Characterizing AGN Influence on the Calculated Metallicities of Adjacent Star-Forming Spaxels

Aidan Khelil

Oberlin College

Follow this and additional works at: https://digitalcommons.oberlin.edu/honors

Part of the Astrophysics and Astronomy Commons, and the Physics Commons

Repository Citation
https://digitalcommons.oberlin.edu/honors/856

This Thesis - Open Access is brought to you for free and open access by the Student Work at Digital Commons at Oberlin. It has been accepted for inclusion in Honors Papers by an authorized administrator of Digital Commons at Oberlin. For more information, please contact megan.mitchell@oberlin.edu.
Characterizing AGN Influence on the Calculated Metallicities of Adjacent Star-Forming Spaxels

Aidan Khelil
Advisor: Professor Jillian Scudder

Department of Physics and Astronomy
Oberlin College
Oberlin, OH
United States of America
31 May, 2022
Abstract

In this thesis, I introduce a method to identify and characterize the effects of active galactic nuclei (AGN) on the spectra of nearby star-forming regions. I analyze spatially-resolved areas of galaxies called "spaxels" within Data Release 15 of the Sloan Digital Sky Survey (SDSS) with the goal of locating those which are physically close to AGN. I find those spaxels with calculated metallicities which lie adjacent to AGN-flagged spaxels and characterize their metallicity values relative to the spaxels which are not adjacent to AGN-flagged spaxels, using a total of 11 separate metallicity calibrations. I find that the current methods to mask AGN-influenced regions for large-scale investigation are, in general, robust, as the largest median deviation between metallicities in border spaxels and those in non-border spaxels is 0.0467 dex. The largest mean difference in metallicity between border and non-border spaxels is 0.0522 dex with a standard deviation of 0.0590 dex. However, on a spaxel-by-spaxel basis, I find that the differences in metallicity between border spaxels and non-border spaxels can be as large as 0.9350 dex. These results are concerning for spaxel-by-spaxel analysis, and indicate the need for an improved masking process in the future.
Acknowledgments

Sitting here now, I am finding it exceptionally difficult to piece together the words necessary to describe the immense gratitude I feel towards my peers and mentors at Oberlin College. It is very easy to trivialize these kinds of statements, but in the hopes that someone may one day crack open this random undergraduate’s senior thesis and read this section, I think I’ll try to do my best to lay it out here. It is nothing more than what they deserve.

Firstly, I want to thank Professor Yumi Ijiri, in whose lab I gained my very first undergraduate research experience. It became clear to me right away that Yumi was not only an outstanding physicist but a better research advisor than one could reasonably ask for. In her lab, I got my first look into what it means to do physics research and the joys (and sorrows) that come with it. Yumi is one of the kindest and most understanding people I have ever met (even though she can sometimes be intimidatingly smart), and I consider myself fortunate beyond belief to have had her as my first ever research advisor.

I can’t write another word before expressing my thanks to yet another research advisor, Professor Jillian Scudder. I asked Professor Scudder to supervise my Honors research, and doing so was, without hyperbole, one of the best decisions of my life. My time working with Professor Scudder has not only been profoundly formative in my academic career, but it has informed many of the goals I have set for myself in the future. Professor Scudder is a rare kind of mentor that blends general academic brilliance with inspiring science communication skills. It was not until I met and worked with Professor Scudder than I truly knew what I wanted to do with my career and with my life.

I want to thank the entire Oberlin Physics & Astronomy Department for providing an incredibly welcoming and supportive environment for my entire undergraduate experience. I thank Professor Rob Owen and Professor John Scofield, with whom I have shared many laughs both inside and outside of the classroom. During my time at Oberlin, I have gone from studying under great teachers to collaborating with who I now consider my great friends.
Finally, I thank all of my peers, especially my fellow Honors students, Harry Mayrhofer, Dhruv Tandon, and James Sullivan. I would not have been able to accomplish anywhere near what I have done without their friendship and support. They are all exceptional physicists, and I truly cannot wait to see all the success they achieve in their future endeavors. I have no doubt that greatness is brewing for each and every one of you.

Thank you.
A  All Results (Figures)  

B  Python Code  

B.1  border_spaxels.py - KE08  

B.2  radius_compare.py - KE08
1 Galaxies, black holes, and SMBHs

Galaxies are some of the largest and most complex structures in the observable universe. Massive gravitationally bound systems of gas, dust, and stars, galaxies promise otherwise unattainable insights into the macroscopic mechanics of the cosmos. Galaxies can contain a number of stars between the order of $10^8$ and $10^{14}$. Filling the space between the stars in a galaxy is the vast interstellar medium, which cumulatively contains massive amounts of gas and dust of diverse composition. Much of the material found in the interstellar medium may contribute to the creation of future stars in areas known as “star-forming” regions (Stahler & Palla, 2004). The gas and dust in the interstellar medium may also give us insight into the history of the region in question (Kewley et al., 2019). Because stars fuse lighter elements into heavier isotopes through the process of nuclear fusion, we can reason that a large concentration of heavier elements implies an older cloud of gas which has been “enriched” by stars for a longer time.

Galaxies reside in enormous gravitational wells produced by the sheer mass of their constituent parts. One contributor to the gravitational well in galaxies is often a black hole, a region of space where an infinitely dense concentration of mass produces such a strong gravitational field that, at a certain distance, not even electromagnetic radiation can escape.

