrically the same size (see Figure 3). I considered many different bin sizes and in the end adopted 0.25 dex in $\log(R_{23})$ and $\log(O_{32})$. However, the number of galaxies varied in each bin ranging from 12 spectra in the low density bins to 340 or more spectra in the densest bins.

This method was not ideal because it did not consider the range in density of galaxies on the R_{23} – O_{32} plane. Thus, it resulted in eight stacks with detections and the rest as non-detections. Of those eight detections, only five were high signal-to-noise ($S/N \geq 5$) detections. In order to maximize the number of reliable detections, the bin size must vary such that higher density regions should be split up into smaller bins. This adaptive binning approach is discussed below.

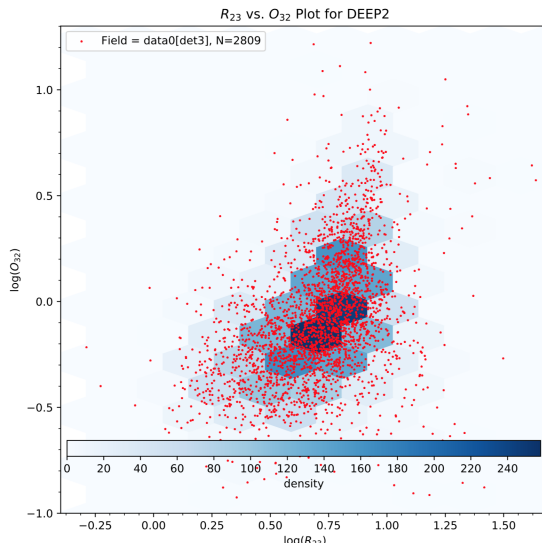


Figure 3: Illustration of binning in the R_{23} – O_{32} plane using a fixed grid method with sizes of 0.25 dex. Red points illustrate individual galaxies with a density distribution shown by the blue colormap.

Steps for Data Analysis Process

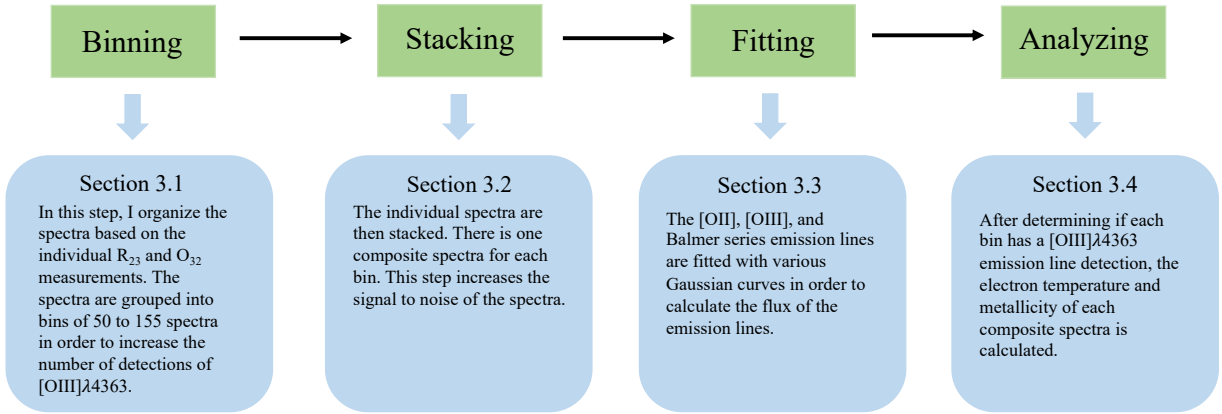


Figure 2: Flow chart to illustrate methodology for determining gas metallicity using the [O III] λ 4363 emission line with composite DEEP2 spectra.

3.1.2 ADAPTIVE BINNING USING THE VORONOI TESSELLATION ALGORITHM

To account for the density of galaxies in the R_{23} – O_{32} plane, I implemented the adaptive binning method of Cappellari & Copin [2003], which uses a Voronoi tessellation algorithm. This adaptive binning method starts at the area of highest density. It determines the number of measurements to bin based on a specified target signal-to-noise value, TARGET_SN. The design of their algorithm was to consider the reliability of any measurement such that regions of high signal-to-noise could be adaptively binned further while regions of lower signal-to-noise would require more measurements to bin. This binning method sums up in quadrature the signal-to-noise of each individual measurement as it is included in a bin. When the summed signal-to-noise equals the inputted target signal-to-noise, the binning method starts a new bin. In Cappellari & Copin [2003] work, they applied their algorithm toward two-dimensional data of extended galaxies. Here in the spectral stacking analysis, I apply this toward the distribution on the R_{23} – O_{32} plane, and treat each measurement (i.e., spectra) with uniform weights of unity. By adopting such a weighting scheme, the TARGET_SN is dependent on the square root of the number of spectra in each stack, $\sqrt{N_{\text{stack}}}$. I illustrate the application of the Voronoi method in the R_{23} – O_{32} plane in Figure 4.

Similar to the grid method, this method yielded detections for the dense regions on the R_{23} – O_{32} plane. It also yielded more composite spectra with detections because the denser regions were divided into additional bins. I found the optimal TARGET_SN

to be 14 ($N_{\text{stack}} = 200$), which yielded the most number of stacked spectra with [O III] λ 4363 detections. This yielded bins ranging in size from N_{stack} of 14 to 711 spectra with an average of 150. The properties of each bin—minimum, maximum, average, and median R_{23} and O_{32} values—are given in Table 1.

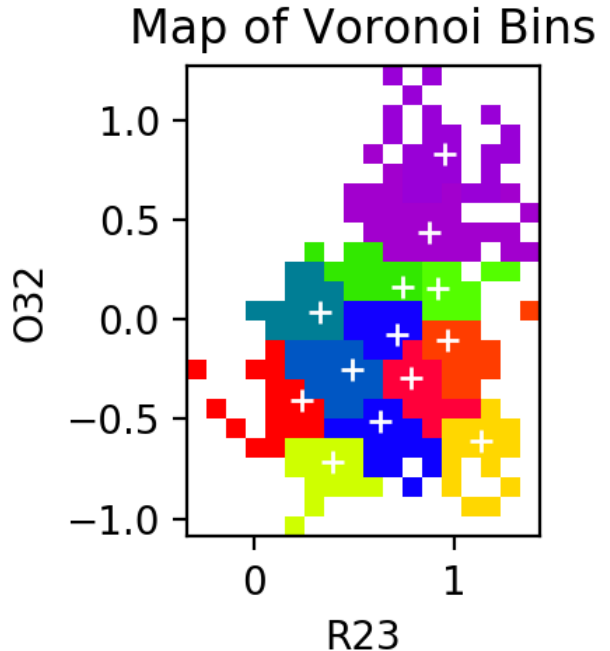


Figure 4: Illustration of the Voronoi tessellation algorithm in the R_{23} – O_{32} plane. Each bin is distinguished by different colors with a white cross indicating the densest region of the bin.

