

The SAMI Galaxy Survey: reconciling strong emission line metallicity diagnostics using metallicity gradients

Henry Poetrodjojo¹,^{2,3}★ Brent Groves^{1,2,4} Lisa J. Kewley^{1,2} Sarah M. Sweet^{1,2,5,6}
 Sebastian F. Sanchez⁷ Anne M. Medling⁸ Ángel R. López-Sánchez^{9,10} Sarah Brough¹¹
 Luca Cortese¹²,^{2,4} Jesse van de Sande¹³,^{2,3} Sam Vaughan¹³,^{2,3} Samuel N. Richards¹²,¹²
 Julia J. Bryant^{2,3,13} Scott M. Croom¹³,^{2,3} Joss Bland-Hawthorn¹³,^{2,3} Michael Goodwin⁹
 Jon S. Lawrence¹⁴ Matt S. Owers¹⁵,^{10,15} and Nicholas Scott¹⁵,^{2,3}

¹Research School of Astronomy and Astrophysics, The Australian National University, Cotter Road, Weston, ACT 2611, Australia

²ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

³Sydney Institute for Astronomy, School of Physics, The University of Sydney, A28, NSW 2006, Australia

⁴International Centre for Radio Astronomy Research, The University of Western Australia, 35 Stirling Hw, Crawley, WA 6009, Australia

⁵Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia

⁶School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia

⁷Instituto de Astronomía, Universidad Nacional Autónoma de México, A.P. 70-264, 04510 México, D.F., México

⁸Ritter Astrophysical Research Center University of Toledo, Toledo, OH 43606, USA

⁹Australian Astronomical Optics – Macquarie, 105 Dehli Road, North Ryde, NSW 2113, Australia

¹⁰Department of Physics and Astronomy, Macquarie University, NSW 2109, Australia

¹¹School of Physics, University of New South Wales, NSW 2052, Australia

¹²SOFIA Science Center, USRA, NASA Ames Research Center, Building N232, M/S 232-12, P.O. Box 1, Moffett Field, CA 94035-0001, USA

¹³Australian Astronomical Optics, AAO-USydney, School of Physics, University of Sydney, NSW 2006, Australia

¹⁴Australian Astronomical Optics – Macquarie, Macquarie University, NSW 2109, Australia

¹⁵Astronomy, Astrophysics and Astrophotonics Research Centre, Macquarie University, Sydney, NSW 2109, Australia

Accepted 2021 January 8. Received 2021 January 8; in original form 2020 February 12

ABSTRACT

We measure the gas-phase metallicity gradients of 248 galaxies selected from Data Release 2 of the SAMI Galaxy Survey. We demonstrate that there are large systematic discrepancies between the metallicity gradients derived using common strong emission line metallicity diagnostics. We determine which pairs of diagnostics have Spearman’s rank coefficients greater than 0.6 and provide linear conversions to allow the accurate comparison of metallicity gradients derived using different strong emission line diagnostics. For galaxies within the mass range $8.5 < \log (M/M_{\odot}) < 11.0$, we find discrepancies of up to 0.11 dex/ R_e between seven popular diagnostics in the metallicity gradient–mass relation. We find a suggestion of a break in the metallicity gradient–mass relation, where the slope shifts from negative to positive, occurs between $9.5 < \log (M/M_{\odot}) < 10.5$ for the seven chosen diagnostics. Applying our conversions to the metallicity gradient–mass relation, we reduce the maximum dispersion from 0.11 dex/ R_e to 0.02 dex/ R_e . These conversions provide the most accurate method of converting metallicity gradients when key emission lines are unavailable. We find that diagnostics that share common sets of emission line ratios agree best, and that diagnostics calibrated through the electron temperature provide more consistent results compared to those calibrated through photoionization models.

Key words: galaxies: abundances – galaxies: fundamental parameters – galaxies: ISM .

1 INTRODUCTION

The global gas-phase metallicity (hereafter metallicity) of a galaxy correlates strongly with its stellar mass, giving rise to the well known mass–metallicity relation (Lequeux et al. 1979; Garnett & Shields 1987; Tremonti et al. 2004; Sweet et al. 2014; Sánchez et al. 2019). With advancements in integral field spectroscopy (IFS), the metallicity of a galaxy on intragalactic scales can now

be resolved for large numbers of galaxies. Local disc galaxies typically exhibit negative metallicity gradients, where the metallicity of a galaxy decreases radially from the centre (Vila-Costas & Edmunds 1992; Zaritsky, Kennicutt & Huchra 1994; Moustakas et al. 2010; Rupke, Kewley & Chien 2010). The existence of a common metallicity gradient among isolated galaxies, when normalized by scalelength, has been found with the introduction of current generation IFS surveys such as the Calar Alto Legacy Integral Field Area (CALIFA) survey, the Sydney Australian Astronomical Observatory Multi-Object Integral Field Spectrograph (SAMI) Galaxy Survey, and the Mapping Nearby Galaxies at Apache

* E-mail: henry.poetrodjojo@sydney.edu.au

Point Observatory (MaNGA) survey (Sánchez et al. 2012, 2014; Ho et al. 2015; Sánchez-Menguiano et al. 2016, 2018; Poetrodjojo et al. 2018). Other studies indicate that the metallicity gradient of a galaxy steepens with stellar mass up to $\log(M/M_\odot) \sim 10.5$, beyond which the metallicity gradients begin to flatten (Belfiore et al. 2017). In contrast, galaxies that show evidence of a major merger or merger remnants at some point during their evolution show shallower metallicity gradients than isolated galaxies (Kewley et al. 2010; Rich et al. 2012; Sánchez et al. 2014; López-Sánchez et al. 2015).

The gas-phase abundance of oxygen, i.e. the ratio between oxygen and hydrogen, O/H, is often used to measure the metallicity of a galaxy. Oxygen is the most abundant element in the Universe after hydrogen and helium, having very strong optical emission lines which are easy to detect even in very distant objects. Since these strong emission lines are collisionally excited, the abundance of the emitting ion can be determined directly if the electron temperature is known. In particular, measuring the electron temperature through the [O III] λ 4363 and λ 5007 emission lines provides a theoretically very reliable diagnostic for the abundance of O^{+2} , and thus metallicity (known as the 'direct temperature method'; Alloin et al. 1979; Pagel et al. 1979). The practical weakness of the direct temperature method lies in the relative faintness of the [O III] λ 4363 and other auroral emission lines, especially at high metallicities (since the intensity of the auroral lines anticorrelates with metallicity). Because of this limitation, metallicity diagnostics involving strong emission lines have become more widely adopted for their ability to be used with low surface brightness, distant and metal-rich sources.

Strong emission line metallicity diagnostics are usually calibrated using:

- (i) Theoretical models (e.g. Kewley & Dopita 2002; Kobulnicky & Kewley 2004; Dopita et al. 2016; Thomas et al. 2018).
- (ii) Empirical calibrations based on the direct temperature method (e.g. Alloin et al. 1979; Pagel et al. 1979; Pilyugin 2001; Pilyugin & Grebel 2016; Ho 2019).
- (iii) A combination of the above (e.g. Denicoló, Terlevich & Terlevich 2002; Pettini & Pagel 2004).

Metallicity calibrations are generally a simple polynomial mapping of emission line ratios to the gas-phase metallicity, but more complex metallicity calibrations can involve the use of Bayesian inference (Blanc et al. 2015; Thomas et al. 2018) or neural network machine learning (Ho 2019; Wu & Boada 2019). Despite all metallicity diagnostics aiming to measure the same quantity (O/H), large discrepancies of up to 0.6 dex exist between diagnostics calibrated through the direct method and theoretical models (Yin et al. 2007; Kewley & Ellison 2008; López-Sánchez & Esteban 2010; López-Sánchez et al. 2012). Yin et al. (2007) also found discrepancies of 0.2 dex between the direct method calibrations from Tremonti et al. (2004), Pilyugin (2001), and Pilyugin & Thuan (2005). These variations in metallicity means that the absolute metallicity scales of galaxies are highly uncertain and cannot be compared between different methods.

Another source of uncertainty that could lead to differences between metallicity diagnostics is the Diffuse Ionized Gas (DIG). The vast majority of strong emission line metallicity diagnostics rely on the assumption that the emission lines are produced from star-forming H II regions. However, H II regions are not the only source of emission in a galaxy. Other sources of emission include active galactic nuclei (AGN), shocks, and the DIG (Kewley et al. 2006). While AGN are relatively simple to identify and separate from star-forming regions (Davies et al. 2016, 2017; D'Agostino

et al. 2019a, b), the DIG is difficult to remove from the spectrum. The DIG is generally found over the entire disc of the galaxy as well as above and below the galactic plane, making it hard to isolate at low spatial resolution scales where the boundaries between H II regions and the DIG are blurred (Walterbos & Braun 1994; Ferguson, Wyse & Gallagher 1996; Hoopes, Walterbos & Greenwalt 1996; Greenawalt et al. 1998). Boettcher, Gallagher & Zweibel (2017) were able to isolate the extraplanar DIG (eDIG) by using high-spectral resolution ($R = 5490$) and Markov chain Monte Carlo methods. Oey et al. (2007) found a mean fraction of 0.59 ± 0.19 of H α surface brightness originating from the DIG. At high spatial resolution scales [such as CALIFA, Physics at High Angular resolution in Nearby Galaxies (PHANGS) or TYPHOON], the isolation of the DIG from H II regions is much simpler and can be done with a variety of techniques (e.g. [S II] λ 6717, λ 6731/H α , or the equivalent width of H α). At the spatial resolution scales of SAMI or MaNGA, completely removing the DIG from H II region emission is difficult. With such a large portion of emission line flux originating from the DIG, DIG contamination inevitably affects our measurements of metallicity (Zhang et al. 2017). Poetrodjojo et al. (2019) showed how the metallicity gradient of M83 was affected by the DIG for 5 different metallicity diagnostics, showing clear differences in how each diagnostic responds to the contamination of DIG emission. Using the direct temperature method, Richards et al. (2014) found that the metallicity of an offset unresolved H II complex within a dwarf galaxy, to be 0.2 dex lower than the surrounding DIG. Conversely, Sweet et al. (2014) found that for a sample of star-forming dwarf galaxies that including emission from the surrounding DIG gave mean metallicities consistent with those measured using emission only from the H II regions with metallicity calibrations by Dopita et al. (2013). The consistent metallicities between the H II regions and the DIG are likely because the large gas reservoirs in the sampled dwarf galaxies were well mixed (Kobulnicky & Skillman 1997; Lee & Skillman 2004). However, even without the contamination of DIG, the H II regions of M83 show very different metallicity gradients depending on which metallicity diagnostic is used (Poetrodjojo et al. 2019).

Recent studies on the mass–metallicity relation attempt to remove any biases caused by the choice of gas-phase metallicity calibrator by performing their analysis with a wide range of calibrations (Barrera-Ballesteros et al. 2017; Sánchez et al. 2017, 2019). Each of these three studies find little to no evidence of any secondary dependence of the mass–metallicity relation with star formation rate, regardless of which metallicity diagnostic or calibrator is used. Not only is the robustness of conclusions increased by utilizing many different metallicity calibrations, Barrera-Ballesteros et al. (2017) and Sánchez et al. (2017, 2019) demonstrate that although systematics plague the measurement of gas-phase metallicity, overall trends are conserved. However, the comparison of individual objects from different samples still remains difficult without the luxury of a wide wavelength coverage.

Although Kewley & Ellison (2008) provided conversions between the current available diagnostics at the time, the conversions were calibrated using SDSS aperture spectroscopy. With the rise of spatially resolved IFS observations, we are moving away from measuring global metallicities and instead measuring metallicity *gradients*. With all the issues presented, it is inadvisable to compare metallicity gradients determined from different diagnostics. With the large amount of different instruments observing a wide range of redshifts currently available, we cannot reasonably expect there to always be overlapping wavelength coverage. In these situations, it becomes essential to develop a way to convert and compare

metallicity gradients determined from different diagnostics. This is especially important in understanding how metallicity gradients evolve as a function of redshift, where it becomes extremely difficult to obtain a broad range of optical emission lines.

In this paper, we use galaxies from SAMI Data Release 2 (DR2; Scott et al. 2018) to measure metallicity gradients using 13 different metallicity diagnostics and calibrations. We then compare the measured metallicity gradients to determine which diagnostics and calibrations can be empirically converted from one another and provide the conversion fits in a table. We discuss the differences between the diagnostics which lead to their inconsistencies and compare different calibration methods.

We structure this paper in the following way. Section 2 describes the SAMI Galaxy Survey and how we select our sub-sample from the data available. We outline the methods we use for determining the gas-phase metallicity based on popular strong emission line diagnostics in Section 3. In Sections 4 and 5, we present and briefly discuss our results on the metallicity gradients and provide a method for converting between different diagnostics. Finally, we summarize our findings and outline future work that we hope will stem from these results. Throughout the entire paper, we assume the following values for cosmological constants, $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$ (Hinshaw et al. 2009).

2 SAMPLE SELECTION

2.1 SAMI Galaxy survey

The SAMI Galaxy survey (Bryant et al. 2015) is an integral field spectroscopic survey of ~ 3000 low-redshift ($z < 0.095$) galaxies primarily selected from the Galaxy and Mass Assembly (GAMA) survey (Driver et al. 2011), with the addition of eight galaxy clusters to extend the sampling of environmental density (Owers et al. 2017). The Sydney Australian Astronomical Observatory Multi-Object Integral Field Spectrograph (SAMI; Croom et al. 2012) is located on the 3.9-m Anglo-Australian Telescope at Siding Spring Observatory. The SAMI Galaxy Survey covers a wide range of galaxies with stellar masses ranging between 10^7 and $10^{12} M_\odot$, and redshifts between $0.004 < z < 0.095$.