Supermassive black holes, or those with hundreds of millions of times the mass of the sun, are theorized to lie at the centers of almost all known galaxies (Kormendy & Richstone, 1995). These objects are capable of such energetic outbursts of electromagnetic radiation that they can outshine the rest of the galaxy, greatly interfering with our ability to obtain spectra from the fainter regions. For this reason, we must be sure that we account for oversaturation caused by a supermassive black hole to maintain the reliability of our data. Due to their potential to influence the accuracy and overall reliability of scientific measurements, it is crucial to identify and adjust for any regions which appear affected by the presence of a supermassive black hole (Baldwin et al., 1981).
2 Emission Spectra and Metallicities

The most common and useful galactic observations are those conducted on the emission spectra collected from galaxies. When the atoms in clouds of gas and dust inside a galaxy are excited by the light from nearby stars or from other energetic light sources, that radiation may be absorbed by the electrons in the atoms of the gas, promoting them to a higher energy level in their atomic structure. Electrons then spontaneously de-transition from a higher-energy state to a lower state to achieve the lowest possible energy state, and that excess energy is released as a photon with a wavelength proportional to the potential difference between the two levels. Because the potential differences between atomic energy levels are discrete and unique to a given element, this energy transition produces consistent patterns in the wavelengths of emitted light. The light we see in these patterns appear in the form of “spectral lines” at given wavelengths, which we combine to piece together the entire spectrum. Figure 1 displays an example of a galactic emission spectrum. The spectral lines of individual elements are clearly labeled, and a difference in relative intensity, or line strength, is observable. The elements present in a cloud of gas and dust can be precisely identified based on the measured wavelengths of light emitted during their energy level transitions. When observing nearby galaxies, it is oftentimes possible to obtain separate spectra from different spatially-resolved regions inside a single galaxy (e.g. Bundy et al., 2015). The acquisition of spatially-resolved spectral data allows us to characterize the internal behavior of a galaxy and improve our understanding of the mechanics of such large-scale systems.

Of particular interest in any spectroscopic study is the presence and concentration of more complex elements, or elements heavier than helium. In astronomy, any element heavier than helium is classified as a “metal”. The concentration of metals relative to “non-metals,” or hydrogen and helium, is called the “metallicity” of the region. From the wavelengths of light observed in these emission spectra, we can measure the intensity of spectral lines from metals and compare those to the intensity of hydrogen and helium lines. The resultant ratio
implies an elemental composition and allows us to obtain the metallicity for a particular region in a galaxy.

2.1 Spectral Lines of Interest

When looking to calculate metallicities for various spatially-resolved regions in a galaxy, we compare the ratios between the intensities of multiple specific emission lines. The most important of these lines are the $[O_{III}]\lambda 5007$, $[O_{II}]\lambda 3727$, $[N_{II}]\lambda 6584$, Hα (λ6563), Hβ (λ6564), and occasionally the $[S_{II}]\lambda 6717, 6731$ lines (Kewley & Ellison, 2008; Scudder et al., 2021). In this naming convention, the element or ion is listed first. For example, $O_{III}$ implies doubly-ionized oxygen is producing the spectral lines, while $O_{II}$ implies singly-ionized oxygen, $N_{II}$ indicates singly ionized nitrogen, and H represents atomic hydrogen. Next listed is the wavelength of the spectral line in question, measured in angstroms (Å), which are defined as 0.1 nanometers or $10^{-10}$ meters. Square brackets, such as those seen around $[O_{III}]$
and \([N_{II}]\), indicate a “forbidden line,” which are lines produced by low-density gas in outer space, and which are not easily reproduced in the lab on Earth. Spectral line ratios not only give us metallicity information but can also allude to factors that can negatively affect the reliability of our measurements. Our measurements concern spatially-resolved areas in a galaxy, therefore we are able to identify objects and phenomena taking place physically nearby to the regions under investigation (Bundy et al., 2015). As described later in this thesis, our methods for accounting for the impacts of nearby objects on emission spectra can be a detriment to the reliability of our data.

Analysis done on the galactic emission spectra in any spectroscopic study assume the excitation of the interstellar gas is done solely through photoionization processes (e.g. Baldwin et al., 1981), meaning all energy absorbed by the interstellar material came in the form of electromagnetic radiation from nearby O/B-type stars. Energy produced through other means may impact the emission line intensities observed and significantly change the apparent composition of the material by changing relative line ratios. Therefore, the characteristics of a particular spectrum could indicate the presence of a secondary energy source if they are not consistent with what we would expect from photoionization only. A simple way to determine the existence of an additional energy source is to compare the relative intensities of certain spectral lines (Baldwin et al., 1981; Kewley et al., 2001; Kauffmann et al., 2003; Stasinska et al., 2006).

### 2.2 Metallicities and Inference to Age

Metallicities provide a useful means of determining the age and star-forming capabilities of regions in galaxies. We can safely infer that because a star turns hydrogen and helium into heavier elements, any region in a galaxy with a high metallicity is most likely the result of many stellar lifecycles, with each star enriching the surrounding gas with heavier elements and contributing to the large metallicity value (Kewley et al., 2019). Likewise, we may infer that a region of space containing a large amount of gas and dust with a low metallicity is an
ideal candidate for new or ongoing star formation, as it contains the building blocks necessary for stars to form. Such regions will not have gone through as many stellar lifecycles and will therefore not contain many of the products of hydrogen fusion, such as carbon, nitrogen, or oxygen. The characteristic spectral lines of these elements are important in determining the metallicity of a region.

2.3 Metallicity Calibrations

There are 11 metallicity calibrations that I have considered for the purposes of this inquiry, grouped into 4 categories based on the significant spectral lines observed (Scudder et al., 2021). Metallicity calibrations are abbreviated according to the paper in which they were first introduced, with the author’s initials followed by the year of publication. The four calibrations that require \([O_{III}]\lambda 5007, [O_{II}]\lambda 3727, [N_{II}]\lambda 6584,\) and H\(\beta\) are called “\(R_{23}\)” calibrations (e.g. Zaritsky et al., 1994; McGaugh, 1991; Kobulnicky & Kewley, 2004; Kewley & Ellison, 2008). They are the only of our calibrations which consider singly-ionized oxygen, \(O_{II}\). Z94, M91, KK04, and KE08 are grouped under the \(R_{23}\) category.