3.1.3 USER-DEFINED ADAPTIVE GRID BINNING

The current binning method used in this study, which is illustrated in Figure 5, is a hybrid of the two methods described above. Here I develop a custom grid that is adapted to (1) the density of spectra in the R_{23} – O_{32} plane, and (2) the likelihood that the composite spectra will have an [O III] λ 4363 detection. Specifically, bins that contain lower metallicity galaxies will require *fewer* spectra to reach the minimum signal-to-noise ($S/N = 3$) on [O III] λ 4363. First, galaxies are binned based on their R_{23} values.

Here, I vary the number of spectra in each R_{23} bin. As R_{23} increases, the number of spectra needed for a detection decreases. After considering different bin sizes, I adopted N_{stack} of 458, 450, 400, 300, 300, 275, 250, 200, and 176 (with increasing R_{23}).

Next, each R_{23} bin is split based on the O_{32} values of the spectra. Initially, this split occurred at the median O_{32} value for each bin. However, I found that this was not optimal because the O_{32} range was wide such that the lower O_{32} bin would generally result in a non-detection. After some evaluation with a focus on robust sample sizes in each R_{23} – O_{32} bin to yield a maximum number of detections, I determined that splitting each R_{23} bin equally into three sub-bins is ideal. The properties of each bin—the minimum, maximum, average, and median R_{23} and O_{32} values—are given in Table 2.

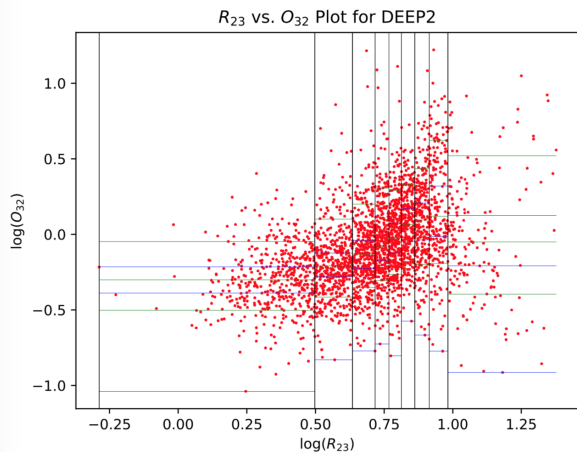


Figure 5: First, the spectra are organized by R_{23} (black lines). Then, each R_{23} bin is split into three sub bins by the O_{32} values. In the figure above, the blue lines mark the bottom of each bin and the green lines marks the average O_{32} value of each bin. I will refer to this as Manual Binning.

3.2 SPECTRAL STACKING

The rest-frame spectra in each bin are then stacked such that the composite spectrum is an average for all spectra in a given bin, as demonstrated in Figure 6. Stacking decreases random variations or noise across the spectra, which ultimately increases the signal-to-noise to detect weak lines. In the stacking process, I mask regions of the observed spectra for night skyline emission to eliminate contamination.³

3.3 FITTING EMISSION LINES

Once the spectra are stacked, the flux of the emission lines are determined by fitting Gaussian profiles using `scipy.optimize.curve_fit`. The fitted lines include [O II] λ 3727, [O III] λ 4363, $H\beta$, [O III] λ 4959, and [O III] λ 5007. Depending on the transition, either a single or double Gaussian profile is used in the fitting. For example, the [O III] lines can be fitted with a single Gaussian profile (Figure 7A). However, Balmer lines such as $H\beta$, must be approximated with the combination of two Gaussian lines, one positive and one negative (Figure 7B). The latter is for stellar absorption. The [O II] emission requires two positive Gaussian peaks to approximate the emission at 3726.16 and 3728.91 Å (Figure 7C).

The `curve_fit` tool allows for a range of inputted values for the parameters to best optimize the fitting. In order to fit the spectra, I first normalized them to 10^{-17} erg s $^{-1}$ cm $^{-2}$ Å $^{-1}$. This normalization allows for a large dynamic range (by a factor of 10^{17}) in the fitting parameters, making `curve_fit` more accurate. Inputs for `curve_fit` include:

- A line profile that matches the expected shape of the emission line
- The range of wavelengths and fluxes for the composite spectrum fluxes where the emission line(s) is expected
- Initial guesses for the fitting parameters

The initial parameter values include:

1. The central wavelength of the emission line(s)
2. Expected width of the emission line(s)
3. Maximum amplitude
4. Continuum level

I determined the range that the initial guesses are allowed to vary through trial and error.

³This night skyline masking code was created by Rafia Bushra, a member of our research team.