The SAMI data are sampled at $0.25 (0.5 \times 0.5)$ arcsec² spaxels covering the 14.7 arcsec diameter aperture of the SAMI hexabundle (Bland-Hawthorn et al. 2011; Bryant et al. 2014) with an average seeing of 2.16 arcsec (Green et al. 2018). Since the average seeing is much larger than the individual spaxel sizes, this leads to an oversampling of our data, resulting in a covariance between neighbouring spaxels. We take this covariance into account when implementing the various binning schemes on the SAMI data. The SAMI fibres are fed into the double-beam AAOmega spectrograph (Sharp et al. 2006). The blue cube covers a wavelength range between 3700 and 5700 Å with a spectral resolution of $R = 1812$ and the red cube covers a wavelength range between 6300 and 7400 Å with a spectral resolution of $R = 4263$ (van de Sande et al. 2017). The spectral range of the AAOmega spectrograph allows us to observe the important metal-sensitive emission lines: [O II]λ3726, λ3729, Hβλ4861, [O III]λ5007, Hαλ6563, [N II]λ6583, and [S II]λ6717, λ6731.

Poetrodjojo et al. (2018) found that the relatively weak detection of the [O II] emission line compared to the other metal-sensitive emission lines significantly reduced the ability to measure the metallicity using the R₂₃ and N2O2 emission line diagnostic. To increase the detection of the weaker emission lines, we will use the sector-binned data cubes released in the SAMI DR 2 (Scott

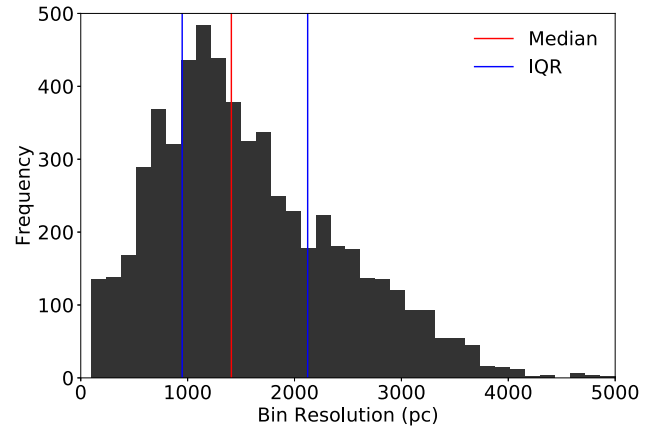


Figure 1. Spatial resolution distribution of the sector-binned data cubes in our final galaxy sample. The median spatial resolution of our sample is 1409 pc with a 25th and 75th percentile of 735 and 2507 pc, respectively.

et al. 2018). The row stacked spectra (RSS) are first binned into five linearly spaced elliptical annuli based on their position angle (PA) and ellipticity. The cubes are then further azimuthally subdivided into eight regions, resulting in sector-binned cubes. The sector-binned data cubes have the advantage over unbinned data of increasing the S/N of emission lines while maintaining both azimuthal and radial spatial resolution. Fig. 1 shows the spatial resolution distribution of the sector-binned data cubes of our final galaxy sample. For full details on the generation of the binned data cubes (see Scott et al. 2018).

The emission line fitting was done using the SAMI line fitting routine LZIFU (Ho et al. 2016; Medling et al. 2018). We first subtract the underlying stellar continuum using PPXF (Cappellari & Emsellem 2004; Cappellari 2017) and the MUSECAT simple stellar population models (Vazdekis et al. 2012). The dominant emission lines are then fit using up to 3 Gaussian profiles with the Levenberg–Marquardt least-squares method implemented in MPFIT (Markwardt 2009). Each emission line is constrained to have the same kinematic velocity and velocity dispersion. The flux ratios of [O III]λλ4959, 5007, and [N II]λλ6548, 6583 are fixed to those given by quantum mechanics.

2.2 Well-resolved radial profiles

While the SAMI DR2 consists of 1559 galaxies, many are observed at high inclination angles or found to have significant non-stellar emission contaminating their spectra. To measure reliable metallicity gradients, we select galaxies with relatively face-on profiles so that the minor axis can be well sampled. We select galaxies with inclinations of $< 60^\circ$, which reduces our sample from 1559 galaxies to 941 galaxies. We calculate the inclination angle of the galaxy using the standard Hubble formula (Hubble 1926)

$$\cos^2(i) = \frac{(b/a)^2 - q_0^2}{1 - q_0^2}, \quad (1)$$

where i is the inclination angle, $q_0 = 0.2$, and b/a is the ratio between the minor and major axis as measured in the r band by the Sloan Digital Sky Survey (SDSS; York et al. 2000) using GALFIT (Peng et al. 2010). To sufficiently sample the entire scalelength over which we measure the metallicity gradients, we require at least one sector every $0.25R_e$ up to $1R_e$ (as measured by GALFIT), further reducing our sample to 257 galaxies. This large drop in our galaxy sample is mostly driven by requiring a signal-to-noise ratio (S/N) > 3 in all

of the strong emission line fits within the spectral range of SAMI in each sector. The lower throughput in the blue arm of the spectrograph causes the S/N of the [O II] emission line to be relatively weak compared to the other optical emission lines. However, the use of sector-binned cubes increases our sample size tenfold compared to the 25 galaxies used in Poetrodjojo et al. (2018) (who used unbinned data).

2.3 Star-forming cuts

Blue cloud galaxies typically emit strong emission lines from H II regions surrounding recently formed massive stars. However, H II regions are not the only possible source of emission. Gas excited from the DIG, shocks, or AGN may contribute to the overall emission line profile (e.g. Groves, Dopita & Sutherland 2004). The large majority of strong emission line metallicity diagnostics are calibrated on the assumption that all of the emission is produced from H II regions. Recent work by Kumari et al. (2019) and Vale Asari et al. (2019) allows for the measurement of gas-phase metallicity in spaxels (integrated spectra in the case of Vale Asari et al. 2019) dominated by the DIG by providing correction factors to remove its effects. At the spatial resolution of multiplexing IFS surveys, contamination by the DIG is inevitable and causes systematic biases in the measurement of metallicity gradients (Mast et al. 2014; Erroz-Ferrer et al. 2019; Poetrodjojo et al. 2019). In some cases, it is possible to separate the star formation dominated and other ionizing sourced line emission using high spatial resolution observations (e.g. Davies et al. 2014, 2016; D’Agostino et al. 2018; Lacerda et al. 2018), but in our case we chose to remove all sectors that showed significant non-star-forming emission.

We use the classification scheme of Kewley et al. (2006) to distinguish when non-star-forming emission is present using the following strong emission line ratio diagnostic curves:

$$\log \left(\frac{[\text{O III}]}{\text{H}\beta} \right) > \frac{0.61}{\log \left(\frac{[\text{N II}]}{\text{H}\alpha} \right) - 0.05} + 1.30, \quad (2)$$

$$\log \left(\frac{[\text{O III}]}{\text{H}\beta} \right) > \frac{0.72}{\log \left(\frac{[\text{S II}]}{\text{H}\alpha} \right) - 0.32} + 1.30, \quad (3)$$

$$\log \left(\frac{[\text{O III}]}{\text{H}\beta} \right) > \frac{0.73}{\log \left(\frac{[\text{O II}]}{\text{H}\alpha} \right) + 0.59} + 1.33. \quad (4)$$

Sectors with an S/N > 3 in the emission line fluxes that satisfy any of (fail all of) these criteria are classified as non-star forming (star forming). After this cut is performed, we still require at least one sector every $0.25R_e$ up to $1R_e$ to reliably measure the metallicity gradient. We are left with a final sample of 248 galaxies (5832 sectors) for which we can measure metallicity gradients using any of the metallicity diagnostics outlined in the following section. Fig. 2 shows the stellar mass distribution of our galaxies after each sample selection cut. For galaxies $\log(M/M_\odot) < 8.5$ and $\log(M/M_\odot) > 10.5$, the ability to sufficiently sample the galaxy to $1R_e$ is reduced. With low-mass galaxies ($\log(M/M_\odot) < 8.5$), we are unable to sample to fine enough resolution scales (at least $0.25R_e$) to reliably measure metallicity gradients. Due to the limited redshift range, the effective radii of massive galaxies ($\log(M/M_\odot) > 10.5$) are larger than the SAMI field of view and are therefore excluded from the sample.

3 MEASURING METALLICITY GRADIENTS

3.1 Extinction correction

Before being used in diagnostic ratios, emission lines must be first corrected for attenuation by dust in the interstellar medium (ISM).

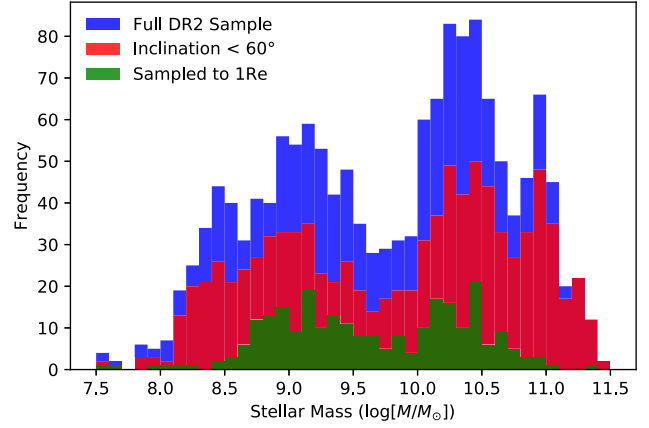


Figure 2. Stellar mass distribution of our galaxy sample. Between a stellar mass range of $8.5 < \log(M/M_\odot) < 10.5$ we have a similar mass distribution as the full DR2 sample.

The attenuation of emission lines is wavelength dependent, meaning that diagnostics that use emission lines that are widely separated in wavelength are most heavily affected, such as N2O2, R₂₃, and O32. To correct the emission lines, we create maps of the observed Balmer ratio, $(\text{H}\alpha/\text{H}\beta)_{\text{obs}}$, and solve for $E(B - V)$ by using the relation

$$E(B - V) = \log_{10} \left(\frac{(\text{H}\alpha/\text{H}\beta)_{\text{obs}}}{(\text{H}\alpha/\text{H}\beta)_{\text{int}}} \right) / (0.4(k(\text{H}\beta) - k(\text{H}\alpha))), \quad (5)$$

where $(\text{H}\alpha/\text{H}\beta)_{\text{int}}$ is the intrinsic ratio of 2.86 for case B recombination (Osterbrock 1989). We note that the intrinsic Balmer line ratio is also a function of metallicity (López-Sánchez et al. 2015). We use the Cardelli, Clayton & Mathis (1989) extinction curve and assume a typical $R(V)$ value of 3.1 to determine k values for H α and H β . We then use the calculated $E(B - V)$ to determine $A(\lambda)$ at our emission line wavelengths to de-redden the emission line fluxes. Although emission line ratios with small wavelength separations are not significantly affected by dust extinction, we apply an extinction correction to all emission line ratios to maintain a fair comparison when using them in metallicity diagnostics.

3.2 Metallicity diagnostics

Due to the enormous amount of metallicity diagnostics and calibrations available, it is unrealistic to analyse every single one. We therefore limit this study to popular metallicity diagnostics, which are used extensively by the gas-phase metallicity community. To cover as much parameter space as possible, we select diagnostics which use unique combinations of strong emission lines, calibrations using different photoionization codes and electron temperature methods, as well as exploring calibration mapping methods including polynomial mapping, Bayesian inference and machine learning algorithms. The equations for all diagnostics and calibrations used in this study are given in the Appendix.

3.2.1 R₂₃

The $([\text{O II}]\lambda\lambda 3726, 3729 + [\text{O III}]\lambda\lambda 4959, 5007)/\text{H}\beta$ (R₂₃) emission line ratio measures the oxygen abundance through the direct use of oxygen emission lines. Due to its popularity, many calibrations for the R₂₃ diagnostic exist (Pagel et al. 1979; Pagel, Edmunds & Smith 1980; Edmunds & Pagel 1984; McCall, Rybski & Shields 1985; Dopita & Evans 1986; Torres-Peimbert, Peimbert & Fierro

1989; McGaugh 1991; Zaritsky et al. 1994; Pilyugin 2000; Charlot & Longhetti 2001; Kewley & Dopita 2002; Kobulnicky & Kewley 2004). However, one major complication with the R_{23} emission line ratio is its strong dependence on the ionization state of the gas, as quantified by the ionization parameter. We use two popular calibrations by Kobulnicky & Kewley (2004, hereafter **K04R₂₃**) and Curti et al. (2017, hereafter **C17R₂₃**). Kobulnicky & Kewley (2004) used the R_{23} emission line ratio in conjunction with the $[\text{O III}]\lambda\lambda 4959, 5007/[\text{O II}]\lambda\lambda 3726, 3729$ (O32) emission line ratio to simultaneously constrain the metallicity and ionization parameter through an iterative method. Curti et al. (2017) take a more conventional approach in which the R_{23} emission line ratio is mapped to the direct temperature metallicities through a high-order polynomial. This approach is much less computationally expensive and provides an easy way to calculate the gas-phase metallicities using the oxygen emission lines.

3.2.2 N2O2

Unlike R_{23} , the $[\text{N II}]\lambda 6583/[\text{O II}]\lambda 3726, \lambda 3729$ (N2O2) emission line ratio is relatively insensitive to the ionization parameter because of the similar ionizing potentials of the nitrogen and oxygen species. The biggest drawback of the N2O2 emission line ratio is the strong extinction required due to the large wavelength differences between the $[\text{N II}]$ and $[\text{O II}]$ emission lines. We use the calibration outlined in Kewley & Dopita (2002), where they calibrate the N2O2 emission line ratio using the MAPPINGS III photoionization models (Sutherland et al. 2013) shown in equation (3).