The calculation for the \(R_{23}\) ratio goes as follows:

\[
R_{23} = \frac{[O_{II}]\lambda 3727 + [O_{III}]\lambda 4959, 5007}{H\beta} \tag{1}
\]

It is important to note that the \(R_{23}\) calibrations have the unique issue of double-valued solutions. When calculating an \(R_{23}\) metallicity of a spatially-resolved region, we get both a high-metallicity solution and a low-metallicity solution (Scudder et al., 2021). To break this degeneracy of solutions, we use the \([N_{II}]/[O_{II}]\) line ratio. Because the \([N_{II}]/[O_{II}]\) ratio is most sensitive to the metallicity of a region while avoiding sensitivity to the ionization level, it can be used to determine whether a region falls into the high-metallicity or low-metallicity solution “branch” (Kewley & Ellison, 2008). Once it is determined which bracket a region falls under, a metallicity value can be calculated.
The second category, called "O3N2," consists of three metallicity calibrations (e.g. Pettini & Pagel, 2004; Curti et al., 2016; Marino, R. et al., 2013). These calibrations require the \([O_{III}]\lambda5007\), \([N_{II}]\lambda6584\), H\(\alpha\), and H\(\beta\) lines. The three members of this category are M13-O3N2, PP04-O3N2, and C17-O3N2.

The O3N2 ratio is calculated as follows:

\[
O3N2 = \frac{[O_{III}]\lambda5007/H\beta}{[N_{II}]\lambda6584/H\alpha} \tag{2}
\]

The third group, the "N2" group, require only the \([N_{II}]\lambda6584\) and H\(\alpha\) lines (e.g. Pettini & Pagel, 2004; Curti et al., 2016; Marino, R. et al., 2013). The metallicity calibrations which fall under this category are M13-N2, PP04-N2, and C17-N2.

The N2 ratio is calculated as follows:

\[
N2 = \frac{[N_{II}]\lambda6584}{H\alpha} \tag{3}
\]

The final category of metallicity calibrations consists only of the D16 calibration (Dopita et al., 2016). We consider the \([N_{II}]\lambda6584\), H\(\alpha\), and \([S_{II}]\lambda\lambda6717,6731\) lines for this calibration. D16 is unique in considering the sulfur doublet \([S_{II}]\lambda\lambda6717,6731\).

To convert from line ratios into proper metallicities, we plug the line ratios calculated above into their respective polynomial calibrations. If we want to calculate a metallicity value for an \(R_{23}\) calibration such as M91, we must use the \([N_{II}]/[O_{II}]\) line ratio to determine whether we are looking for the high-metallicity or low-metallicity solution. If \(\log[N_{II}]/[O_{II}] \gtrsim -1.2\), the region is determined to lie in the high-metallicity solution range. In this case, the metallicity for the region is calculated as follows:

\[
Z = 12 - 4.944 - 0.767x + 0.602x^2 - y(0.29 - 0.332x - 0.331x^2) \tag{4}
\]

where \(Z\) is the metallicity, measured in dex, \(x = \log(R_{23})\) and \(y = \log([O_{III}]\lambda4959 +\)
\[ \frac{[O_{III}]\lambda 5007}{[O_{II}]\lambda 3727} \] (McGaugh, 1991). Similarly, if \( \log[N_{II}]/[O_{II}] \lesssim -1.2 \), we consider the low-metallicity solutions for this region. The value for a low-metallicity solution is calculated as follows:

\[
Z = 12 - 2.939 - 0.2x - 0.237x^2 - 0.305x^3 - 0.0283x^4 - y(0.0047 - 0.0221x - 0.102x^2 - 0.0817x^3 - 0.00717x^4)
\]

(5)

with \( Z, x, \) and \( y \) defined the same as above. The polynomials used to calculate M91, KE08, KK04, Z94, PP04_O3N2, and PP04_N2 metallicities are listed fully in Kewley & Ellison (2008), as well as Pettini & Pagel (2004), McGaugh (1991), Kobulnicky & Kewley (2004), and Zaritsky et al. (1994).
3 Active Galactic Nuclei

It is important to identify those supermassive black holes (SMBHs) with an exceptional effect on nearby regions in space and therefore will have the largest impact on our observations. While almost all galaxies contain SMBHs, not all SMBHs can be considered “active”. If the center of a galaxy proves far more luminous than would be ordinarily expected from exclusively starlight, it is classified as an active galactic nucleus (AGN). AGN are typically supermassive black holes closely orbited by highly-energized clouds of gas and dust called accretion disks.

AGN provide an obstacle to accurate metallicity measurements by skewing many of the observed values (Baldwin et al., 1981; Stasinska et al., 2006; Kewley et al., 2001; Kauffmann et al., 2003). As noted in Section 2.1, the spectra observed in a galaxy are assumed to be produced by gas undergoing only photoionization processes. However, the ultra-excited accretion disks around SMBHs emit a large number of high-energy particles and ionizing radiation. These particles can collide with nearby interstellar gas, heating and exciting it, which produces spectra which are inconsistent with our photoionization assumption. This process is one thing we wish to account for by identifying regions of a galaxy dominated by light from an AGN. Another AGN-related complication is that the high-energy ionizing radiation produces a large partially ionized regions, where strong \([N_{\text{II}}]\) lines are produced as well as increasing the overall ionization traced by \([O_{\text{III}}]\) (Kewley et al., 2019). Because AGN have such a large potential to damage the reliability of our metallicity calculations, regions containing AGN should be masked from the dataset before metallicity calculations are conducted.

3.1 The Mass-Metallicity Relationship

A general trend in galactic metallicities is described by the mass-metallicity, or M-Z relationship (Lequeux et al., 1979). As displayed in Figure 2, there is typically a positive
correlation between the measured mass of a galaxy and the observed metallicity of the gas inside the galaxy, as more massive galaxies display higher metallicities on average. The mass-metallicity relationship can be explained by a depletion model, where galactic winds accelerate metallic gas to velocities greater than the galactic escape velocity (Tremonti et al., 2004). In this model, galaxies with a greater mass produce a deeper potential well with a higher escape velocity, leading to the retention of a greater volume of gas.