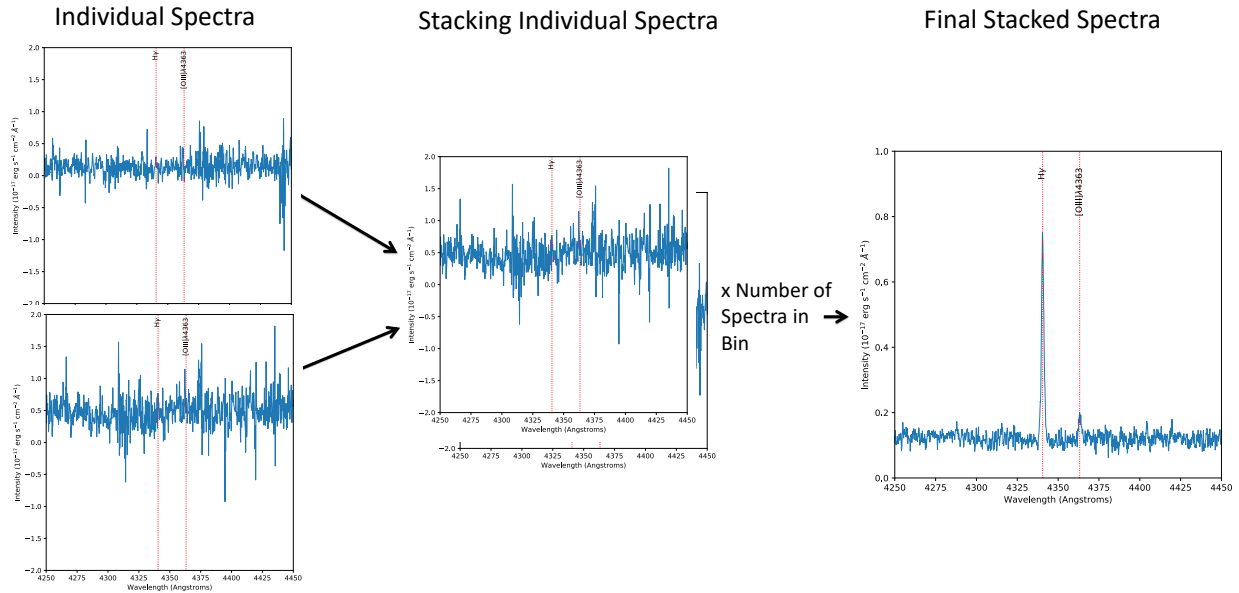


Figure 6: Illustration of the spectral stacking procedure. The spectra (left) in each bin are overlaid and stacked (middle) using an averaging approach. This produces a composite spectrum (right) with lower noise to enable detection of the weak $[O\ III]\lambda 4363$ line. The right composite spectrum, which is a stack of 83 individual spectra, yielded a S/N of 8.25 on $[O\ III]\lambda 4363$.

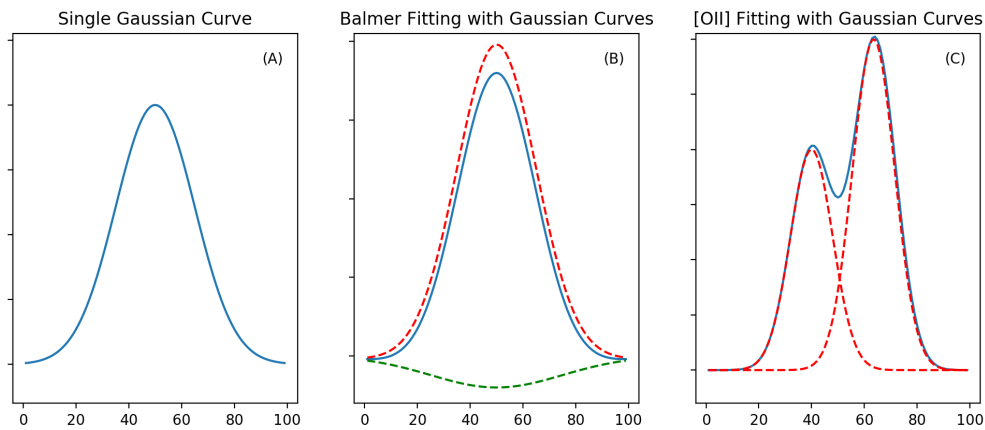


Figure 7: Illustration of different Gaussian curves used to fit different emission lines: (A) single emission line (e.g., the $[O\ III]$ lines); (B) single emission (dashed red) and absorption (dashed green) lines for Balmer lines (e.g., $H\beta$); and (C) double-peak emission lines (dashed red) to fit the $[O\ II]$ lines. The final profiles are shown by the solid blue lines.

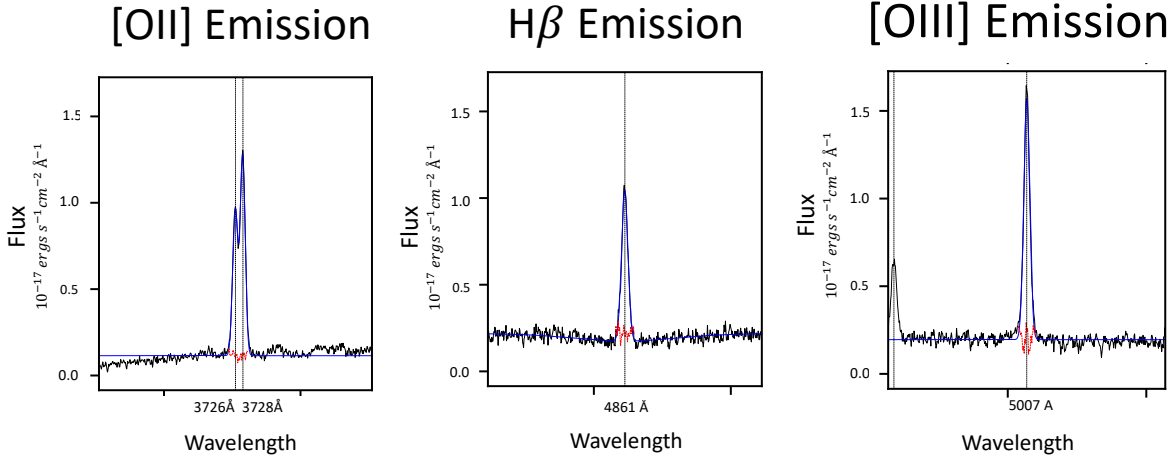


Figure 8: Illustration of Gaussian fitting (blue lines) for three different emission lines in a stacked spectrum (black line): [O II] λ 3727 (left), H β (center), and [O III] λ 5007 (right). The red line indicates the difference between the spectrum and the fit.

An illustration of these `curve_fit` fits is provided in Figure 8 for one of the composite spectra. Once the optimal fits were found, the line fluxes were calculated by integrating the curve fits.