3.2.3 N2H α

Perhaps one of the mostly widely used metallicity diagnostics is the $[\text{N II}]\lambda 6583/\text{H}\alpha$ (N2H α) emission line ratio. Due to the small wavelength separation of the $[\text{N II}]$ and $\text{H}\alpha$ emission line, the N2H α emission line ratio is popular among high-redshift studies where extinction correction may be difficult or only a small wavelength coverage is available (e.g. Storchi-Bergmann, Calzetti & Kinney 1994). We use two popular calibrations of the N2H α emission line ratio from Pettini & Pagel (2004, hereafter **P04**) and Marino et al. (2013, hereafter **M13**). The main difference between the calibrations presented by Pettini & Pagel (2004) and Marino et al. (2013) occurs at the high-metallicity end. Due to the limitations of constraining the electron temperature, Pettini & Pagel (2004) used photoionization models by Diaz et al. (1991) and Castellanos, Díaz & Terlevich (2002a, b) to calibrate the N2H α diagnostic at high metallicities. With advances in the sensitivity of spectrographs, Marino et al. (2013) were able to measure the electron temperature using the $[\text{N II}]\lambda 5755$ emission line, leading to a more consistent calibration.

3.2.4 O3N2

Similar to the N2H α diagnostic, $([\text{O III}]\lambda 5007/\text{H}\beta)/([\text{N II}]\lambda 6583/\text{H}\alpha)$ (O3N2) uses emission lines that have minimal wavelength separation to remove the need for an extinction correction. O3N2 uses the same emission lines that are commonly presented on the BPT diagram (Baldwin, Phillips & Terlevich 1981), a method for separating star-forming regions from active galactic nuclei (AGN) dominated regions. The calibrations by Pettini & Pagel (2004) and Marino et al. (2013) show that the gas-phase metallicity is a linear function of the O3N2 emission line ratio, allowing for efficient calculations. We use the calibrations from Pettini & Pagel (2004)

and Marino et al. (2013), the same studies as our N2H α calibrations, for our analysis.

3.2.5 N2S2H α

The N2S2H α emission line diagnostic is a combination of the $[\text{N II}]\lambda 6583/\text{H}\alpha$ and $[\text{N II}]\lambda 6583/[\text{S II}]\lambda\lambda 6717, 6731$ emission line ratios. As with N2H α and O3N2, N2S2H α uses emission lines close in wavelength to remove the need for extinction corrections. The N2S2H α emission line diagnostic is sensitive to the metallicity through the $[\text{N II}]/\text{H}\alpha$ emission line ratio and removes the ionization parameter dependence through the $[\text{N II}]/[\text{S II}]$ emission line ratio. This combination of emission lines is calibrated through the MAPPINGS V photoionization models by Dopita et al. (2016, hereafter **D16**).

3.2.6 ONS and ON

The emission line diagnostics introduced so far have been a combination of one or two emission line ratios. The oxygen–nitrogen–sulphur (ONS) emission line diagnostic incorporates four different emission line ratios to measure gas-phase metallicity. Similar to **K04R₂₃**, the ONS calibration is split into multiple branches. The calibration is split into three classes of H II regions; cool, warm, and hot. The conditions of the H II regions are determined from the $[\text{N II}]/\text{H}\beta$ and $[\text{N II}]/[\text{S II}]$ emission line ratios. The ONS calibration also includes an explicit excitation parameter, $[\text{O III}]/([\text{O II}]+[\text{O III}])$, allowing for changes in the ionization parameter. Pilyugin, Vílchez & Thuan (2010) also provide an additional metallicity calibration, the oxygen–nitrogen (ON) diagnostic that provides comparable quantities as the ONS calibration for all H II region classes without using the $[\text{S II}]$ emission line. Although the metallicity calibrations themselves do not use the $[\text{S II}]$ emission line, it is still required to classify the H II region.

3.2.7 Rcal and Scal

Similar their previous works, Pilyugin & Grebel (2016) created two new metallicity diagnostics involving more than two emission line ratios. The R-calibrations (Rcal) and S-calibrations (Scal) were calibrated using the counterpart method (Pilyugin, Grebel & Mattsson 2012), which ultimately derives metallicity from the T_e method. The emission lines used in Rcal and Scal only differ by swapping out the $[\text{O II}]$ emission line in Rcal for the $[\text{S II}]$ emission line in Scal. Like the **K04R₂₃** diagnostic, the calibration is split into low- and high-metallicity branches based on the $[\text{N II}]/\text{H}\beta$ emission line ratio. The Rcal and Scal diagnostics agree with directly measured abundances to within 0.1 dex, comparable to the expected uncertainties of the abundances themselves.

3.2.8 NebulaBayes

The use of Bayesian Inference to constrain the ionization parameter and metallicity was first introduced by Blanc et al. (2015) with the Interactive Data Language (IDL) code IZI. The advantage of using Bayesian inference to constrain the metallicity and ionization parameter is the ability to include additional emission line information to improve estimates. For example, with simple emission line ratios such as R_{23} and N2O2, we are unable to include the $[\text{S II}]$ emission line to better predict the gas-phase metallicity. With Bayesian analysis, we are able to include the entire suite of available emission lines to determine the metallicity. For this study we use

NebulaBayes (NB; Thomas et al. 2018), a Bayesian Inference code inspired by IZI with more generalized capabilities. We use the following emission lines to constrain the metallicity and ionization parameter using NB: [O II] $\lambda\lambda$ 3726, 3729, H β λ 4861, [O III] λ 5007, H α λ 6563, [N II] λ 6583, and [S II] $\lambda\lambda$ 6717, 6731.

3.2.9 Machine learning

Recently, the use of machine learning algorithms has spiked with the ability to dedicate large portions of time to training neural networks. Two recent examples include implementations by Wu & Boada (2019) and Ho (2019). Wu & Boada (2019) use a convolutional neural network (CNN) to predict the gas-phase metallicity of SDSS galaxies using only the optical *gri* images. Ho (2019) uses a multilayer perceptron (MLP) model to predict the gas-phase metallicity using the extinction-corrected [O II], H β , [O III], [N II], and [S II] emission lines. For this study, we use OxygenMLP (MLP; Ho 2019) as our machine learning fiducial model. Unlike NB, MLP requires that all emission lines are available for use.

3.2.10 Photoionization models

Pérez-Montero (2014) presents HII-CHI-mistry (hereafter, HII-CHI), a method of determining metallicity through photoionization models calculated with CLOUDY v17.00 (Ferland et al. 2017). HII-CHI-mistry v4.0 is a publicly available PYTHON routine which aims to constrain O/H, N/O, and log(U) using several of the already introduced extinction corrected optical emission lines [O II], [Ne III] λ 3868, [O III] λ 4363, H β , [O III] λ 5007, [N II], and [S II]. Due to the relative weakness of the [Ne III] λ 3868 and [O III] λ 4363 emission lines, we do not provide these fluxes to HII-CHI, limiting its usage to the ‘log U limited’ photoionization grids.

3.3 Error propagation

To propagate the line flux errors produced by LZIFU through to the metallicity calculations, we simulate 1000 maps for all emission lines used in the calculation. The maps are created such that the fluxes are Gaussian distributed within the LZIFU standard deviation for that emission line.

Using the simulated line maps, metallicity maps are created for each metallicity diagnostic. The non-linearity of some of the metallicity diagnostics means that the metallicity distributions are not necessarily Gaussian. To represent the spread of metallicity, we determine the distance from the true value to the 16th and 84th percentiles and calculate the average. This average provides us with a measure of the error of the metallicity maps, which are then propagated to the gradient errors.

3.4 Metallicity gradients

When determining the radial metallicity gradients, we first correct for the observed inclination of the galaxy using its ellipticity and position angle as measured in the *r* band by the GAMA survey (Driver et al. 2011). We then use a linear least-squares algorithm to fit a linear trend to the metallicity gradient, propagating the uncertainty in metallicity through the linear parameters. We use FITEXY (Press et al. 1992), available for use with IDL, to perform the fitting because it does not detect and remove outliers. We have specifically avoided more robust line fitting algorithms such as LTS_LINEFIT (Cappellari et al. 2013) because they automatically remove any detected outliers.

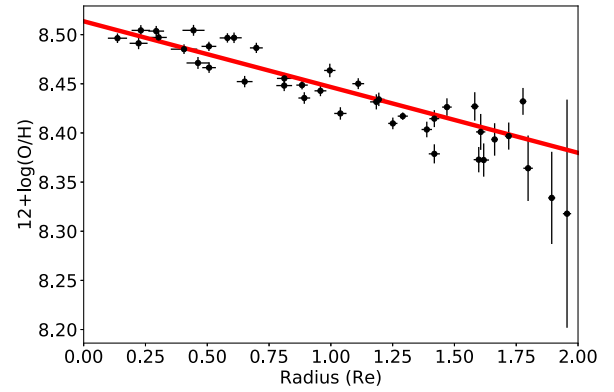


Figure 3. A typical negative radial metallicity gradient of GAMA-492414 using the Scal metallicity diagnostic. The red line represents the best linear fit to the sector metallicities.

Since we are investigating how the metallicity gradient varies among different metallicity diagnostics, using an algorithm that may detect a sector as an outlier in one diagnostic but not another is undesirable and introduces a potential source of uncertainty. The gradients are then normalized by the effective radius (R_e) of the galaxy to remove the size dependence of metallicity gradients (Sánchez et al. 2014; Ho et al. 2015; Sánchez-Menguiano et al. 2016). Sánchez-Menguiano et al. (2018) found wide-spread deviations from single linear metallicity gradients and instead used multiple linear gradients to more accurately fit the radial metallicity distribution. However, for this study we adopt a single linear fit to our metallicity gradients because at the spatial resolution of the sector binned cubes, we are unable to resolve a broken linear fit to the metallicity gradient.

Several studies (Yuan, Kewley & Rich 2013; Mast et al. 2014; Poetrodjojo et al. 2019) have shown that there is a systematic flattening of the metallicity gradient at lower spatial resolution scales. Although the sector-binning of the SAMI data cubes is effectively reducing the spatial resolution, we find that the metallicity gradients are not affected in a significant way. This is because the flattening of metallicity gradients is most noticeable at higher spatial resolutions and the effect is heavily diminished at the kiloparsec resolution scales of most SAMI galaxies. We show an example of a radial metallicity gradient in Fig. 3. We can see that there is sufficient spatial resolution to characterize a single linear fit to the metallicity gradient, but finer resolutions scales are needed for multilinear fits.

4 RESULTS

Before we compare metallicity gradients derived from different metallicity diagnostics, we first compare our data to the fits determined by Kewley & Ellison (2008). Fig. 4 shows the metallicities of the individual sectors from all the galaxies in our sample. The red line in Fig. 4 shows the empirical conversions determined by Kewley & Ellison (2008) using single aperture spectra from SDSS. In the bottom row of Fig. 4, the blue line shows the empirical conversion between Scal and three other metallicity diagnostics determined by De Vis et al. (2019) using spectra obtained from DustPedia and MUSE. We also show the median metallicities in green as a comparison. The upturn in the K04R₂₃ metallicities around $12 + \log(\text{O}/\text{H}) \approx 8.5$ is caused by the dual branched nature of the K04R₂₃ metallicity diagnostic. We find that although the fits by Kewley & Ellison (2008) and De Vis et al. (2019) follow the general trend in all cases, there appear to be significant deviations from the median metallicities in some cases (e.g. N2O2 versus P04N2H α). The spread

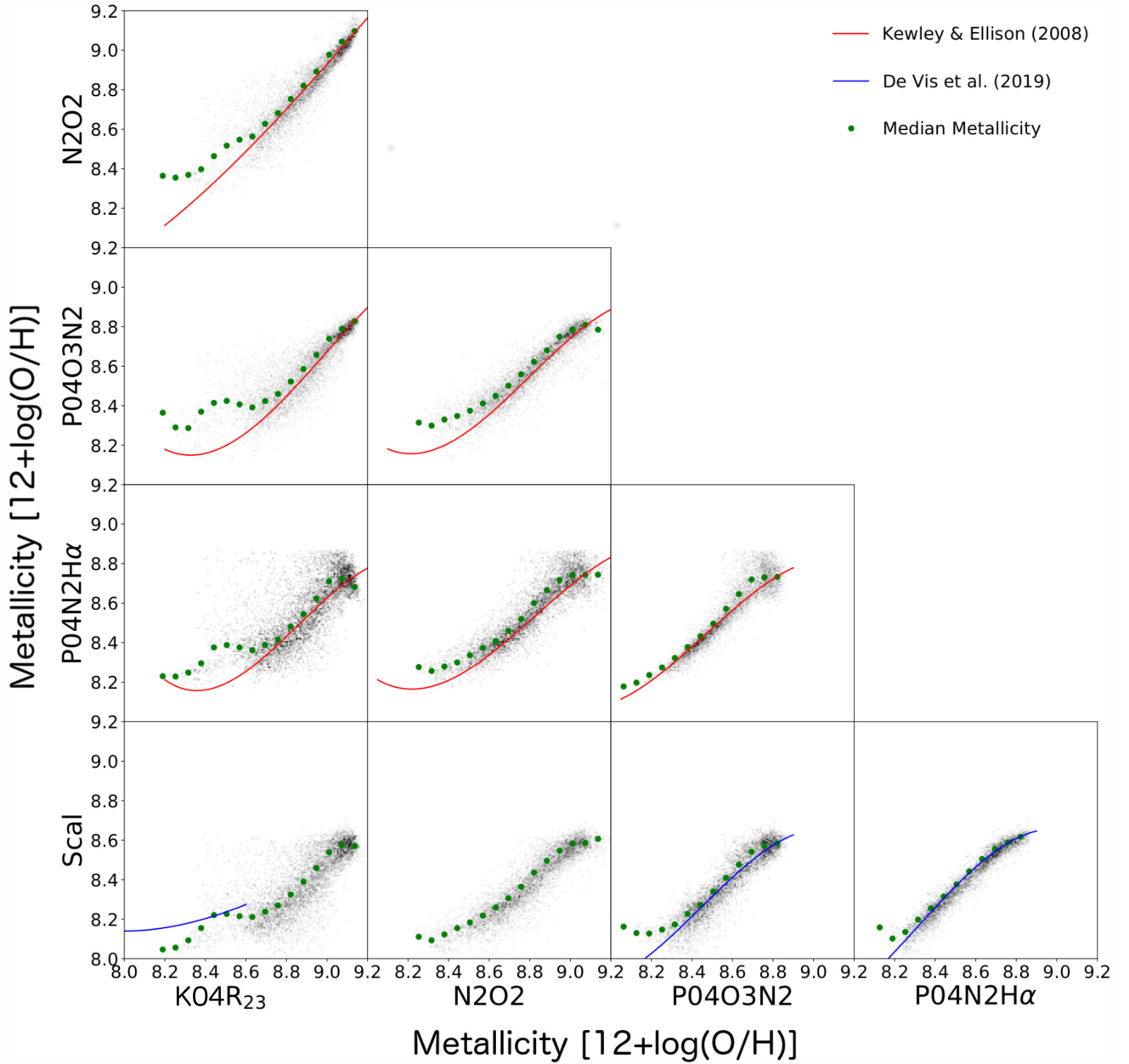


Figure 4. Metallicities of individual sectors calculated through different metallicity diagnostics as a function of each other. The red line represents the empirical conversions determined by Kewley & Ellison (2008) and the blue line represents the empirical conversions determined by De Vis et al. (2019). While the conversions by Kewley & Ellison (2008) and De Vis et al. (2019) follow the general trend, a noticeable scatter exists around each fit.