Another important contributor to the observed M-Z relationship is the increased likelihood of higher-mass galaxies to contain an Active Galactic Nucleus (AGN) (Kauffmann et al., 2003). Galaxies which are host to an AGN are removed from M-Z plots, which skews the results in the high-mass section of the diagram.
3.2 Radial-Dependent Metallicity Values

An important observation of note is that regions of space closer to the galactic center typically have higher metallicity values than those at the outer edges of the galaxy (Rich et al., 2012). This is due to the inside-outward construction of galaxies, where new gas that falls into a galaxy’s gravitational well is more concentrated around the outer regions of the structure. This means that gas found nearer to the center has likely been a part of the star-forming material in the galaxy for longer and therefore has a higher metallicity.

Unusual behavior in expected metallicity trends could indicate the presence of an outer influence such as a galaxy merger event, where two galaxies collide and combine to form one aggregate system (e.g. Mihos & Hernquist, 1996; Rupke et al., 2010a,b; Torrey et al., 2012; Scudder et al., 2012; Ellison et al., 2008; Barrera-Ballesteros, J. K. et al., 2015; Sánchez, S. F. et al., 2015). For instance, a large amount of low-metallicity gas near the center of a galaxy would imply that such material was added to the galaxy recently or fell from outer regions inward. On the other hand, older, higher metallicity gas concentrated on the outside of a galaxy could imply that some major event pushed older gas outwards, or that older gas from another galaxy was captured in that atypical configuration.

3.3 Identifying AGN

One characteristic unique to photoionization processes, and therefore one we may use to identify AGN in a galaxy, is the “hardness” of the energy source (Baldwin et al., 1981). A harder spectrum is characterized by much higher energy light with the potential to ionize the atoms in nearby clouds of gas and dust. Areas affected by AGN have hard spectra and high-energy light that would be unexpected from a photoionized region. AGN-impacted regions are characterized by large amounts of partially ionized gas created by the hard emissions of an AGN. A strong $[N\text{II}]\lambda$6584 line would imply AGN activity, as $N\text{II}$, or singly-ionized nitrogen, is formed in partially ionized gas. Further, a large amount of doubly-ionized oxygen
would point to the fact that high-energy ionizing radiation from an AGN’s accretion disk is having an effect on nearby oxygen-rich regions of space and ionizing them to a greater degree than expected. We can take the ratio of the intensity of the \([O_{III}]\lambda5007\) line to that of the H\(\beta\) line (henceforth referred to as \([O_{III}]/H\beta\) based off their wavelengths) and compare it to the ratio of the \([N_{II}]\lambda6584\) line to the H\(\alpha\) line (henceforth referred to as \([N_{II}]/H\alpha\) (Kewley et al., 2001; Kauffmann et al., 2003; Stasinska et al., 2006). One convenient way to display these ratios is by using a so called “BPT Diagram,” which is a plot of the relative strengths of these lines (Baldwin et al., 1981).

Figure 3 shows an example of such a diagram, showing the \(O_{III}/H\beta\) ratio versus the \(N_{II}/H\alpha\) ratio of a typical galaxy, giving us a simple visualization of the relative line intensities. Using this comparison, we can not only determine the existence of AGN but also begin to characterize the AGN themselves. If we notice a very large intensity ratio for \(N_{II}/H\alpha\), we can infer that the presence of a powerful energy source is influencing the material in that region, and we can flag that region accordingly. Somewhat similarly, if we notice a high-intensity of the doubly-ionized oxygen line, \([O_{III}]\), when compared with the H\(\beta\) line, we can conclude that there is a large amount of overall ionization in that particular location. Regions of space which do not appear to have large \([O_{III}]\) and \([N_{II}]\) intensities are classified as “star-forming” regions. If a spatial region displays a strong \([O_{III}]\) line, it may still be classified as star-forming provided the \([N_{II}]\) line is weak enough. Star-forming regions form the dataset from which we can extract accurate metallicity values, due to their characteristiclly unaffected emission lines (Baldwin et al., 1981; Kewley et al., 2001; Kauffmann et al., 2003; Stasinska et al., 2006).
Figure 3: A density histogram of observed spectra from the SDSS DR15 in the form of a BPT diagram (Baldwin et al., 1981), with a y-axis displaying the $O_{III}\lambda5007/H\beta$ intensity ratio and the x-axis showing the $N_{II}\lambda6584/H\alpha$ line ratio, both on a log scale. Each purple dot represents a spectrum. We may infer that purple dots nearer to the top-right corner of the diagram are almost definitely host to AGN, as they have a large value for both the $O_{III}\lambda5007/H\beta$ line ratio and the $N_{II}/H\alpha$ line ratio.
4 MaNGA Survey

The source of our spectral data is the Mapping Nearby Galaxies at Apache Point Observatory (MaNGA) Survey (Bundy et al., 2015). The MaNGA Survey is one of the three main programs in the Sloan Digital Sky Survey’s fourth generation (SDSS-IV). Beginning in July, 2014 and concluding in December 2021, the MaNGA Survey sought to provide spectral “maps” of 10,000 nearby galaxies, where spatially-resolved regions in each galaxy are individually observed for their respective complete spectra in the range of 3600 Å to 10300 Å. MaNGA uses an integral field spectrograph consisting of 17 fiber-bundle integral field units. An integral field unit, or IFU, is a bundle of fiber-optic cables, with each cable tuned to capture the emission spectrum of one spatial region of a galaxy. The fiber bundles used in the survey consist of between 19 to 127 fibers arranged in a hexagonal configuration, which produces a hexagonal pattern observed in each galaxy map. The representations of spatially-resolved regions imaged by these fibers, each with their own respective spectra, are called “spaxels”. While each IFU bundle contains up to 127 fibers, each fiber’s orientation is slightly dithered to produce images of higher resolution than 127 spaxels. Each spaxel’s spatial location is recorded in the form of (X,Y) coordinates. In this thesis, I will examine the SDSS Data Release 15 (DR15), which was released in February 2019 and contains a total of 4,621 nearby galaxies (Aguado et al., 2019).