3.4 DERIVING TEMPERATURES AND METALLICITIES

Once the fitting is complete, I can calculate the free electron temperature (T_e) and oxygen abundance (O/H) for the composite spectra. Oxygen emission lines detectable in the optical are illustrated in Figure 1. The [O III] λ 4363 line is the transition from the 1S_0 to the 1D_2 state, and is a higher excitation transition than [O III] λ 4959 and [O III] λ 5007, which are the transitions from 1D_2 to the 3P_1 and 3P_2 states, respectively. Oxygen is the dominant heavy element in H II regions and therefore, oxygen abundance is used as a proxy for the total metallicity. First, the ratio of [O III] λ 4363 to the combination of the two other [O III] emission lines, [O III] λ 4959 and [O III] λ 5007, is determined:

$$R \equiv \frac{[\text{O III}]\lambda 4363}{[\text{O III}]\lambda 4959 + [\text{O III}]\lambda 5007}. \quad (3)$$

Next, following the work described in Nicholls et al. [2013], the electron temperature is calculated.

$$T_e = a(-\log_{10}(R) - b)^{-c}. \quad (4)$$

For O^{++} , I use $a = 13205$, $b = 0.92506$, $c = 0.98062$. These values were derived by Nicholls et al. [2014]

using the Mappings IV photoionization code. This method is a simplified approximation, derived in Nicholls et al. [2013], that fits a power law to the curves of the computed flux ratio versus the equilibrium temperature from the Maxwell Boltzmann temperature distributions. The approximation is within 0.5% of the fully calculated value.

Following Izotov et al. [2006], the gas metallicity is derived from T_e and the flux ratios [O II]/H β and [O III]/H β with the following equations:

$$12 + \log\left(\frac{\text{O}^+}{\text{H}}\right) = \log\left(\frac{[\text{O II}]}{\text{H}\beta}\right) + 5.961 + \frac{1.676}{t_2} - 0.4 \log(t_2) - 0.034t_2 + \log(1 + 1.35x) \quad (5)$$

$$12 + \log\left(\frac{\text{O}^{++}}{\text{H}}\right) = \log\left(\frac{[\text{O III}]}{\text{H}\beta}\right) + 6.200 + \frac{1.251}{t_3} - 0.55 \log(t_3) - 0.014t_3. \quad (6)$$

Here, t_3 , t_2 , and x are dependent on the electron temperature:

$$t_3 = 10^{-4} \cdot T_e, \quad (7)$$

$$t_2 = 0.7 \cdot t_3 + 0.17, \text{ and} \quad (8)$$

$$x = 10^{-4} \cdot n_e \cdot t_2^{-0.5}, \quad (9)$$

where n_e is the O^+ electron density. In general, n_e is not higher than 10^3 cm^{-3} . For simplicity, this study adopts $n_e = 10^3 \text{ cm}^{-3}$ because the effect of n_e is small compared to other terms in the equation such

that it does not greatly affect the metallicity estimates. Finally, the t_2 – t_3 relation that is adopted is from [Andrews & Martini \[2013\]](#). By solving for O^+/H and O^{++}/H and combining them together, the total oxygen abundance is determined:

$$\frac{O}{H} = \frac{O^+}{H} + \frac{O^{++}}{H}. \quad (10)$$

The User-defined Adaptive Binning approach yielded seven detections of the $[O\text{ III}]\lambda 4363$ emission line. While most composite spectra do not have an $[O\text{ III}]\lambda 4363$ detection ($S/N \geq 3$), the spectra may be reliable enough to derive an upper limit on $[O\text{ III}]\lambda 4363$ (hence an upper limit on T_e and a lower limit on metallicity) or robust limits. To identify such spectra, I consider those with $S/N \geq 100$ for $[O\text{ III}]\lambda 5007$. This limit was set to ensure a robust measurement while still keeping it conservative to increase the sample of “reliable $[O\text{ III}]\lambda 4363$ non-detections.” Generally, the fluctuations across the composite spectra are roughly the same. Thus a $S/N=3$ upper limit on $[O\text{ III}]\lambda 4363$ with a $S/N=100$ on $[O\text{ III}]\lambda 5007$ would correspond to approximately 0.03 for the R ratio. The $[O\text{ III}]\lambda 4363$ flux limit is calculated based on the adjacent $H\gamma$ Balmer emission line:

$$F([O\text{ III}]\lambda 4363) = 3 \cdot \sigma_{H\gamma}, \quad (11)$$

where $\sigma(H\gamma)$ is the measured rms in the composite spectra around $H\gamma$. Here, the rms is measured based on the line width of $H\gamma$; it can be calculated in a straight-forward manner from the $H\gamma$ flux and the S/N on $H\gamma$. This yielded an additional eight composite spectra with robust $[O\text{ III}]\lambda 4363$ non-detections. I provide the electron temperatures, metallicities, and emission-line flux ratios in [Table 3](#) for the User-defined Adaptive Binning approach described in [Section 3.1.3](#).

4 RESULTS

With more robust metallicity measurements using the T_e method, I can examine the reliability of existing strong-line calibrations. Below I describe the main results from such comparisons for recent calibrations that are believed to describe high-redshift galaxies: [Bian et al. \[2018\]](#) ([Section 4.1](#)) and [Jiang et al. \[2019\]](#) ([Section 4.2](#)).

4.1 THE LOCAL ANALOG CALIBRATION

Rest-frame optical spectra of high redshift galaxies are difficult to obtain, which increases the challenge

of detecting the $[O\text{ III}]\lambda 4363$ weak emission line. To address this deficiency, [Bian et al. \[2018\]](#) identified a sample of local galaxies that are analogous to high-redshift galaxies from the Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7). This sample consists of star-forming galaxies that are located in the same region of the Baldwin–Phillips–Terlevich (BPT) diagram as typical $z \approx 2$ galaxies. Galaxies are classified as star-forming using the BPT diagram. Each galaxy is selected to have detections ($S/N > 10$) of $[O\text{ II}]\lambda 3727$, $H\beta$, $[O\text{ III}]\lambda\lambda 4959, 5007$, $H\alpha$ and $[N\text{ II}]\lambda 6583$.