around the empirical conversions by Kewley & Ellison (2008) and De Vis et al. (2019) means that the simple solution of doing a spaxel by spaxel conversion may not produce realistic metallicity gradients.

Fig. 5 compares the metallicity gradients derived from five different diagnostics. The red line represents the one-to-one line where we expect points to lie when the metallicity diagnostics are in agreement. It is obvious in Fig. 5 that significant scatter exists between the diagnostics, with some diagnostics agreeing better than others. In some cases (e.g. $K04R_{23}$ versus $N2S2H\alpha$), no trend can be discerned from the scatter, meaning that converting between these two diagnostics is unreliable. Fig. 4 shows that the relationship between metallicity diagnostics tends to be monotonic, which also holds for the remaining diagnostics not shown in Fig. 4.

We therefore expect metallicity gradients to also be monotonic between diagnostics. The Spearman's rank coefficient provides a measure of how well the relationship between two variables can be represented by a monotonic function, with a coefficient of 1 indicating a perfect one-to-one relation. To determine which diagnostics can be reasonably compared, we calculate the Spearman's rank coefficient between the metallicity gradients of all diagnostics and list them in Table 1. We consider diagnostics with a Spearman's rank coefficient greater than 0.6 able to be reliably described by an empirical fit. Approximately 40 per cent (36/91) of all possible diagnostic pairs have a Spearman's rank coefficient greater than 0.6. Using `LTS_LINEFIT`, we fit linear functions to pairs of diagnostics with Spearman's rank coefficients greater than 0.6. Here, we switch

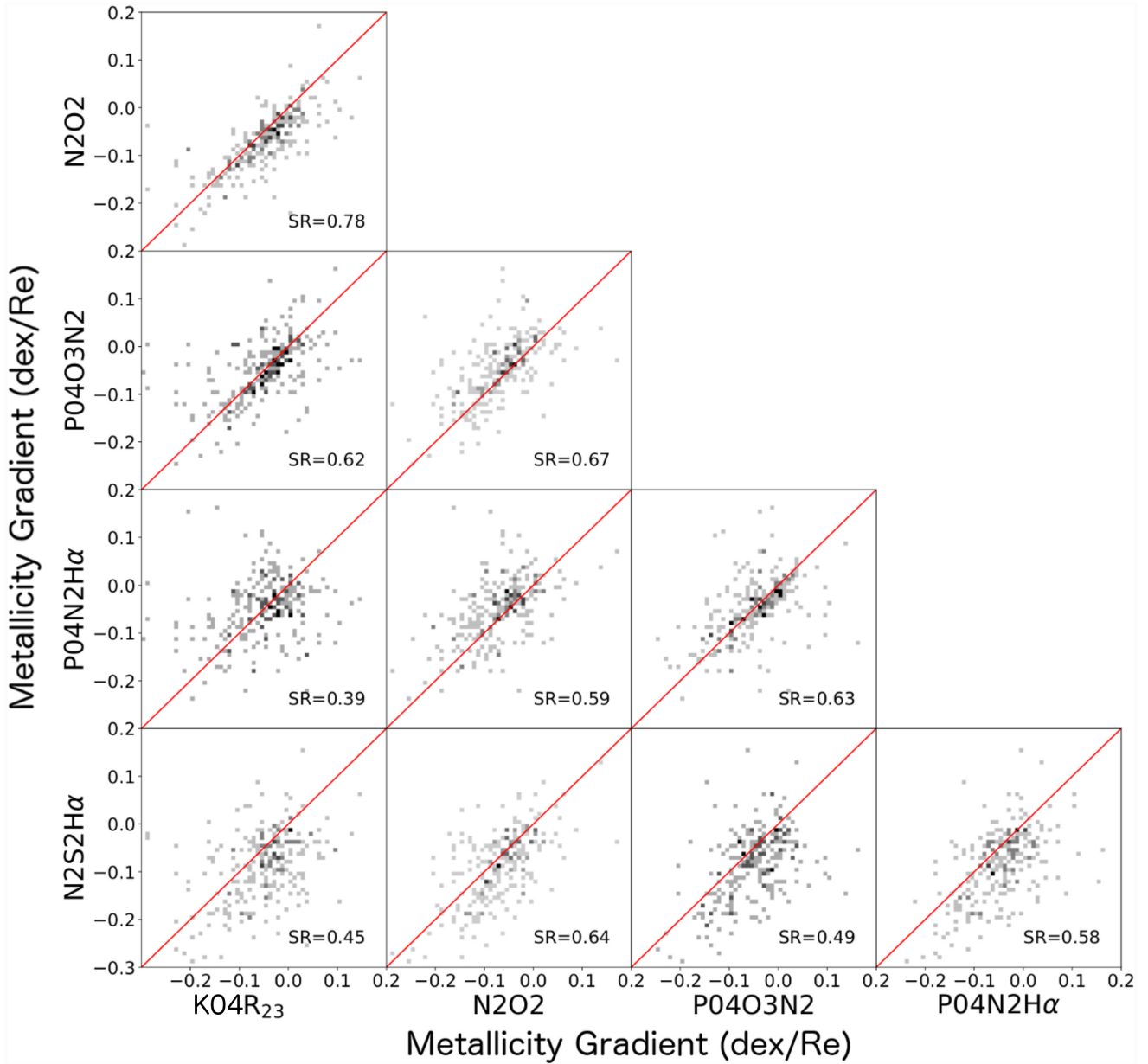


Figure 5. Metallicity gradient of each galaxy calculated through different metallicity diagnostics as a function of each other. The red line represents the one-to-one line indicating perfect agreement between the diagnostics. The Spearman rank coefficient of each diagnostic pair is calculated to test the level of correlation and tabulated in Table 1. For pairs of diagnostics with a Spearman rank greater than 0.6, there exists obvious systematic deviations away from the one-to-one line which are corrected for.

to using `LTS_LINEFIT` instead of `FITEXY` as it is now desirable to remove outliers. The outlier detection of `LTS_LINEFIT` uses the least trimmed squares (LTS) regression approach by Leroy & Rousseeuw (1987). The algorithm aims to find the subset of h data points which produces the smallest χ^2 amongst all possible subsets of $h > N/2$. This computational intensive algorithm provides accurate linear fits to empirical conversions between the different metallicity diagnostics.

Figs 6 and 7 show specific examples of Fig. 5, where we fit our own linear functions to more accurately describe the relationships between diagnostics. These two examples are specifically chosen as they have Spearman ranks greater than 0.6 with gradients significantly different from unity, which clearly demonstrates the sys-

tematic uncertainties between metallicity diagnostics. Fig. 6 shows the metallicity gradients from MLP against the metallicity gradients from Rcal. It is clear that assuming a one-to-one relation (red line) between the diagnostics would introduce significant systematic biases. The best-fitting linear curve, shown in blue, indicates that a much shallower gradient is needed when converting between these two diagnostics. We show the residuals in the lower panel of the figure, and find a standard deviation dispersion of 0.026 dex/ R_e . Another example is shown in Fig. 7 between N2S2H α and Scal. While only small zero-point offsets are present (flat gradients remain flat gradients between the diagnostics), there is a clear difference in slope from the one-to-one relation in both examples. This highlights the danger in assuming that metallicity gradients obtained through

Table 1. Spearman rank coefficients (1σ error) between all metallicity diagnostics and calibrations. Pairs of diagnostics with a Spearman rank coefficient greater than 0.6 are selected for further analysis and are highlighted in bold font. We deem diagnostic pairs with a Spearman rank coefficient less than 0.6 to have no significant correlation.

K04R ₂₃														
P04N2H α	0.39 (0.06)	P04N2H α												
M13N2H α	0.39 (0.06)	0.96 (0.01)	M13N2H α											
P04O3N2	0.62 (0.05)	0.63 (0.05)	0.63 (0.04)	P04O3N2										
M13O3N2	0.61 (0.05)	0.62 (0.05)	0.61 (0.05)	0.98 (0.01)	M13O3N2									
N2O2	0.78 (0.03)	0.59 (0.05)	0.59 (0.05)	0.67 (0.05)	0.68 (0.04)	N2O2								
N2S2H α	0.45 (0.06)	0.58 (0.05)	0.52 (0.06)	0.49 (0.06)	0.51 (0.06)	0.64 (0.05)	N2S2H α							
ONS	0.51 (0.06)	0.39 (0.07)	0.40 (0.07)	0.62 (0.05)	0.60 (0.05)	0.61 (0.05)	0.57 (0.05)	ONS						
ON	0.52 (0.06)	0.17 (0.07)	0.18 (0.07)	0.39 (0.06)	0.40 (0.06)	0.57 (0.06)	0.53 (0.05)	0.79 (0.04)	ON					
Scal	0.44 (0.06)	0.82 (0.03)	0.79 (0.03)	0.53 (0.06)	0.52 (0.06)	0.65 (0.05)	0.88 (0.03)	0.57 (0.06)	0.44 (0.06)	Scal				
Rcal	0.60 (0.05)	0.84 (0.02)	0.85 (0.02)	0.58 (0.05)	0.57 (0.05)	0.83 (0.03)	0.65 (0.05)	0.52 (0.06)	0.41 (0.06)	0.85 (0.02)	Rcal			
NeBayes	0.52 (0.06)	0.37 (0.06)	0.30 (0.07)	0.55 (0.05)	0.56 (0.06)	0.59 (0.05)	0.44 (0.06)	0.34 (0.06)	0.22 (0.07)	0.38 (0.06)	0.46 (0.06)	NeBayes		
MLP	0.53 (0.05)	0.86 (0.02)	0.85 (0.02)	0.65 (0.04)	0.65 (0.04)	0.77 (0.03)	0.68 (0.04)	0.52 (0.06)	0.40 (0.07)	0.83 (0.03)	0.90 (0.02)	0.45 (0.06)	MLP	
HII-CHI	0.37 (0.07)	0.39 (0.06)	0.42 (0.06)	0.71 (0.04)	0.71 (0.04)	0.40 (0.06)	0.26 (0.07)	0.39 (0.06)	0.21 (0.07)	0.30 (0.06)	0.32 (0.06)	0.38 (0.07)	0.39 (0.06)	HII-CHI
C17R ₂₃	0.69 (0.04)	0.27 (0.06)	0.20 (0.07)	0.59 (0.05)	0.60 (0.05)	0.66 (0.05)	0.50 (0.05)	0.50 (0.05)	0.46 (0.06)	0.37 (0.06)	0.41 (0.06)	0.58 (0.05)	0.37 (0.06)	0.42 (0.06)

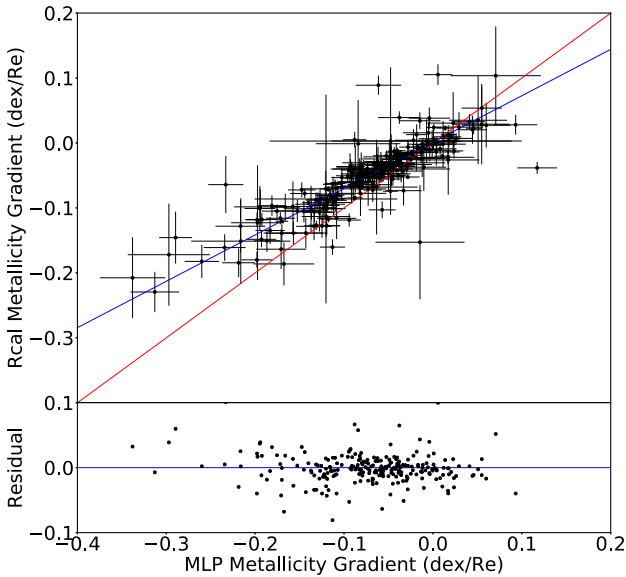


Figure 6. Metallicity gradient calculated using Rcal as a function of the MLP metallicity gradient. The red line represents the one-to-one relation and the blue line shows the best linear fit to the data.