The MaNGA data is formatted as the measure of intensity of spectral lines in each observed spaxel. From this raw data, we must convert these spectral intensities into metallicities using the relationships described in Section 2.3. The galaxies in MaNGA are indexed based on their “Object ID,” which is an 18-digit identification number specific to each galaxy. Inside each object, the spaxels are indexed based on their “Spaxel ID,” which includes the Object ID and the X and Y coordinates for each spaxel observed in the galaxy. These coordinates are used to compare spaxels based on their physical location inside the object. Further, MaNGA includes each spaxel’s radial distance from the center of their galaxy, which allows
us to adjust for the expected radial dependence on calculated metallicities.

4.1 AGN Flags

MaNGA provides data on the emission lines from each spaxel in a galaxy, which allow us to calculate the intensity ratios of these lines. While subjectively comparing these two line ratios provides a means to broadly categorize regions as star-forming or AGN-influenced, there have been efforts to more concretely divide the two populations. Three major lines are added to the BPT diagram, the Kewley 2001 (abbreviated K01 in the future) line, the Kauffman 2003 (K03) line, and the Stasinska 2006 (S06) line. An identical diagram to Figure 3, but with the K01, K03, and S06 lines added, is displayed in Figure 4 (Kewley et al., 2001; Kauffmann et al., 2003; Stasinska et al., 2006).

These three lines serve as boundaries for definitively classifying spatially-resolved regions as star-forming or AGN-influenced. The S06 line serves as a conservative estimate of the number of regions that we may confidently consider “star-forming”. All points to the left of this line almost certainly pertain to star-forming regions. To the right of the S06 line are the K03 and K01 lines. The K01 line serves as the furthest point to which we may consider a region star-forming. Points to the right of the K01 line are considered almost definitely impacted by AGN activity.

A set of AGN “flags” were constructed for quality assurance for Scudder et al. (2021) to delineate all observed spectra regions. Those found to the right of the K01 line will be flagged “AGN,” as they are highly unlikely to contain unaffected spectral line ratios. Points found between the K01 and K03 line will be flagged “K01_SF.” As the K01 line serves as a truly optimistic upper-bound on star-forming regions, material in this region on the diagram is also quite likely to feel the effects of AGN activity. Therefore, objects in this region probably would not produce accurate metallicity values. Points which lie between the K03 and S06 line are flagged “K03_SF.” These are more likely star-forming regions than not, so metallicity data from these areas can be trusted. We flag regions to the left of the S06
Figure 4: BPT diagram with included K01, K03, and S06 lines. Created for quality assurance for Scudder et al. (2021), points to the left of the S06 line are considered definitely star-forming, while those to the right of the K01 line are considered definitely AGN-affected. Points in region i, ii, iii, and iv are given the flags S06_SF, K03_SF, K01_SF, and AGN respectively. These flags reflect their likelihood to be influenced by a nearby AGN.

4.2 Galaxy Maps

Using each spaxel’s spatial location as well as their calculated metallicities and AGN flags, we can construct our own maps of each galaxy, with which we can visualize both its metal-
Figure 5: Each spatially resolved region, spaxel, in the galaxy identified by Object ID “587724197746311240”. X and Y spatial coordinates are labeled on their respective axes. Spaxels are color-coded based on their associated AGN flags, with the key shown on the right. A “Low S/N” marked spaxel is one in which a low signal-to-noise ratio prevented the acquisition of accurate emission line strengths.

ity distribution and possible AGN-affected regions. Having attached AGN flags onto each spatially-resolved region of a galaxy, we construct a “footprint” like the one displayed in Figure 5. This is an image of the spaxels of the galaxy identified by Object ID “587724197746311240”. The spaxels’ X and Y coordinates are labeled on the X and Y axes, respectively. Each spaxel is given a color-coded AGN flag according to the key on the right. The purple regions indicate spaxels where a low signal-to-noise ratio prevented the acquisition of trustworthy data. For a spaxel to be considered for an AGN flag, the signal-to-noise (S/N) ratio, or the ratio of emission line intensity to background, must be greater than 1. A diagram like this provides a convenient visualization of the spaxels’ AGN flag distribution in any observed galaxy. The AGN footprint shown in Figure 6a is quite typical and expected for those galaxies containing active galactic nuclei. A large AGN-flagged region is clear to see in the center of the galaxy, with “arms” of AGN-impacted material moving out towards the outside of the structure.

Just as each spaxel has an associated AGN flag, those flagged as more-likely star-
forming, those with either a S06_SF flag or a K03_SF flag, have an associated metallicity. Further, just as we can construct a “footprint” of AGN flags for a particular galaxy, we can also construct a footprint of metallicity distributions in that galaxy. Figure 6 is an side-by-side comparison of both types of footprints. The metallicities displayed in Figure 6b are calculated by Scudder et al. (2021) with the KE08 metallicity calibration and polynomial fit. As with the AGN footprint, the metallicity footprint for this galaxy is typical of a galaxy with an AGN at the center. Spaxels in a large area around the middle of the galaxy are masked, as are those in the vicinity of the AGN-impacted “arms” of the object. These spaxels are, as expected, flagged K01_SF or AGN. Metallicity values are only attached to the spaxels in the S06_SF and K03_SF regions near the top and bottom of the galaxy. These spaxels, and particularly those which border an AGN-flagged or K01_SF-flagged spaxel, are the subject of my investigation regarding the effects of AGN on their displayed metallicities. Another expected behavior on display in this figure is the radial dependence of metallicity, which is easy to notice in the image on the right. Spaxels near the center of the galaxy clearly display larger metallicity values on average than those at the outskirts. However, there remain spaxels in the outer regions of the galaxy which border AGN-affected spaxels. The radial dependence on calculated metallicities must be taken into account when attempting to characterize the severity of AGN impact on these measurements.