To detect the weak $[O\text{ III}]\lambda 4363$ emission line, [Bian et al. \[2018\]](#) adopted a stacking method similar to my approach described in [Section 3.3](#). However, instead of binning in the R_{23} – O_{32} plane, their composite spectra are based on the $[N\text{ II}]\lambda 6584/H\alpha$ flux ratio. There is a direct inverse correlation between the $[N\text{ II}]\lambda 6584/H\alpha$ ratio and O_{32} ratio. Their analysis yielded 14 stacked spectra; seven of which consist of high-redshift local analogs. The other seven are identified as a reference sample of local typical galaxies. To determine electron temperatures and oxygen abundances, [Bian et al. \[2018\]](#) used the relations of [Izotov et al. \[2006\]](#) and include, in their calculation, estimates for n_e from the $[S\text{ II}]\lambda\lambda 6717, 6731$ doublet ratio. They are unable to calculate an electron temperature for singly-ionized oxygen so they adopt the following relationship between doubly- and singly-ionized oxygen from [Garnett, D. R. \[1992\]](#), $t_2 = 0.7 \cdot t_3 + 0.3$. Using the composite spectral measurements for the local analog sample, [Bian et al. \[2018\]](#) derived the following equations for a relationship between metallicity and $\log(O_{32})$ and metallicity and $\log(R_{23})$, hereafter the Local Analog Calibration (LAC):

$$\begin{aligned} 12 + \log(O/H) &= 8.54 - 0.59 \log(O_{32}), \text{ and} \\ \log(R_{23}) &= a + bx + cx^2 - dx^3 \end{aligned} \quad (12)$$

where $x \equiv 12 + \log O/H$, $a = 138.0430$, $b = -54.828$, $c = 7.2954$, and $d = 0.32293$.

I illustrate these relations in [Figure 9](#) against my composite measurements. Also in this illustration, I include individual T_e -based metallicity measurements from the DEEP2 ([Ly et al. 2015](#)) and $\mathcal{M}ACT$ ([Ly et al. 2016](#)) samples. As seen in [Figure 9](#), the LAC R_{23} calibration agrees with most of the composite and individual spectra for metallicity values ranging from $7.8 < 12 + \log(O/H) < 8.4$. However, the LAC sample consists of galaxies at intermediate metallicities. Thus, any estimates computed for higher or lower metallicity is an extrapolation. I also examined the LAC O_{32} calibration illustrated in [Figure 9](#), and conclude that it is not a reliable indicator

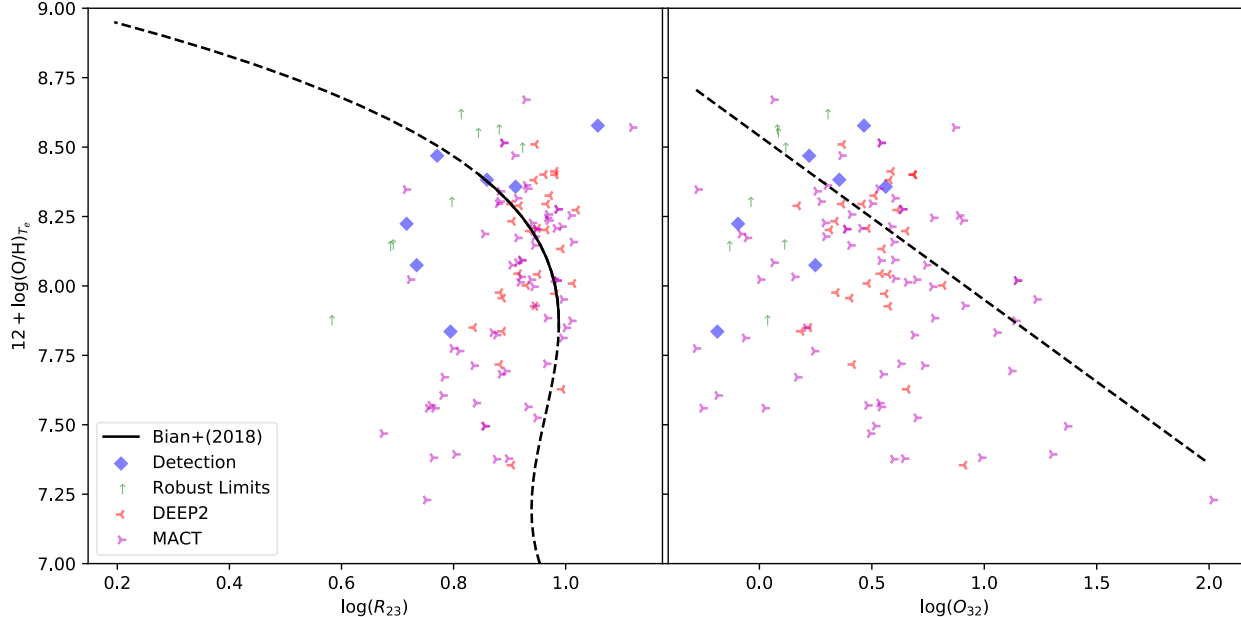


Figure 9: Illustration of the LAC defined by [Bian et al. \[2018\]](#). Dashed lines illustrate the relationship between metallicity and $\log(R_{23})$ (left) and $\log(O_{32})$ (right). The solid line indicates where our measurements and the calibration have a correlation. Composite measurements with $[\text{O III}]\lambda 4363$ detections and robust non-detections are shown as filled blue diamonds and green up arrows, respectively. Individual measurements from DEEP2 and *MACT* are shown by red “tri left” and purple “tri right” markers, respectively.

of metallicity. This is most likely a result of the fact that O_{32} is an ionization indicator.

4.2 THE GREEN PEAS CALIBRATION

In order to detect the weak $[\text{O III}]\lambda 4363$ emission line, spectra must have strong $[\text{O III}]$ emission lines. Green pea galaxies, first noted by volunteers in the Galaxy Zoo project, are dwarf galaxies with strong $[\text{O III}]$ emission lines. [Jiang et al. \[2019\]](#) argued that these galaxies can be used to construct a metallicity calibration for higher redshift galaxies because of their similar properties, such as low metallicity ($7.5 < 12 + \log(\text{O}/\text{H}) < 8.5$) and strong R_{23} and O_{32} flux ratios. The [Jiang et al. \[2019\]](#) calibration, hereafter the Green Peas Calibration (GPC), was derived from a sample of ≈ 800 green pea galaxy spectra with redshifts of $z = 0.011\text{--}0.411$. This sample was assembled from the SDSS DR13 ([Albareti et al. 2017](#)). The green pea galaxies used in the GPC had $[\text{O III}]\lambda 4363$ detections and thus T_e -based metallicity estimates.

In deriving electron temperature and metallicity, [Jiang et al. \[2019\]](#) used equations from [Izotov et al. \[2006\]](#).⁴ In addition, they use the $t_2\text{--}t_3$ relation from [Izotov et al. \[2006\]](#) and the $[\text{S II}]$ flux ratios to determine n_e , which varied between 10 cm^{-3} and 10^4

cm^{-3} .