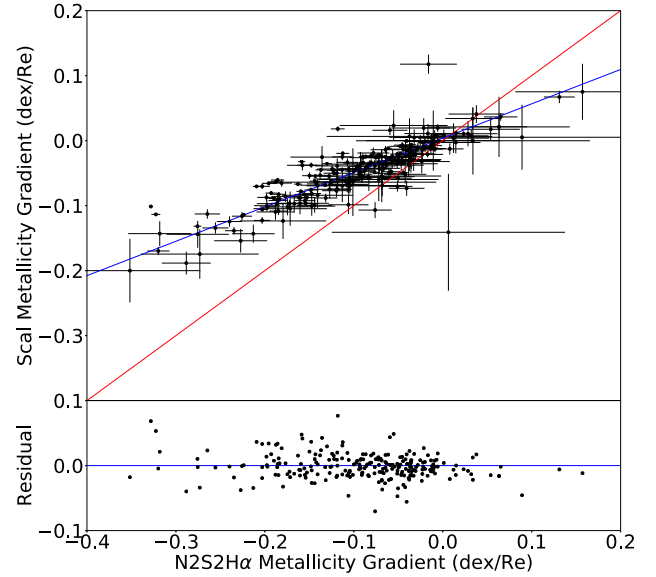


Figure 7. Same as Fig. 6 but for Scal metallicity gradients as a function of N2S2 metallicity gradients. In this example, the difference in slope is even greater.

different metallicity diagnostics are consistent. We provide all the linear conversions between significant diagnostic pairs in Table 2.

5 DISCUSSION

5.1 Comparing previous method

The purpose of empirically fitting metallicity gradients from different diagnostics is to allow us to freely compare metallicity gradients obtained from different methods. Prior to this study, a logical approach to this problem would be to convert individual spaxel metallicities using the conversion functions presented in either Kewley & Ellison (2008) or De Vis et al. (2019) and remeasure the metallicity gradient with the converted metallicities, we refer to this as ‘spaxel-converted gradients’. To compare this solution to the one outlined in this study, we convert the P04O3N2 metallicities to P04N2H α and N2O2 metallicities using the polynomial functions outlined in Kewley & Ellison (2008) and measure their metallicity gradients. We then compare the residuals between the spaxel-converted gradients and the metallicity gradients derived from the traditional emission lines (i.e. deriving the P04N2H α metallicity gradients using the [N II]/H α emission line ratio rather than converting from P04O3N2).

Fig. 8 shows the distribution of the difference between the metallicity gradients derived from the traditional emission lines and the spaxel-converted metallicity gradients in blue, when converting from P04O3N2 to P04N2H α metallicity gradients. Although the mode of the distribution is close to 0, there is a significant standard deviation of 0.09 dex/ R_e and an obvious positive skewness. We then compare the spaxel-converted distribution with the distribution of error obtained by using our empirically converted metallicity gradients in red. The standard deviation of the errors is significantly reduced and the positive skewness disappears. The gradient conversion method produces a higher concentration of galaxies with correctly converted metallicity gradients (error of 0) with fewer significant outliers on the wings. Another example of Fig. 8 is shown in Fig. 9, where instead of a decrease in scatter, we correct for the systematic offset present when converting between the P04O3N2 and the N2O2 metallicity gradients. Although the scatter is comparable between the two distributions, there is a clear systematic shift when using the spaxel-converted metallicity gradients.

5.2 Comparing Pettini & Pagel (2004) and Marino et al. (2013) calibrations

The N2H α and O3N2 line ratios are amongst the most popular diagnostics used to derive metallicity because the wavelength separation of the emission lines is small enough to minimize the effect of dust reddening. The strength and wavelength proximity of the emission lines used in the N2H α and O3N2 diagnostics make them relatively easy to observe at multiple redshifts. This convenience has resulted in many calibrations to the diagnostic (e.g. Pettini & Pagel 2004; Maiolino et al. 2008; Marino et al. 2013), each of which provide slightly different estimates of the metallicity. In this work we have chosen the N2H α and O3N2 calibrations outlined in Pettini & Pagel (2004) and Marino et al. (2013) due to their popularity. The calibration by Pettini & Pagel (2004) has been used extensively in metallicity studies, especially at high redshift where extinction correction can be difficult. Recently, the calibrations by Marino et al. (2013) have become the preferred calibration for N2H α and O3N2 diagnostics due to their better reliability at high metallicities.

From Table 1, we can see that both calibrations of the N2H α and O3N2 diagnostics produce extremely similar Spearman rank

coefficient, with the largest difference of 0.03 occurring between the N2H α and Scal. Table 2 shows the root mean square (rms) error of each of the linear fits applied to the pair of diagnostics with Spearman ranks greater than 0.6. For both the N2H α and O3N2 diagnostics, the calibrations by Marino et al. (2013) agree significantly better than those by Pettini & Pagel (2004) for all the metallicity diagnostics chosen for this study. We therefore recommend that the Marino et al. (2013) calibration be used over the Pettini & Pagel (2004) calibration to allow for more reliable comparisons to other diagnostics. Users of the N2H α and O3N2 metallicity diagnostics should also be aware of their limitations at low metallicities when being applied to high-redshift objects (López-Sánchez et al. 2012).

5.3 Comparing metallicity diagnostics

Pilyugin et al. (2010) provided two empirically calibrated metallicity diagnostics, ONS and ON. From Table 1, we can see that the ON metallicity diagnostic can only be converted to the ONS diagnostic (Spearman rank = 0.78). This similarity is expected as Pilyugin et al. (2010) found that the ON diagnostic was comparable to the ONS diagnostic across all classes (cool, warm, and hot) of H II regions. Although the ON diagnostic may be comparable to the ONS diagnostic, the addition of the [S II] emission line into the ONS diagnostic allows it be converted to O3N2 and N2O2 metallicity gradients, giving the ONS diagnostic more versatility in terms of being able to compare it to metallicity gradients from other studies.

In addition, the calibrations for the ON and ONS diagnostic are given in three separate equations for each class of H II region. To distinguish these classes, Pilyugin et al. (2010) used the [N II]/H β and [N II]/[S II] emission line ratios. These definitions mean that although the ON calibration does not directly use the [S II] emission line to derive the metallicity, it is still required to distinguish which H II regime the emission lines fall under. It is therefore better to use the ONS diagnostic because of its ability to be easily compared to the O3N2 and N2O2 diagnostic without the need for extra information over the ON diagnostic.

As mentioned in Section 1, there are generally two methods for calibrating metallicity diagnostics, the direct temperature method, or through photoionization models. Scal and Rcal are calibrated using the counterpart method outlined in Pilyugin et al. (2012). The counterpart method involves matching the observed H II region with that of a reference sample containing H II regions of known metallicity. The metallicity of the reference sample used by Pilyugin et al. (2012) is determined through the direct temperature method. Therefore, we classify the Scal and Rcal metallicity diagnostics as being calibrated through the direct temperature method.

The Scal and Rcal metallicity diagnostics differ only by replacing the [S II] emission line in the Scal diagnostic with the [O II] emission line in the Rcal diagnostic. This simple change effectively highlights that diagnostics that utilize similar emission line ratios tend to agree better than those that use different sets of emission lines. The Scal and Rcal diagnostics can be converted to the N2H α , N2O2, N2S2H α diagnostics, as well as to the machine learning algorithm and to each other. Additionally, with the [O II] emission line, the K04R₂₃ calibration can be converted to the Rcal diagnostic. Similarly, The N2S2H α diagnostic provides a lower rms error when converting to the Scal diagnostic compared to the Rcal diagnostic. Interestingly, the N2O2 diagnostic has a higher Spearman rank with the Rcal diagnostic but a lower rms error with the Scal diagnostic. This apparent discrepancy could be due to the sensitivity of the Spearman

Table 2. Linear conversion fits to convert metallicity gradients from one metallicity diagnostic to another. The conversion relations follow the general form of $y = mx + c$, where x is the base metallicity that can be calculated through available emission lines, and y is the final metallicity that you wish to convert to. For completeness, we have included redundant metallicity conversions (e.g. O3N2 to N2H α) by calculating the inverse function (shown in red), but advise against their usage as simply calculating the metallicity gradient with the simpler metallicity diagnostic will yield better results.

Base metallicity (x)	Final metallicity (y)	K04R ₂₃	P04N2H α	M13N2H α	P04O3N2	M13O3N2	N2O2	N2S2H α	ONS	ON	Scal	Rcal	MLP	HII-CHI	C17R ₂₃
K04R ₂₃ (Photoionization)	m (Gradient)	–	–	–	1.49	0.73	1.04	–	–	–	–	0.50	–	–	–
	c (Intercept)	–	–	–	0.01	0.00	–0.02	–	–	–	–	–0.02	–	–	–
	rms scatter	–	–	–	0.047	0.034	0.080	–	–	–	–	0.036	–	–	–
P04N2H α (Mixed)	m (Gradient)	–	–	–	1.29	0.55	–	–	–	–	0.56	0.65	1.20	–	–
	c (Intercept)	–	–	–	0.01	–0.01	–	–	–	–	–0.02	–0.02	–0.01	–	–
	rms scatter	–	–	–	0.034	0.026	–	–	–	–	0.024	0.029	0.030	–	–
M13N2H α (T_e)	m (Gradient)	–	–	–	2.87	2.04	–	–	–	–	1.39	1.39	2.03	–	–
	c (Intercept)	–	–	–	0.03	0.02	–	–	–	–	–0.01	–0.01	–0.02	–	–
	rms scatter	–	–	–	0.027	0.026	–	–	–	–	0.019	0.019	0.015	–	–
P04O3N2 (Mixed)	m (Gradient)	0.67	0.78	0.35	–	–	0.95	–	0.53	–	–	–	1.62	0.46	–
	c (Intercept)	–0.01	–0.01	–0.01	–	–	–0.02	–	–0.01	–	–	–	0.01	0.02	–
	rms scatter	0.047	0.034	0.027	–	–	0.065	–	0.034	–	–	–	0.041	0.036	–
M13O3N2 (T_e)	m (Gradient)	1.37	1.83	0.49	–	–	1.35	–	2.49	–	–	–	2.39	1.22	1.79
	c (Intercept)	0.00	0.00	–0.01	–	–	–0.01	–	0.03	–	–	–	0.01	0.03	0.02
	rms scatter	0.034	0.026	0.026	–	–	0.032	–	0.030	–	–	–	0.027	0.031	0.033
N2O2 (Photoionization)	m (Gradient)	0.96	–	–	1.06	0.74	–	1.67	0.66	–	0.55	0.78	1.34	–	0.79
	c (Intercept)	0.02	–	–	0.02	0.01	–	0.01	0.00	–	–0.01	0.01	0.02	–	0.01
	rms scatter	0.080	–	–	0.065	0.032	–	0.078	0.054	–	0.049	0.062	0.077	–	0.071
N2S2H α (Photoionization)	m (Gradient)	–	–	–	–	–	0.60	–	–	–	0.53	0.42	0.58	–	–
	c (Intercept)	–	–	–	–	–	–0.01	–	–	–	0.00	–0.01	–0.01	–	–
	rms scatter	–	–	–	–	–	0.078	–	–	–	0.017	0.036	0.043	–	–
ONS (T_e)	m (Gradient)	–	–	–	1.89	0.40	1.51	–	–	1.03	–	–	–	–	–
	c (Intercept)	–	–	–	0.02	–0.01	0.00	–	–	0.00	–	–	–	–	–
	rms scatter	–	–	–	0.034	0.030	0.054	–	–	0.042	–	–	–	–	–
ON (T_e)	m (Gradient)	–	–	–	–	–	–	–	0.97	–	–	–	–	–	–
	c (Intercept)	–	–	–	–	–	–	–	0.00	–	–	–	–	–	–
	rms scatter	–	–	–	–	–	–	–	0.042	–	–	–	–	–	–
Scal (T_e)	m (Gradient)	–	1.80	0.72	–	–	1.83	1.89	–	–	–	1.18	1.71	–	–
	c (Intercept)	–	0.01	0.01	–	–	0.01	0.00	–	–	–	0.00	0.01	–	–
	rms scatter	–	0.024	0.019	–	–	0.049	0.017	–	–	–	0.024	0.023	–	–
Rcal (T_e)	m (Gradient)	1.99	1.53	0.72	–	–	1.29	2.35	–	–	0.84	–	1.40	–	–
	c (Intercept)	0.02	0.02	0.01	–	–	0.00	0.01	–	–	0.00	–	0.00	–	–
	rms scatter	0.036	0.029	0.019	–	–	0.062	0.036	–	–	0.024	–	0.021	–	–
MLP (T_e)	m (Gradient)	–	0.83	0.49	0.62	0.42	0.74	1.71	–	–	0.59	0.71	–	–	–
	c (Intercept)	–	0.01	0.01	–0.01	0.00	–0.02	0.02	–	–	–0.01	0.00	–	–	–
	rms scatter	–	0.030	0.015	0.041	0.027	0.077	0.043	–	–	0.023	0.021	–	–	–
HII-CHI (Photoionization)	m (Gradient)	–	–	–	2.16	0.82	–	–	–	–	–	–	–	–	–
	c (Intercept)	–	–	–	–0.03	–0.03	–	–	–	–	–	–	–	–	–
	rms scatter	–	–	–	0.036	0.031	–	–	–	–	–	–	–	–	–
C17R ₂₃ (T_e)	m (Gradient)	–	–	–	–	0.56	1.27	–	–	–	–	–	–	–	–
	c (Intercept)	–	–	–	–	–0.01	–0.01	–	–	–	–	–	–	–	–
	rms scatter	–	–	–	–	0.033	0.071	–	–	–	–	–	–	–	–

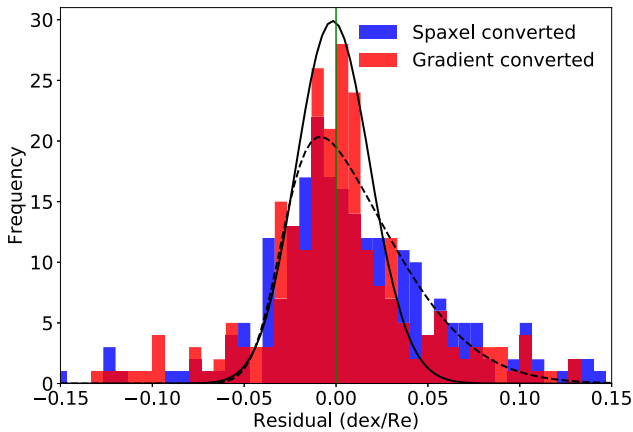


Figure 8. Distribution of residuals when comparing spaxel-converted (blue) and empirically converted metallicity gradients (red) to the true metallicity gradient. The fitted skewed Gaussians (solid line for red and dashed line for blue) clearly show the reduction in residual dispersion when converting P04O3N2 metallicity gradients empirically to P04N2H α instead of spaxel-converted metallicity gradients.