Galaxy maps are diverse in shape and structure, and many display properties which are unexpected, unusual, or which could hint at other processes in play. Figure 7 displays two such galaxies with very unusual AGN footprints. Both of these galaxies display a footprint almost somewhat antithetical to that which we would expect from and ordinary AGN-occupied galaxy. Both have regions flagged as AGN or K01_SF, but neither show any of these AGN-flagged spaxels at the galactic center. Both Figures 6a and 6b exhibit very little AGN behavior, as there are only a dozen or so spaxels flagged AGN. However, there are large regions flagged K01_SF, which flank the north and south of the galactic center. As described above, spaxels flagged K01_SF are thought to be likely AGN-impacted.
Figure 6: AGN footprint (a, left) and Metallicity footprint (b, right) of the galaxy denoted by Object ID “58772550658322550”. This galaxy displays a general shape and metallicity distribution that is highly typical of galaxies in the MaNGA survey. This object has a clearly marked AGN in the center and two AGN-impacted “arms” on the left and right. All regions flagged as K01 SF or AGN have their metallicities masked, and metallicities are only available for those spaxels flagged S06 SF or K03 SF. Further, the expected radial dependence on observed metallicity is clear to see in (b), as the associated metallicity values of spaxels at the outskirts are on average lower than those nearer to the center. The metallicities in (b) were calculated for Scudder et al. (2021).
Figure 7: Two atypical AGN footprints. If there is an AGN present in a galaxy, we would expect it to be located near the center, such as in Figure 6a, with AGN-impacted regions close-by. However, in both (a) and (b), there are star-forming spaxels near the center, with AGN-flagged spaxels surrounding the top and bottom. These figures suggest that some non-AGN process is producing a hard ionization field in specific regions of this galaxy.

While Figure 7 offers behaviors slightly different from those that we would expect, the two galaxies pictured are far from the most unusual objects in the dataset. Figure 8a displays an exceptionally strange AGN footprint. The center of the galaxy is characterized by a large expanse of star-forming spaxels entirely encased in a shell of AGN-affected spaxels. This is close to the opposite distribution that we would expect, as there is no AGN at the center of this galaxy but rather a “ring” of AGN-flagged areas around it. A footprint such as those in Figure 8 could indicate that a galaxy merger had taken place in this galaxy’s past. It is possible that the gravitational influence between two galaxies, at least one of which contained an AGN, caused much of the AGN-affected, ionized material to disperse and distribute along the outside of the galaxy. At the same time, it is possible that the unaffected star-forming material coalesced nearer to the center during the merger. Another indicator of an extragalactic influence in the past is the unusual metallicity distribution shown in Figure 8b. The expected radial dependence is not observed, with higher-metallicity gas frequently most
Figure 8: Highly unusual AGN (a) and metallicity (b) footprints. A vast expanse of star-forming space comprises the center of this galaxy, while there is a large number of AGN-impacted spaxels forming a “ring” around the outskirts, as shown in (a). A footprint like this could hint at the possibility that a galactic merger event threw AGN-affected gas towards the outside of the galaxy while allowing unaffected gas to coalesce in the center. Another hint that such an event once took place is the unusual metallicity distribution in (b), where the usual radial dependence breaks down. Here, the higher-metallicity spaxels are in part concentrated along the outer edge of the star-forming region. The metallicities in (b) were calculated for Scudder et al. (2021).

4.3 AGN “Bleed-Through”

Masking out data in any dataset runs the risk of removing too much or too little and impacting your eventual conclusion. This can be especially problematic when considering AGN. It is unclear how effective our AGN selection criteria (described in Section 3.3) are in omitting all impacted emission spectra for spatially-resolved data. An unidentified discrepancy in a spectrum will naturally imply a difference in composition of the nearby material and could therefore mislead as to the history and characteristics of the region in question. This is an
example of an incorrect assumption, such as the idea that photoionization is the only source of excitation in a gas, prompting an incorrect conclusion. To prevent these kinds of faulty interpretations of the data, regions suspected of containing an active galactic nucleus must be masked out of consideration so that their spectra do not influence the spectra of other regions in the galaxy. Data from regions flagged as K01\_SF or AGN do not have their metallicities calculated in order to minimize the negative distortion effects of AGN on aggregate galactic spectra. These are the masks I am investigating. This thesis seeks to identify how effectively our masking procedure minimizes AGN influence, as well as pinpoint and describe any metallicity calibration-specific issues that arise from it.

4.4 The Database

From these MaNGA data, we have constructed a SQL database which includes the Object IDs, Spaxel IDs, AGN flags, metallicities, and radial distance from the galactic center. All investigation in this thesis will be structured around SQL database queries to extract relevant information. For the purposes of this investigation, the use of a SQL database is preferred due to the large volume of data in our set. Because the quantity of spaxels is on the order of $10^7$, the construction of our database is necessary to facilitate quicker indexing and selection of the relevant data.