Unlike the LAC (see Section 4.1), the GPC takes into account other physical parameters such as the ionizing radiation with the O_{32} flux ratio. This approach is similar to other studies (e.g., [Pilyugin 2000, 2001a,b](#)). Using a least squares fitting for 789 galaxies, [Jiang et al. \[2019\]](#) find a functional form of

$$\log(R_{23}) = a + bx + cx^2 - d(e + x)y, \quad (14)$$

where $x \equiv 12 + \log(\text{O}/\text{H})$, $y \equiv \log(O_{32})$, $a = -24.135$, $b = 6.1532$, $c = -0.37866$, $d = -0.147$, and $e = -7.071$.

I illustrate, in Figure 10, a comparison of GPC with my composite spectra measurements and individual measurements from the DEEP2 ([Ly et al. 2015](#)) and *MACT* ([Ly et al. 2016](#)) samples. Using the observed $12 + \log(\text{O}/\text{H})$ and $\log(O_{32})$, I derive estimates for $\log(R_{23})$ values from the GPC and compare them against the observed $\log(R_{23})$. This quantitative comparison is illustrated in Figure 11. Overall, I find a small statistical average offset of 0.03 dex with a dispersion of 0.07 dex.

My results on the GPC is similar to those for the LAC with some differences. First, the GPC holds for a specific range of intermediate metallicities

⁴This approach differs slightly from my work since I used a different relation to derive T_e (see Equation 4).

($7.7 < 12 + \log(\text{O}/\text{H}) < 8.5$), corresponding to the range of metallicities for the Jiang et al. [2019] data set. This is seen in Figures 10–11 with data points indicated by the green, cyan, and blue colors. Second, the calibration does not hold for galaxies with lower or higher $12 + \log(\text{O}/\text{H})$ measurements, as seen in the bottom panel of Figure 11. In the GPC sample, Jiang et al. [2019] noted a subset of 15 galaxies that have extremely low metallicities ($12 + \log(\text{O}/\text{H}) < 7.6$). Although there is a loose correlation between the GPC curves and data points at lower metallicities, the samples sizes are too small such that the GPC is robust for $12 + \log(\text{O}/\text{H}) < 7.5$.

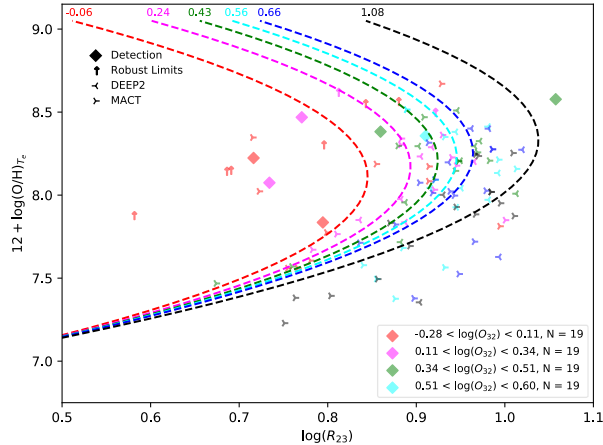


Figure 10: Illustration of the GPC defined by Jiang et al. [2019]. Dashed lines illustrate the relationship between T_e -based metallicity and $\log(R_{23})$ for constant $\log(O_{32})$. The composite $[\text{O III}]\lambda 4363$ detections and robust non-detections are shown by the blue filled diamonds and green filled diamonds, respectively. Individual measurements from DEEP2 and MACT are shown by the “tri left” and “tri right” markers, respectively.

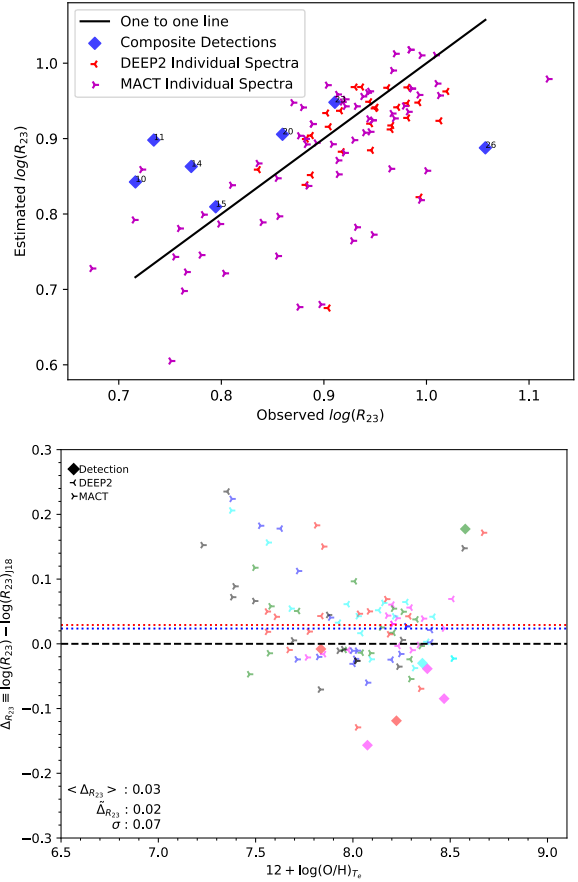


Figure 11: Top: Comparison between observed and estimated $\log(R_{23})$ using the GPC. Bottom: Illustration of the difference in $\log(R_{23})$ as a function of metallicity. The average and median of the difference are illustrated by the red and blue dashed lines, respectively. For both panels, the composite spectra are illustrated by the diamond points while individual measurements from DEEP2 and MACT are illustrated by the “tri left” and “tri right” markers, respectively. The black line is the one-to-one relation.

5 CONCLUSIONS

To study galaxy evolution, extragalactic studies use metallicity calibrations to understand the interstellar gas in galaxies. These calibrations are often based on strong-line diagnostics of the nebular gas, such as the R_{23} and O_{32} flux ratios. In the past, calibrations have been formalized using strong emission line measurements in nearby galaxies, and it has been assumed that these calibrations would hold at higher redshifts. However, subsequent studies have shown that this assumption is incorrect. To address this issue, this study utilized data from the DEEP2 Galaxy

Evolution Survey to examine existing calibrations for high-redshift galaxies.