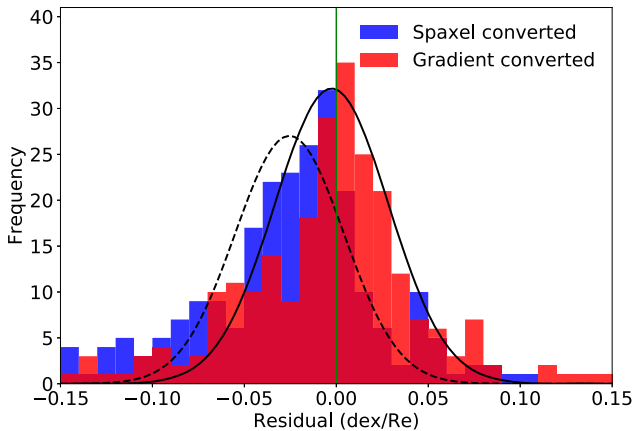


Figure 9. Same as Fig. 8 but converting the P04O3N2 sector metallicities to N2O2. In this example, an offset is corrected for by converting metallicity gradients directly using the conversion relations presented here.

rank to outliers that are being removed during the linear-fitting process.

From Table 2, we see that the diagnostics calibrated through the direct temperature always have a lower rms error than diagnostics calibrated through photoionization models regardless of whether or not the reference diagnostic was calibrated by direct temperature or through photoionization models. The exception to this is between the Scal and N2S2H α diagnostics, which agree remarkably well.

It is unsurprising that when comparing direct temperature calibrations to other direct temperature calibrations, we obtain a lower rms error compared to diagnostics calibrated through photoionization models. This is partially due to several diagnostics sharing a considerable fraction of H II regions as calibration points. However, the diagnostics calibrated through photoionization models have a larger rms error when compared to other photoionization model calibrations relative to the direct temperature calibrations. The obvious answer to this problem is that different photoionization models with different assumptions were used to calibrate the diagnostics.

For example, the N2O2 diagnostic was calibrated through MAPPINGS III photoionization models (Sutherland et al. 2013) while the

N2S2H α diagnostics was calibrated through MAPPINGS V photoionization models (Sutherland et al. 2018). Significant changes have been made between MAPPINGS III and MAPPINGS V, that may cause the discrepancies we see here. Differences in model conditions may also contribute to the differences we see, such as pressure and electron temperature. The large number of free parameters available to theoretical models is likely the cause of the rms error seen between N2O2 and N2S2H α .

However, different versions of MAPPINGS does not fully explain the scatter between the K04R₂₃ and N2O2 diagnostic. Although these two diagnostics were calibrated by different authors (Kewley & Dopita 2002; Kobulnicky & Kewley 2004), they both use the same photoionization grids outlined in Kewley & Dopita (2002). This highlights that even if consistent photoionization models are used, different metallicity diagnostics produce greater scatter compared to direct temperature methods.

5.4 Machine learning versus emission line ratio

We see from Table 1 that MLP is convertible (Spearman's rank > 0.6) for diagnostics that include the N2 ([N II]/H α or [N II]/H β) emission line ratio as well as the N2O2 diagnostic. Although the N2O2 diagnostic can be converted to MLP, the rms scatter indicates that it does not perform as well as the other diagnostics that can also be converted to MLP (i.e. N2H α , O3N2, N2S2H α , Scal, and Rcal). MLP also does significantly better with the M13 calibrations compared to the P04 calibrations. For both N2H α and O3N2 diagnostics, MLP is amongst the worst when compared to the P04 calibration, but has the smallest rms error for the M13N2H α calibration and is only marginally (0.01) beaten by the N2 calibrations for M13O3N2. MLP performing better for the M13 calibration rather than the P04 calibration is consistent with the testing performed by Ho (2019) when developing MLP.

Diagnostics that do not include the [N II]/H α emission line ratio (R₂₃, ONS, and ON) do not correlate well with MLP with the exception of N2O2 in which MLP is outperformed by a number of other diagnostics. This dependence on the [N II]/H α emission line ratio arises due to the large variation of the ratio with metallicity, visible in the classic BPT diagram. This large variation is caused by the secondary nucleosynthetic pathway of nitrogen (see e.g. Nicholls et al. 2017).

5.5 Metallicity gradient–mass relation

One of the most well-known galaxy correlations is the mass–metallicity relation (Tremonti et al. 2004). In the era of integral field spectroscopy and spatially resolved metallicity maps, this correlation has evolved into the metallicity gradient–mass relation. The main point of contention with the metallicity gradient–mass relation is whether or not metallicity gradients evolve as a function of mass. Some studies show the existence of a characteristic metallicity gradient once it has been normalized to the size of the disc, i.e. the metallicity gradient is the same regardless of the mass (Sánchez et al. 2012, 2014; Ho et al. 2015; Sánchez-Menguiano et al. 2016, 2018). Others show that the metallicity gradients become steeper for higher mass galaxies up to a stellar mass of about $10^{9.5} - 10^{10.5} M_{\odot}$. Beyond a mass of $10^{10.5} M_{\odot}$, a break in the metallicity gradient–mass relation appears and the metallicity gradients become shallower with increasing mass (Belfiore et al. 2017).

Several studies have proposed and clearly shown that single linear fits do not accurately represent the metallicity as a function of radius

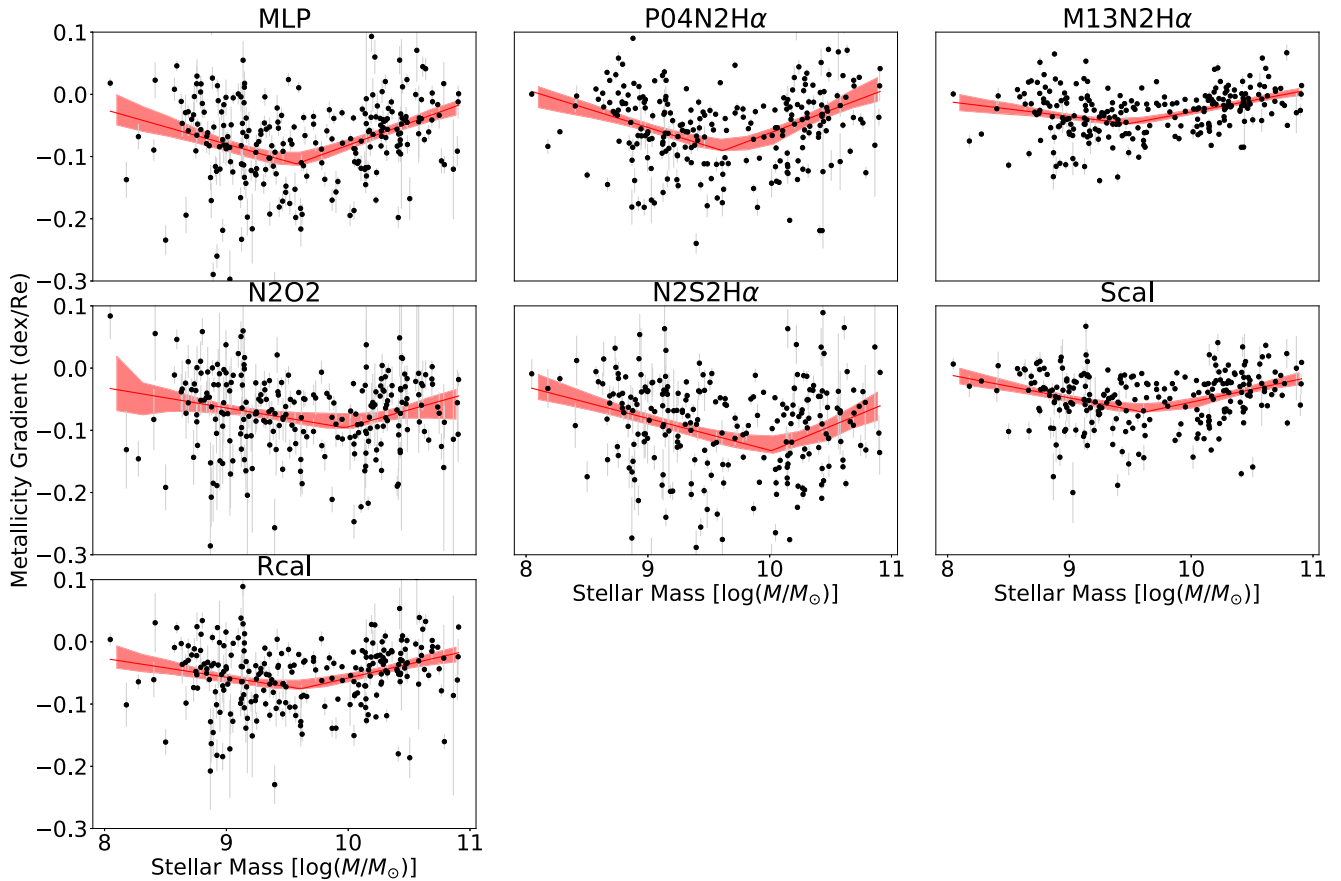


Figure 10. Metallicity gradient–mass relation for calibrations that can be converted to Scal. We apply a broken linear-fit to each of the panels, allowing the location of the break to vary between diagnostics. Red bands indicate the 1σ uncertainties of the broken linear-fits. Even though we do not force a break in the linear trend, each of the panels shows a broken linear-fit where the measured gradient switches from negative to positive for galaxies with stellar masses between $10^{9.5}$ and $10^{10.5} M_{\odot}$.

(Vila-Costas & Edmunds 1992; Sánchez et al. 2012, 2014; Sánchez-Menguiano et al. 2016; Belfiore et al. 2017). Using MUSE data, Sánchez-Menguiano et al. (2018) showed the presence of inner and outer flattening of the metallicity radial gradient, especially for high-mass galaxies. This flattening suggests that a broken linear-fit may be a better method for characterizing metallicity radial gradients. The inner and outer flattening of massive galaxies may be artificially flattening the overall metallicity radial profile when fit with single linear gradients.

Fig. 10 shows the metallicity gradient–mass relation of seven different metallicity calibrations, each of which can be converted to the Scal metallicity diagnostic based on the linear fits provided in this paper. We perform a broken linear regression to each of the panels. We require the broken linear-fit to be continuous, but allow the position of the break to be a free variable. We find that for each of the calibrations, a break in the metallicity gradient–mass relation occurs between $10^{9.5}$ and $10^{10.5} M_{\odot}$. As we are not directly forcing a break in the linear trend while fitting, this indicates a consistent trend in the metallicity gradient–mass relation when using single linear fits to metallicity gradients. Due to the relatively low spatial resolution of our data, we are unable to test how this trend changes if a broken linear-fit is used to fit metallicity gradients.

We summarize Fig. 10 in the left-hand panel of Fig. 11, where we show the broken linear-fits to the metallicity gradient–mass

relation of each calibration. We can see that the metallicity gradient–mass relation varies substantially depending on which metallicity calibration is used. We have chosen these diagnostics specifically as they can be converted to the Scal diagnostic based on the analysis presented in this paper. We have chosen Scal to be our fiducial metallicity diagnostic because apart from MLP and NB, it requires the availability of the majority of the strong optical emission lines alongside Rcal and ONS. Selecting a fiducial metallicity diagnostic that requires a large number of emission lines is important because the only situation where these conversions should be used is when going from a basic (in terms of how many emissions lines are required) metallicity diagnostic such as $N2H\alpha$ to a more complicated one. Although the ONS diagnostic uses more emission lines, the majority of metallicity diagnostics do not agree with its calculations, meaning it would not be a useful case study. In the right-hand panel of Fig. 11, we convert each of the metallicity gradient–mass relations to the Scal diagnostic. By applying the metallicity calibration conversions, we reduce the maximum deviations between calibrations from $0.11 \text{ dex}/R_e$ to $0.02 \text{ dex}/R_e$, a significant improvement. The reduction in maximum deviation demonstrates the effectiveness of using the metallicity calibrations conversions to convert from single emission line ratio diagnostics ($N2H\alpha$ and $N2O2$) to multiple emission line ratio diagnostics.