Querying a SQL database is done by providing a list of properties for the database to consider. For example, we could look to select only those spaxels which have a non-null metallicity value and also have an associated AGN flag of K03\_SF. The database then collects all spaxels which match all desired properties and sequences them into an array. The Spaxel ID of a given spatially-resolved region is unique to that region so, it is helpful to sequence each data point by its Spaxel ID. When extracting data from multiple columns of the database, we must match all columns by Spaxel ID to ensures that only those data with matching associated Spaxel IDs are considered together. Otherwise, it is possible to apply a metallicity calculation to an incorrect spaxel.
5 Method

5.1 Extracting from the Database

The SQL database was queried to extract the Spaxel ID, Object ID, and metallicity values of spaxels which had associated, non-null metallicity values. Only those spaxels flagged “S06_SF” or “K03_SF” have associated, non-null metallicity values because only these points have not had their metallicities masked by the process described in Section 4.1. These spaxels will hereon be called “metallicity-marked” spaxels, of which there are about 1.1 million, with a small quantity dependence on the metallicity calibration that is under inquiry. Another query extracted the Spaxel ID, Object ID, and AGN flag for all spaxels which were flagged as “AGN” or “K01_SF.” These AGN flags had no associated metallicity values and were used as AGN spaxels. They will hereon be called “AGN-flagged” spaxels. There are 730,317 AGN-flagged spaxels. Because the presence of an AGN flag is not dependent on the metallicity calibration being investigated, this number remains constant for all calibrations.

The Spaxel ID of any given spaxel is formatted as a string of the following structure: “[Object ID]_[Spaxel X-Coordinate]_[Spaxel Y-Coordinate]”. To extract the X and Y coordinates from this, the string was chopped according to the delimiter “_” and broken into columns of Object ID, X-Coordinate, and Y-Coordinate. This was done to both the metallicity-marked spaxels and AGN-flagged spaxels.

5.2 Finding “Border” Spaxels

To begin an iteration through all relevant galaxies to extract border spaxels, I select only those galaxies with at least one AGN-flagged spaxel and at least one metallicity-flagged spaxels. A new array is created consisting of all unique Object IDs that are found with both AGN and metallicity-marked spaxels. A simple mask allows us to locate those object IDs that appear in both lists and therefore contain at least one of each type of spaxel. This
array consists of an average of 2602 galaxies, with different calibrations having as few as 2597 galaxies or as many as 2608, and it will serve as the list of galaxies through which I iterate.

For each galaxy, I create a mask to select all spaxels which reside inside that galaxy. I then select all AGN-flagged spaxels which lie close to a metallicity-marked spaxel. This is done by removing all AGN-flagged spaxels that do not have a metallicity-marked spaxel within one pixel separation in either the X or Y directions. I next iterate through all these “nearby” AGN-flagged spaxels, individually testing each one for metallicity-marked spaxels within one pixel of separation in both the X and Y directions. If an AGN-flagged spaxel is adjacent to a metallicity-marked spaxel in both directions, that metallicity-marked spaxel is added to a dynamic array, along with its coordinates, metallicity value, Spaxel ID, and Object ID. Following the iteration through all galaxies, the result is a 2-dimensional array containing all metallicity-marked spaxels with an adjacent AGN-flagged spaxel.

From this array of, on average, 57,187 border spaxels, I can extract the individual metallicity values of these spaxels for comparison to those metallicity-marked spaxels which do not lie on a border with an AGN, of which there are about $1.03 \times 10^7$ on average. I can also determine the proportion of metallicity spaxels which border at least one AGN-flagged spaxel when compared to the total number of metallicity-marked spaxels which are in the dataset. Each of these comparisons can be done on a per-galaxy basis or as a general comparison regardless of galaxy.

5.3 Adjusting for the Radial Dependence of Metallicities

To compare metallicity-marked spaxels on an AGN border to those at similar radii but not on a border, I query the database to extract the object ID, Spaxel ID, metallicity, and radius from the galactic center of all metallicity-marked spaxels. All relevant object IDs were identified while setting up the iterative loop described above, and those are loaded in from a text file to maintain the same set of IDs.
Iterating through all relevant galaxies again, all spaxels are placed into bins by radius. Each bin corresponds to a 0.5 kpc range from 0 kpc up to the maximum detected radius of spaxels in that galaxy. This can be anywhere up to 29.5 kpc, but is typically between 5 to 10 kpc. I may now check which metallicity spaxels in each bin are found in our list of “border” spaxels (those which lie next to an AGN-flagged spaxel). Separating these from our full radial bins, we end up with two sets of data per radial bin: one with all border spaxels and one with “non-border” spaxels.

Within these bins, I compare the metallicities of border spaxels versus non-bordering spaxels regardless of the effects that distance from the galactic center may have on either. I iterate through all bins and find the average metallicities calculated for border spaxels and the average metallicities calculated for non-border spaxels in each bin. I then subtract the non-border average from the border average to get the difference between the two for each bin. The result is an array of about 12,300 metallicity differences, with which I calculate the median, mean, and standard deviation with respect to each metallicity calibration. I also find the metallicity difference with the largest magnitude to check if any highly effected spaxels managed to slip through our masking process undiscovered.
6 Results/Discussion

6.1 Prevalence of Border Spaxels

The first major point to make the frequency at which border spaxels appear. Of the roughly 2600 galaxies surveyed, an average of 2124 of them, with slight variation depending on metallicity calibration, contained at least one spaxel bordering an AGN-flagged region, representing a rate greater than 80%. Figure 9 displays four histograms describing the number of galaxies that contain different numbers of border spaxels. In general, the number of galaxies decreases exponentially with the number of border spaxels per galaxy. There is a dearth of galaxies hosting between 320 and 380 border spaxels, with a few containing between 380 and 420.

One interesting result to note is the relative lack of border spaxels present when considering the C17_N2 calibration. The number of border spaxels across all galaxies when considering the other 10 calibrations is about 57,800 on average. However, when considering the C17_N2 calibration, I find only 51,079 metallicity-marked border spaxels. The calibration with the next lowest count is KK04 with 56,899 border spaxels, representing greater than an 11% increase over the number found with C17_N2 metallicities. The total number of C17_N2-marked spaxels is far from the lowest among all calibrations, and the number of galaxies with at least one C17_N2-flagged border spaxel is 2124; which is exactly the average across all calibrations. This discrepancy in spaxel count could be due to a difference in “range of validity” between the C17_N2 calibration and the other calibrations listed (Scudder et al., 2021). If the calculated metallicity for a spaxel lies outside the range of validity for a calibration, then that spaxel’s metallicity is masked. This could be the cause of the relatively lower spaxel count in the C17_N2 calibration’s graph.