One way to determine the gas-phase metallicity is to measure the electron temperature of the gas from the [O III] λ 4363 emission line. However, the [O III] λ 4363 line is a forbidden weak transition, often difficult to detect in individual spectra. To obtain detections, I utilize a spectral stacking method to combine spectra after organizing them by their R_{23} and O_{32} measurements. Several binning techniques were considered. Ultimately I determined that a user-defined adaptive binning approach, where a desired number of galaxies is determined based on the R_{23} and O_{32} measurements, was more effective at yielding composite spectra with detections.

Using these composite spectra, along with individual spectra with [O III] λ 4363 detections from Ly et al. [2015] and Ly et al. [2016], I evaluate two high-redshift metallicity calibrations that were recently published. The first, the Local Analog Calibration (LAC; Bian et al. 2018), was effective at predicting intermediate metallicities using measured R_{23} . However, their O_{32} calibration was less accurate in comparison. This is not surprising as O_{32} was considered a proxy for gas ionization and not gas metallicity. However, because the LAC was based on galaxies at intermediate metallicities, $7.8 < 12 + \log(\text{O}/\text{H}) < 8.4$, it was not reliable for high ($12 + \log(\text{O}/\text{H}) \geq 8.4$) or lower metallicities ($12 + \log(\text{O}/\text{H}) \leq 7.8$). The second, the Green Peas Calibration (GPC; Jiang et al. 2019), is also effective for an intermediate range of metallicity, $7.5 < 12 + \log(\text{O}/\text{H}) < 8.4$. This calibration is more accurate than the LAC with a statistical offset of 0.03 dex and a dispersion of 0.07 dex. However, the GPC is also not accurate for high and low metallicities. Based on my analysis, I recommend using the GPC.

To complete this study, I plan to construct a calibration based on individual and composite measurements for $z \sim 0.8$ galaxies, study the effects of dust attenuation, and provide a comprehensive error analysis budget. I also plan to use the spectral stacking method developed in this work to analyze the *MACT* data set. This will expand my data set to improve the calibration that I develop. It is my hope to develop a calibration that fits low, intermediate, and high metallicity ranges to improve our study of high-redshift galaxies.

ACKNOWLEDGEMENTS

The data presented herein were obtained at the W. M. Keck Observatory, which is operated as a sci-

entific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. I would like to thank NASA Space Grant Consortium for the support for the 2018–2019 academic school year. I would like to thank my thesis project advisor, Dr. Chun Ly, for his mentorship and guidance throughout this process. I would also like to thank my academic advisors and professors for their contributions to my academic progress. Finally, I would like to thank my family for their support. This research made use of the following Python packages: *astropy* (Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *Astropy Collaboration, Price-Whelan, et al. 2018*), *matplotlib* (Hunter, J. D. 2007), *numpy* (Oliphant, T. E. 2006), and *scipy* (Jones, E. et al. 2001).

REFERENCES

- Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, *ApJS*, 233, 25
- Aller, L. H. 1984, *Astrophysics and Space Science Library*
- Andrews, B. H., & Martini, P. 2013, *ApJ*, 765, 140
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- Astropy Collaboration, Price-Whelan, A. M., Sip \tilde{A} tcz, B. M., et al. 2018, *ApJ*, 156, 12
- Bian, F., Kewley, L. J., & Dopita, M. A. 2018, *ApJ*, 859, 175
- Cappellari, M., & Copin, Y. 2003, *MNRAS*, 342, 345
- Davé, R., Finlator, K., & Oppenheimer, B. D. 2011, *MNRAS*, 416, 1354
- Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, *ApJ*, 644, 813
- Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, *Instrument Design and Performance for Optical/infrared Ground-based Telescopes*, 1657
- Garnett, D. R. 1992, *ApJ*, 103, 1330

Table 1: Summary of Adaptive Voronoi R_{23} - O_{32} Bins

Bin ID	N_{stack}	Minimum		Maximum		Average		Median	
		$\log(R_{23})$	$\log(O_{32})$	$\log(R_{23})$	$\log(O_{32})$	$\log(R_{23})$	$\log(O_{32})$	$\log(R_{23})$	$\log(O_{32})$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	711	0.54	-0.27	0.94	0.11	0.75	-0.08	0.74	-0.09
1	468	0.65	0.02	1.37	0.39	0.84	0.18	0.82	0.17
2	152	0.83	-0.57	1.50	-0.11	0.99	-0.26	0.95	-0.26
3	105	0.89	-0.10	1.17	0.15	0.99	0.01	0.97	0.01
4	56	0.29	0.10	0.68	0.40	0.58	0.24	0.58	0.24
5	14	1.02	-0.91	1.42	-0.58	1.22	-0.71	1.20	-0.66
6	556	0.39	-1.54	1.59	-0.16	0.64	-0.32	0.61	-0.32
7	222	-0.02	-0.28	0.56	0.25	0.43	-0.09	0.43	-0.12
8	88	-0.29	-0.43	0.38	-0.17	0.24	-0.30	0.25	-0.31
9	62	0.51	-0.83	0.96	-0.50	0.73	-0.61	0.71	-0.61
10	131	-0.08	-1.04	0.50	-0.37	0.34	-0.52	0.35	-0.49
11	206	0.52	0.30	1.90	0.95	0.96	0.50	0.90	0.46
12	38	0.69	0.70	1.35	1.22	0.97	0.89	0.93	0.85

(1): Adaptive Voronoi Bin ID. (2): Number of stacked spectra, N_{stack} . (3)–(10): Minimum, maximum, average, and median of $\log(R_{23})$ and $\log(O_{32})$ flux ratios.

Hunter, J. D. 2007, *Computing in science & engineering*, 9, 90

Izotov, Y. I., Stasińska, G., Meynet, G., et al. 2006, *A&A*, 448, 955

Jiang, T., Malhotra, S., Rhoads, J. E., et al. 2019, *ApJ*, 872, 145

Jones, E., Oliphant, T., Peterson, P., et al. 2001, <http://www.scipy.org/>

Kewley, L. J., Maier, C., Yabe, K., et al. 2013, *ApJ*, 774, L10

Lilly, S. J., Carollo, C. M., Pipino, A., et al. 2013, *ApJ*, 772, 119

Ly, C., Malhotra, S., Malkan, M. A., et al. 2016, *ApJS*, 226, 5

Ly, C., Rigby, J. R., Cooper, M., et al. 2015, *ApJ*, 805, 45

Masters, D., McCarthy, P., Siana, B., et al. 2014, *ApJ*, 785, 153

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

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5

Nicholls, D. C., Dopita, M. A., Sutherland, R. S., et al. 2013, *ApJS*, 207, 21

Nicholls, D. C., Dopita, M. A., Sutherland, R. S., et al. 2014, *ApJ*, 790, 75

Oliphant, T. E. 2006 *A guide to NumPy*, Vol. 1 (Trelgol Publishing USA)