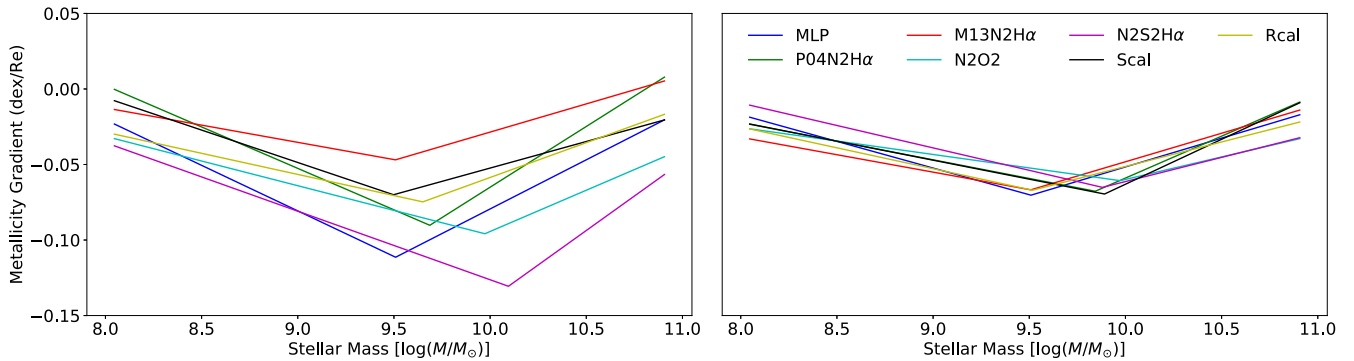


Figure 11. Left: We take each of the metallicity gradient-mass fits from Fig. 10 and put them all on the same axis, highlighting the significant differences that we find in this relation depending on which diagnostic is used. The maximum difference between the different metallicity gradient-mass fits is $0.11 \text{ dex}/R_e$. Right: We convert each of the metallicity gradients to an Scal metallicity gradient based on the conversion factors in Table 2. We then re-fit the metallicity gradient-mass relation using the converted metallicity gradients. One can see that the dispersion from diagnostic to diagnostic is significantly reduced once we convert all the metallicity gradients to Scal. The maximum difference is reduced from $0.11 \text{ dex}/R_e$ to only $0.02 \text{ dex}/R_e$.

6 SUMMARY

Using SAMI DR2 sector-binned cubes, we create metallicity maps and measure their metallicity gradient using 13 popular emission line diagnostics and techniques. We investigate the differences in measured metallicity gradients and determine if we can accurately compare metallicity gradients obtained through different methods. For pairs of diagnostics with a Spearman's rank > 0.6 , we provide empirical fits to allow for the conversion between the two diagnostics. The ability to convert between diagnostics is important for high-redshift studies where a wide range of emission lines are unavailable and there is a need to compare high-redshift metallicity gradients to local galaxies.

We find that the relationship between most of the diagnostics is accurately represented as a linear function with a non-unity slope. The non-unity slope demonstrates that one cannot expect metallicity gradients to be the same across different diagnostics, demonstrating the importance of the conversion functions we provide in this paper. We also find that the zero-point of the linear relations is close to 0, meaning that flat metallicity gradients remain as flat gradients across the diagnostics tested in this study.

We compare two common calibrations of the $\text{N2H}\alpha$ and O3N2 emission line diagnostics by Pettini & Pagel (2004) and Marino et al. (2013) to determine which calibration provides more flexibility in terms of comparing to other metallicity gradients. We find that the calibrations by Marino et al. (2013) provide better agreement than the calibrations by Pettini & Pagel (2004) across all the diagnostics tested here. Better agreement is likely due to the enhanced accuracy of high-metallicity measurements used by Marino et al. (2013) for calibration, which were not available to Pettini & Pagel (2004) at the time of writing.

As expected, emission line diagnostics that share similar emission line ratios tend to agree better than those that use different sets of emission lines. For example, the Rcal diagnostic by Pilyugin & Grebel (2016) provide a lower rms error when converting to the K04R_{23} and N2O2 diagnostic compared to the Scal diagnostic, which replaces the $[\text{O II}]$ emission line for the $[\text{S II}]$ emission. Similarly, the Scal diagnostic provides a better agreement with the $\text{N2S2H}\alpha$ diagnostic compared to the Rcal diagnostic.

The use of machine learning to measure gas-phase metallicity has only recently emerged (Ho 2019; Wu & Boada 2019). We find that OxygenMLP (Ho 2019) agrees well with diagnostics which include

either the $[\text{N II}]/\text{H}\alpha$ or $[\text{N II}]/\text{H}\beta$ emission line ratios. This strongly suggests that the N2 emission line ratio is one of the strongest tracers of gas-phase metallicity.

Our analysis highlights the stark differences present between various metallicity diagnostics and the uncertainties of comparing metallicity diagnostics derived through different methods. When comparing metallicity gradients, we recommend using the same methods as the original author (where possible) in order to reduce systematic errors. If the same emission lines are not available, Table 2 can be used to convert metallicity gradients from one diagnostic to another. We find that directly converting the metallicity gradients empirically provides a closer estimate of the desired metallicity gradient than individually converting spaxels. These conversion relations will be useful tools for analysing the gas-phase metallicities of galaxies at high redshift, where the availability of a wide range of emissions lines is scarce.

ACKNOWLEDGEMENTS

The SAMI Galaxy Survey is based on observations made at the Anglo-Australian Telescope. The Sydney-AAO Multi-object Integral field spectrograph (SAMI) was developed jointly by the University of Sydney and the Australian Astronomical Observatory. The SAMI input catalogue is based on data taken from the Sloan Digital Sky Survey, the GAMA Survey and the VST ATLAS Survey. The SAMI Galaxy Survey is supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013, the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020, and other participating institutions. The SAMI Galaxy Survey website is <http://sami-survey.org/>.

BG gratefully acknowledges the support of the Australian Research Council as the recipient of a Future Fellowship (FT140101202).

SB acknowledges the funding support from the Australian Research Council through a Future Fellowship (FT140101166).

LC is the recipient of an Australian Research Council Future Fellowship (FT180100066) funded by the Australian Government.

JvdS is funded under Bland-Hawthorn's ARC Laureate Fellowship (FL140100278).

JJB acknowledges support of an Australian Research Council Future Fellowship (FT180100231).

MSO acknowledges the funding support from the Australian Research Council through a Future Fellowship (FT140100255).

NS acknowledges support of an Australian Research Council Australian Discovery Early Career Award (project number DE190100375) funded by the Australian Government and a University of Sydney Postdoctoral Research Fellowship.

DATA AVAILABILITY

The data underlying this article are publicly released in the SAMI Data Release 2 data set, available through <https://datacentral.org.au/>.

REFERENCES

- Alloin D., Collin-Souffrin S., Joly M., Vigroux L., 1979, *A&A*, 78, 200
- Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
- Barrera-Ballesteros J. K., Sánchez S. F., Heckman T., Blanc G. A., MaNGA Team, 2017, *ApJ*, 844, 80
- Belfiore F. et al., 2017, *MNRAS*, 469, 151
- Blanc G. A., Kewley L., Vogt F. P. A., Dopita M. A., 2015, *ApJ*, 798, 99
- Bland-Hawthorn J. et al., 2011, *Opt. Express*, 19, 2649
- Boettcher E., Gallagher J. S., III, Zweibel E. G., 2017, *ApJ*, 845, 155
- Bryant J. J., Bland-Hawthorn J., Fogarty L. M. R., Lawrence J. S., Croom S. M., 2014, *MNRAS*, 438, 869
- Bryant J. J. et al., 2015, *MNRAS*, 447, 2857
- Cappellari M., 2017, *MNRAS*, 466, 798
- Cappellari M., Emsellem E., 2004, *PASP*, 116, 138
- Cappellari M. et al., 2013, *MNRAS*, 432, 1709
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Castellanos M., Díaz A. I., Terlevich E., 2002a, *MNRAS*, 329, 315
- Castellanos M., Díaz A. I., Terlevich E., 2002b, *MNRAS*, 337, 540
- Charlot S., Longhetti M., 2001, *MNRAS*, 323, 887
- Croom S. M. et al., 2012, *MNRAS*, 421, 872
- Curti M., Cresci G., Mannucci F., Marconi A., Maiolino R., Esposito S., 2017, *MNRAS*, 465, 1384 (C17R)
- D'Agostino J. J., Poetrodjojo H., Ho I.-T., Groves B., Kewley L., Madore B. F., Rich J., Seibert M., 2018, *MNRAS*, 479, 4907
- D'Agostino J. J., Kewley L. J., Groves B. A., Medling A., Dopita M. A., Thomas A. D., 2019a, *MNRAS*, 485, L38
- D'Agostino J. J. et al., 2019b, *MNRAS*, 487, 4153
- Davies R. L., Rich J. A., Kewley L. J., Dopita M. A., 2014, *MNRAS*, 439, 3835
- Davies R. L. et al., 2016, *MNRAS*, 462, 1616
- Davies R. L. et al., 2017, *MNRAS*, 470, 4974
- De Vis P. et al., 2019, *A&A*, 623, A5
- Denicolò G., Terlevich R., Terlevich E., 2002, *MNRAS*, 330, 69
- Díaz A. I., Terlevich E., Vilchez J. M., Pagel B. E. J., Edmunds M. G., 1991, *MNRAS*, 253, 245
- Dopita M. A., Evans I. N., 1986, *ApJ*, 307, 431
- Dopita M. A., Sutherland R. S., Nicholls D. C., Kewley L. J., Vogt F. P. A., 2013, *ApJS*, 208, 10
- Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, *Ap&SS*, 361, 61 (D16)
- Driver S. P. et al., 2011, *MNRAS*, 413, 971
- Edmunds M. G., Pagel B. E. J., 1984, *MNRAS*, 211, 507
- Erroz-Ferrer S. et al., 2019, *MNRAS*, 484, 5009
- Ferguson A. M. N., Wyse R. F. G., Gallagher J. S., 1996, *AJ*, 112, 2567
- Ferland G. J. et al., 2017, *RMxAA*, 53, 385
- Garnett D. R., Shields G. A., 1987, *ApJ*, 317, 82
- Green A. W. et al., 2018, *MNRAS*, 475, 716
- Greenawalt B., Walterbos R. A. M., Thilker D., Hoopes C. G., 1998, *ApJ*, 506, 135
- Groves B. A., Dopita M. A., Sutherland R. S., 2004, *ApJS*, 153, 75
- Hinshaw G. et al., 2009, *ApJS*, 180, 225
- Ho I.-T., 2019, *MNRAS*, 485, 3569
- Ho I.-T., Kudritzki R.-P., Kewley L. J., Zahid H. J., Dopita M. A., Bresolin F., Rupke D. S. N., 2015, *MNRAS*, 448, 2030
- Ho I.-T. et al., 2016, *Ap&SS*, 361, 280
- Hoopes C. G., Walterbos R. A. M., Greenwalt B. E., 1996, *AJ*, 112, 1429
- Hubble E. P., 1926, *ApJ*, 64, 321
- 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
- Kobulnicky H. A., Kewley L. J., 2004, *ApJ*, 617, 240
- Kobulnicky H. A., Skillman E. D., 1997, *ApJ*, 489, 636
- Kumari N., Maiolino R., Belfiore F., Curti M., 2019, *MNRAS*, 485, 367
- Lacerda E. A. D. et al., 2018, *MNRAS*, 474, 3727
- Lee H., Skillman E. D., 2004, *ApJ*, 614, 698
- Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, *A&A*, 80, 155
- Leroy A. M., Rousseeuw P. J., 1987, *Robust Regression and Outlier Detection*. Wiley Series in Probability and Mathematical Statistics, New York
- López-Sánchez Á. R., Esteban C., 2010, *A&A*, 517, A85
- López-Sánchez Á. R., Dopita M. A., Kewley L. J., Zahid H. J., Nicholls D. C., Scharwächter J., 2012, *MNRAS*, 426, 2630
- López-Sánchez Á. R., Westmeier T., Esteban C., Koribalski B. S., 2015, *MNRAS*, 450, 3381
- Maiolino R. et al., 2008, *A&A*, 488, 463
- Marino R. A. et al., 2013, *A&A*, 559, A114 (M13)
- Markwardt C. B., 2009, in Bohlender D. A., Durand D., Dowler P., eds, *ASP Conf. Ser. Vol. 411, Astronomical Data Analysis Software and Systems XVIII*. Astron. Soc. Pac., San Francisco, p. 251
- Mast D. et al., 2014, *A&A*, 561, A129
- McCall M. L., Rybski P. M., Shields G. A., 1985, *ApJS*, 57, 1
- McGaugh S. S., 1991, *ApJ*, 380, 140
- Medling A. M. et al., 2018, *MNRAS*, 475, 5194
- Moustakas J., Kennicutt R. C., Jr, Tremonti C. A., Dale D. A., Smith J.-D. T., Calzetti D., 2010, *ApJS*, 190, 233
- Nicholls D. C., Sutherland R. S., Dopita M. A., Kewley L. J., Groves B. A., 2017, *MNRAS*, 466, 4403
- Oey M. S. et al., 2007, *ApJ*, 661, 801
- Osterbrock D. E., 1989, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*. University Science Books, Sausalito, California
- Owers M. S. et al., 2017, *MNRAS*, 468, 1824
- Pagel B. E. J., Edmunds M. G., Blackwell D. E., Chun M. S., Smith G., 1979, *MNRAS*, 189, 95
- Pagel B. E. J., Edmunds M. G., Smith G., 1980, *MNRAS*, 193, 219
- Peng Y.-J. et al., 2010, *ApJ*, 721, 193
- Pérez-Montero E., 2014, *MNRAS*, 441, 2663
- Pettini M., Pagel B. E. J., 2004, *MNRAS*, 348, L59 (P04)
- Pilyugin L. S., 2000, *A&A*, 362, 325
- Pilyugin L. S., 2001, *A&A*, 374, 412
- Pilyugin L. S., Grebel E. K., 2016, *MNRAS*, 457, 3678
- Pilyugin L. S., Thuan T. X., 2005, *ApJ*, 631, 231
- Pilyugin L. S., Vilchez J. M., Thuan T. X., 2010, *ApJ*, 720, 1738
- Pilyugin L. S., Grebel E. K., Mattsson L., 2012, *MNRAS*, 424, 2316
- Poetrodjojo H. et al., 2018, *MNRAS*, 479, 5235
- Poetrodjojo H., D'Agostino J. J., Groves B., Kewley L., Ho I. T., Rich J., Madore B. F., Seibert M., 2019, *MNRAS*, 487, 79
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, *Numerical Recipes in FORTRAN. The Art of Scientific Computing*. Cambridge Univ. Press, New York
- Rich J. A., Torrey P., Kewley L. J., Dopita M. A., Rupke D. S. N., 2012, *ApJ*, 753, 5
- Richards S. N. et al., 2014, *MNRAS*, 445, 1104
- Rupke D. S. N., Kewley L. J., Chien L.-H., 2010, *ApJ*, 723, 1255
- Sánchez S. F. et al., 2012, *A&A*, 546, A2