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

Pilyugin, L. S. 2000, *A&A*, 362, 325

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

Pilyugin, L. S. 2001b, *A&A*, 374, 412

Shapley, A. E., Reddy, N. A., Kriek, M., et al. 2015, *ApJ*, 801, 88

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

Table 2: Summary of User-defined Adaptive R_{23} – O_{32} Bins

Bin ID	N_{stack}	Minimum		Maximum		Average		Median	
		$\log(R_{23})$	$\log(O_{32})$	$\log(R_{23})$	$\log(O_{32})$	$\log(R_{23})$	$\log(O_{32})$	$\log(R_{23})$	$\log(O_{32})$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
00	152	-0.29	-1.04	0.50	-0.39	0.35	-0.50	0.35	-0.48
01	153	-0.29	-0.39	0.50	-0.22	0.37	-0.30	0.38	-0.31
02	153	-0.29	-0.21	0.49	0.40	0.38	-0.05	0.39	-0.10
03	150	0.50	-0.83	0.63	-0.28	0.56	-0.42	0.56	-0.40
04	151	0.50	-0.28	0.63	-0.12	0.57	-0.21	0.58	-0.21
05	149	0.50	-0.12	0.63	0.86	0.57	0.10	0.57	-0.02
06	133	0.63	-0.77	0.72	-0.23	0.68	-0.35	0.68	-0.33
07	134	0.63	-0.23	0.72	-0.04	0.68	-0.13	0.68	-0.13
08	133	0.63	-0.04	0.72	1.22	0.68	0.20	0.69	0.08
09	100	0.72	-0.73	0.77	-0.18	0.74	-0.31	0.74	-0.28
10	101	0.72	-0.18	0.77	0.02	0.74	-0.08	0.74	-0.08
11	99	0.72	0.02	0.77	1.09	0.74	0.30	0.74	0.20
12	100	0.77	-0.80	0.81	-0.12	0.79	-0.26	0.79	-0.24
13	101	0.77	-0.12	0.81	0.07	0.79	-0.02	0.79	-0.03
14	99	0.77	0.07	0.81	1.11	0.79	0.29	0.79	0.20
15	91	0.81	-0.57	0.86	-0.05	0.84	-0.18	0.83	-0.17
16	92	0.81	-0.05	0.86	0.17	0.84	0.06	0.84	0.06
17	92	0.81	0.17	0.86	0.73	0.84	0.33	0.84	0.29
18	83	0.86	-0.69	0.91	-0.04	0.89	-0.20	0.89	-0.18
19	84	0.86	-0.04	0.91	0.20	0.88	0.10	0.88	0.09
20	83	0.86	0.21	0.91	1.08	0.89	0.43	0.89	0.34
21	66	0.92	-0.77	0.98	-0.02	0.95	-0.19	0.95	-0.16
22	67	0.92	-0.02	0.98	0.32	0.94	0.12	0.94	0.09
23	67	0.92	0.32	0.98	1.22	0.95	0.62	0.94	0.55
24	55	0.98	-0.91	1.33	-0.22	1.12	-0.40	1.11	-0.38
25	56	0.98	-0.21	1.37	0.13	1.06	-0.04	1.04	-0.05
26	55	0.98	0.14	1.38	1.05	1.12	0.53	1.05	0.37

(1): User-defined Adaptive Bin ID. (2): Number of stacked spectra, N_{stack} . (3)–(10): Minimum, maximum, average, and median of $\log(R_{23})$ and $\log(O_{32})$ flux ratios.

Table 3: Summary of Emission-line Flux Ratios, Electron Temperature, and Metallicity Measurements for User-Defined Adaptive R_{23} - O_{32} Bins

Bin ID (1)	Detection (2)	R (3)	$[\text{O II}]/\text{H}\beta$ (4)	$[\text{O III}]/\text{H}\beta$ (5)	$\log(T_e/\text{K})$ (6)	$12 + \log(\text{O}/\text{H})$ (7)	$\log(\text{O}^+/\text{H})$ (8)	$\log(\text{O}^{++}/\text{H})$ (9)
05	U	0.012	1.837	1.958	4.128	7.88	-4.27	-4.67
07	U	0.010	2.803	2.027	4.091	8.14	-3.96	-4.55
08	U	0.009	2.151	2.729	4.074	8.15	-4.01	-4.37
10	Y	0.009	2.884	2.305	4.074	8.22	-3.88	-4.44
11	Y	0.010	1.955	3.448	4.098	8.08	-4.14	-4.38
13	U	0.009	3.273	2.943	4.071	8.30	-3.82	-4.32
14	Y	0.006	2.213	3.664	4.000	8.47	-3.71	-4.00
15	Y	0.020	3.775	2.441	4.227	7.84	-4.27	-4.82
16	U	0.006	3.162	3.764	4.007	8.55	-3.59	-4.01
17	U	0.004	2.165	4.285	3.966	8.62	-3.58	-3.81
19	U	0.006	3.464	4.083	4.014	8.56	-3.58	-4.00
20	Y	0.007	2.216	4.994	4.037	8.38	-3.86	-3.99
22	U	0.007	3.635	4.668	4.040	8.50	-3.66	-4.03
23	Y	0.008	1.752	6.354	4.044	8.36	-3.99	-3.91
26	Y	0.007	2.913	8.453	4.030	8.58	-3.71	-3.74

(1): User-defined Adaptive Bin ID. (2): $[\text{O III}]\lambda 4363$ detection (Y="Yes"; U="Upper limit"). (3): Oxygen flux ratio between $[\text{O III}]\lambda 4363$ and $[\text{O III}]\lambda\lambda 4959, 5007$, see Equation (3). (4): $[\text{O II}]\lambda 3727/\text{H}\beta$ flux ratio. (5): $[\text{O III}]\lambda\lambda 4959, 5007/\text{H}\beta$ flux ratio. (6): Logarithm of electron temperature. (7): Logarithm of total oxygen abundance. (8)–(9): Logarithm of singly- and doubly-ionized oxygen-to-hydrogen abundance ratios.