- Sánchez S. F. et al., 2014, *A&A*, 563, A49
 Sánchez S. F. et al., 2017, *MNRAS*, 469, 2121
 Sánchez S. F. et al., 2019, *MNRAS*, 484, 3042
 Sánchez-Menguiano L. et al., 2016, *A&A*, 587, A70
 Sánchez-Menguiano L. et al., 2018, *A&A*, 609, A119
 Scott N. et al., 2018, *MNRAS*, 481, 2299
 Sharp R. et al., 2006, in McLean I. S., Iye M., eds, Proc. SPIE Conf. Ser. Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy. SPIE, Bellingham, p. 62690G
 Storchi-Bergmann T., Calzetti D., Kinney A. L., 1994, *ApJ*, 429, 572
 Sutherland R., Dopita M., Binette L., Groves B., 2013, Astrophysics Source Code Library, record ascl:1306.008
 Sutherland R., Dopita M., Binette L., Groves B., 2018, Astrophysics Source Code Library, record ascl:1807.005
 Sweet S. M., Drinkwater M. J., Meurer G., Bekki K., Dopita M. A., Kilborn V., Nicholls D. C., 2014, *ApJ*, 782, 35
 Thomas A. D., Kewley L. J., Dopita M. A., Groves B. A., Hopkins A. M., Sutherland R. S., 2018, *ApJ*, 861, L2
 Torres-Peimbert S., Peimbert M., Fierro J., 1989, *ApJ*, 345, 186
 Tremonti C. A. et al., 2004, *ApJ*, 613, 898
 Vale Asari N., Couto G. S., Cid Fernandes R., Stasińska G., de Amorim A. L., Ruschel-Dutra D., Werle A., Florido T. Z., 2019, *MNRAS*, 489, 4721
 van de Sande J. et al., 2017, *ApJ*, 835, 104
 Vazdekis A., Ricciardelli E., Cenarro A. J., Rivero-González J. G., Díaz-García L. A., Falcón-Barroso J., 2012, *MNRAS*, 424, 157
 Vila-Costas M. B., Edmunds M. G., 1992, *MNRAS*, 259, 121
 Walterbos R. A. M., Braun R., 1994, *ApJ*, 431, 156
 Wu J. F., Boada S., 2019, *MNRAS*, 484, 4683
 Yin S. Y., Liang Y. C., Hammer F., Brinchmann J., Zhang B., Deng L. C., Flores H., 2007, *A&A*, 462, 535
 York D. G. et al., 2000, *AJ*, 120, 1579
 Yuan T.-T., Kewley L. J., Rich J., 2013, *ApJ*, 767, 106
 Zaritsky D., Kennicutt R. C., Jr, Huchra J. P., 1994, *ApJ*, 420, 87
 Zhang K. et al., 2017, *MNRAS*, 466, 3217

APPENDIX A: METALLICITY DIAGNOSTICS AND CALIBRATIONS

Here, we provide the functional forms of all the metallicity calibrations used in this study.

A1 K04R₂₃ (Kobulnicky & Kewley 2004)

The first step involves assigning an initial metallicity guess using the $[\text{N II}]\lambda 6583/[\text{O II}]\lambda 3726, \lambda 3729$ (N2O2) emission line ratio. If $\log([\text{N II}]/[\text{O II}]) < -1.2$, then we place the spaxel on the lower branch and assume an initial $12 + \log(\text{O/H}) = 8.2$. If $\log([\text{N II}]/[\text{O II}]) > -1.2$, then we place the spaxel on the upper branch and assume an initial $12 + \log(\text{O/H}) = 8.7$. The initial estimate of metallicity is then used to calculate the ionization parameter using equation (1)

$$\begin{aligned} \log(q) = & 32.81 - 1.153y^2 + [(O/H)^*] \\ & \times (-3.396 - 0.025y + 0.1444y^2) \\ & \times (4.603 - 0.3119y - 0.163y^2 + [(O/H)^*]) \\ & \times (-0.48 + 0.0271y + 0.02037y^2)^{-1}, \end{aligned} \quad (\text{A1})$$

where $y = \log([\text{O III}]\lambda\lambda 4959, 5007/[\text{O II}]\lambda\lambda 3726, 3729)$, and $(O/H)^* = 12 + \log(\text{O/H})$.

The initial estimate of ionization parameter is then used to calculate the metallicity using the $([\text{O II}]\lambda\lambda 3726, 3729 + [\text{O III}]\lambda\lambda 4959, 5007)/H\beta$ (R_{23}) emission line ratio with the relevant calibration depending on which branch the spaxel was initially assigned by the

N2O2 emission line ratio:

$$\begin{aligned} (O/H)_{\text{lower}}^* = & 9.40 + 4.65x - 3.17x^2 - \log(q) \\ & \times (0.272 + 0.547x - 0.513x^2) \\ (O/H)_{\text{upper}}^* = & 9.72 - 0.777x - 0.951x^2 - 0.072x^3 - 0.811x^4 \\ & - \log(q) \times (0.0737 - 0.0713x - 0.141x^2 \\ & + 0.0373x^3 - 0.058x^4), \end{aligned} \quad (\text{A2})$$

where $x = \log(R_{23})$ and $(O/H)^* = 12 + \log(\text{O/H})$.

The spaxel is then iterated between equations (1) and (2) until convergence is achieved within a predetermined tolerance. Three iterations are typically needed to reach a convergence level of 0.01 dex.

A2 N2O2 (Kewley & Dopita 2002)

$$\begin{aligned} (O/H)^* = & 1106.87 - 532.154x + 96.3733x^2 \\ & - 7.81061x^3 + 0.239282x^4, \end{aligned} \quad (\text{A3})$$

where $x = \log(\text{N2O2})$ and $(O/H)^* = 12 + \log(\text{O/H})$.

A3 N2H α (Pettini & Pagel 2004; Marino et al. 2013)

$$\begin{aligned} (O/H)_{P04}^* = & 9.37 + 2.03x + 1.26x^2 + 0.32x^3, \text{ where} \\ & -2.5 < x < -0.3 \\ (O/H)_{M13}^* = & 8.743 + 0.462x, \text{ where } -1.6 < x < -0.2, \end{aligned} \quad (\text{A4})$$

where $x = \log(\text{N2H}\alpha)$ and $(O/H)^* = 12 + \log(\text{O/H})$.

A4 O3N2 (Pettini & Pagel 2004; Marino et al. 2013)

$$\begin{aligned} (O/H)_{P04}^* = & 8.73 - 0.32x, \text{ where } -1.0 < x < 1.9 \\ (O/H)_{M13}^* = & 8.533 - 0.214x, \text{ where } -1.1 < x < 1.7, \end{aligned} \quad (\text{A5})$$

where $x = \log(\text{O3N2})$ and $(O/H)^* = 12 + \log(\text{O/H})$.

A5 N2S2H α (Dopita et al. 2016)

$$(O/H)_{D16}^* = 8.77 + x + 0.45(x + 0.3)^5, \text{ where } -1.1 < x < 0.5 \quad (\text{A6})$$

where $x = \log([\text{N II}]/[\text{S II}] + 0.264 \times \log(\text{N2H}\alpha))$ and $(O/H)^* = 12 + \log(\text{O/H})$.

A6 ONS and ON (Pilyugin et al. 2010)

$$\begin{aligned} (O/H)_{\text{ONS_Cool}}^* = & 8.277 + 0.657 \times P - 0.399 \times \log(\text{O3H}\beta) \\ & - 0.061 \times \log(\text{N2O2}) + 0.005 \times \log(\text{S2O2}) \\ (O/H)_{\text{ONS_Warm}}^* = & 8.816 - 1.855 \times P + 1.517 \times \log(\text{O3H}\beta) \\ & + 0.710 \times \log(\text{N2O2}) - 0.337 \times \log(\text{S2O2}) \\ (O/H)_{\text{ONS_Hot}}^* = & 8.774 - 1.855 \times P + 1.517 \times \log(\text{O3H}\beta) \\ & + 0.304 \times \log(\text{N2O2}) + 0.328 \times \log(\text{S2O2}) \end{aligned} \quad (\text{A7})$$

$$\begin{aligned}
(\text{O}/\text{H})_{\text{ON.Cool}}^* &= 8.606 - 0.105 \times \log(\text{O3H}\beta) - 0.410 \\
&\quad \times \log(\text{O2H}\beta) - 0.150 \times \log(\text{N2O2}) \\
(\text{O}/\text{H})_{\text{ON.Warm}}^* &= 8.642 + 0.077 \times \log(\text{O3H}\beta) + 0.411 \\
&\quad \times \log(\text{O2H}\beta) + 0.601 \times \log(\text{N2O2}) \\
(\text{O}/\text{H})_{\text{ON.Hot}}^* &= 8.013 + 0.905 \times \log(\text{O3H}\beta) + 0.602 \\
&\quad \times \log(\text{O2H}\beta) + 0.751 \times \log(\text{N2O2}), \quad (\text{A8})
\end{aligned}$$

where P is the excitation parameter $[\text{O III}]/([\text{O II}] + [\text{O III}])$, $\text{O3H}\beta = \log([\text{O III}]/\text{H}\beta)$ and $\text{S2O2} = \log([\text{S II}]/[\text{O II}])$.

$$\begin{aligned}
&\text{Use } (\text{O}/\text{H})_{\text{Cool}}^* \text{ if } \log(\text{N2H}\beta) \geq -0.1 \\
&\text{Use } (\text{O}/\text{H})_{\text{Warm}}^* \text{ if } \log(\text{N2H}\beta) < -0.1 \text{ and if } \log(\text{N2S2}) \geq -0.25 \\
&\text{Use } (\text{O}/\text{H})_{\text{Hot}}^* \text{ if } \log(\text{N2H}\beta) < -0.1 \text{ and if } \log(\text{N2S2}) < -0.25 \\
&\quad (\text{A9})
\end{aligned}$$

where $\text{N2H}\beta = \log([\text{N II}]/\text{H}\beta)$ and $\text{N2S2} = \log([\text{N II}]/[\text{S II}])$.

A7 Rcal and Scal (Pilyugin & Grebel 2016)

$$\begin{aligned}
(\text{O}/\text{H})_{\text{R.Lower}}^* &= 7.932 + 0.944 \times \log(\text{O32}) + 0.695 \times \log(\text{N2H}\beta) \\
&\quad + (0.970 - 0.291 \times \log(\text{O32}) - 0.019 \\
&\quad \times \log(\text{N2H}\beta)) \times (\text{O2H}\beta)
\end{aligned}$$

$$\begin{aligned}
(\text{O}/\text{H})_{\text{R.Upper}}^* &= 8.589 + 0.022 \times \log(\text{O32}) + 0.399 \times \log(\text{N2H}\beta) \\
&\quad + (-0.137 + 0.164 \times \log(\text{O32}) + 0.589 \\
&\quad \times \log(\text{N2H}\beta)) \times (\text{O2H}\beta) \quad (\text{A10})
\end{aligned}$$

$$\begin{aligned}
(\text{O}/\text{H})_{\text{S.Lower}}^* &= 8.072 + 0.789 \times \log(\text{O3S2}) + 0.726 \times \log(\text{N2H}\beta) \\
&\quad + (1.069 - 0.170 \times \log(\text{O3S2}) + 0.022 \\
&\quad \times \log(\text{N2H}\beta)) \times (\text{S2H}\beta) \\
(\text{O}/\text{H})_{\text{S.Upper}}^* &= 8.424 + 0.030 \times \log(\text{O3S2}) + 0.751 \times \log(\text{N2H}\beta) \\
&\quad + (-0.349 + 0.182 \times \log(\text{O3S2}) + 0.508 \\
&\quad \times \log(\text{N2H}\beta)) \times (\text{S2H}\beta) \quad (\text{A11})
\end{aligned}$$

where $\text{O2H}\beta = \log([\text{O II}]/\text{H}\beta)$, $\text{S2H}\beta = \log([\text{S II}]/\text{H}\beta)$ and $\text{O3S2} = \log([\text{O III}]/[\text{S II}])$.

$$\begin{aligned}
&\text{Use } (\text{O}/\text{H})_{\text{Lower}}^* \text{ if } \log(\text{N2H}\beta) < -0.6 \\
&\text{Use } (\text{O}/\text{H})_{\text{Upper}}^* \text{ if } \log(\text{N2H}\beta) \geq -0.6 \quad (\text{A12})
\end{aligned}$$

A8 C17R₂₃ (Curti et al. 2017)

$$\log R_{23} = 0.527 - 1.569y - 1.652y^2 - 0.421y^3 \quad (\text{A13})$$

where $12 + \log(\text{O}/\text{H}) = y + 8.69$

This paper has been typeset from a \LaTeX file prepared by the author.