Another potentially interesting correlation to consider is the proportion of border spaxels to the total number of metallicity-marked spaxels in a particular galaxy. A large propor-
Figure 9: Four histograms describing the count of galaxies versus the number of border spaxels found per galaxy, with the plots indexed based on the metallicity calibration considered. The number of galaxies decreases logarithmically as the number of border spaxels per galaxy increases. There are no galaxies with anywhere between 320 and 380 border spaxels, and only a few galaxies contain between 380 and 420. (d) displays the data for the C17_N2 calibration, which has a comparatively small number of border spaxels per galaxy. This calibration is unique in this regard, and the cause of its deviation from the average likely due to a difference in range of validity when compared with the other calibrations (Scudder et al., 2021).
tion would imply that metallicities are frequently being calculated for star-forming regions very close to AGN-flagged spaxels in that galaxy. Figure 10 displays a histogram relating the number of galaxies to the percentage proportion of border spaxels to total metallicity-marked spaxels per galaxy, with respect to metallicity calibration. There is no clear trend immediately visible from these plots. Figures 9a, c, and d show remarkably similar shapes despite being calculated with different emission line ratios. However, Figure 10b shows a slightly different behavior in the region corresponding to a 50-100% proportion of border spaxel count to total metallicity spaxel count. It appears there are significantly fewer galaxies where 95-100% of the D16 spaxels border an AGN. This could also be due to a difference in range of validity between the D16 calibration and the others pictured. D16 also has a very large scatter when converted from the other calibrations (Scudder et al., 2021), which could couple with the range of validity issue and change the shape of the graph.

6.2 AGN Effects on Border Spaxels

The significance of AGN influence on the metallicities of star-forming regions is strongly dependent upon the scale of the investigation. Large-scale analyses which look at a considerable number of galaxies will find our current AGN-masking methods sufficient, as the largest median deviation between the metallicities of border spaxels and the metallicities of non-border spaxels is small; 0.0467 dex when considering the PP04_N2 metallicity calibration. Figure 11 displays six histograms describing the magnitude and direction of deviation between border spaxel metallicities and non-border spaxel metallicities. The x-axis shows the difference described by “border metallicities-non-border metallicities,” while the y-axis counts the number of radius bins that fall into each difference bin.

The legends in each image list the median, mean, and standard deviation of the metallicity difference rounded to the fourth decimal. The median remains very close to zero in all of these plots, implying that any investigation using large amounts of data should find the current masking process to be non-problematic. However, the maximum magnitude of
Figure 10: Four histograms showing the count of galaxies versus the proportion of border spaxels to all metallicity-marked spaxels per galaxy, with the plots indexed based on the metallicity calibration considered. There is a slight difference in shape at high ratios in the D16 calibration (b). This is likely due to a relatively large scatter when converting between the D16 calibration and another (Scudder et al., 2021). Because the D16 calibration shares only one line with the rest, it is possible that the calibrations do not match particularly well and cause the observed differences in histogram shape.
Figure 11: Six histograms describing the count of radius bins versus the metallicity difference between border spaxels and non-border spaxels at that radius, depending on metallicity calibration. The legend in the top right of each image lists the median, mean, and standard deviation of the metallicity difference in dex, rounded to the fourth decimal place. The largest spread is observed in the Z94 metallicity calibration (c), which displays a standard deviation of about 0.1628 dex. The M13_O3N2 (e) and M13_N2 (f) calibrations display the lowest spread out of the six figures, with standard deviations of 0.0342 and 0.0335 dex respectively. Full sets of plots are available in Appendix A.
difference between spaxels in a radius bin is exceptionally large, at 0.9350 dex in the D16 calibration. This is a cause for concern for any investigation looking to consider spaxels on an individual level, and could necessitate improvements to our AGN-masking procedure in the future. Table 1 displays the maximum metallicity difference between any two spaxels in both the positive and negative direction, as well as the median, mean, and standard deviation from the mean difference.

<table>
<thead>
<tr>
<th>Calibration</th>
<th>Max (dex)</th>
<th>Min (dex)</th>
<th>Range (dex)</th>
<th>Mean (dex)</th>
<th>Median (dex)</th>
<th>StDev (dex)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z94</td>
<td>0.5597</td>
<td>-0.6518</td>
<td>1.2115</td>
<td>-0.0471</td>
<td>-0.0415</td>
<td>0.1145</td>
</tr>
<tr>
<td>M91</td>
<td>0.6069</td>
<td>-0.6598</td>
<td>1.2667</td>
<td>-0.0411</td>
<td>0.0358</td>
<td>0.1082</td>
</tr>
<tr>
<td>KK04</td>
<td>0.6073</td>
<td>-0.5912</td>
<td>1.1985</td>
<td>-0.0374</td>
<td>-0.0296</td>
<td>0.0991</td>
</tr>
<tr>
<td>KE08</td>
<td>0.6856</td>
<td>-0.4039</td>
<td>1.0895</td>
<td>0.0093</td>
<td>0.0027</td>
<td>0.0852</td>
</tr>
<tr>
<td>D16</td>
<td>0.9350</td>
<td>-0.5011</td>
<td>1.4362</td>
<td>0.0205</td>
<td>0.0115</td>
<td>0.0939</td>
</tr>
<tr>
<td>PP04_O3N2</td>
<td>0.4602</td>
<td>-0.2924</td>
<td>0.7527</td>
<td>-0.0110</td>
<td>-0.0123</td>
<td>0.0522</td>
</tr>
<tr>
<td>PP04_N2</td>
<td>0.6293</td>
<td>-0.2588</td>
<td>0.8881</td>
<td>0.0522</td>
<td>0.0467</td>
<td>0.0590</td>
</tr>
<tr>
<td>M13_O3N2</td>
<td>0.2461</td>
<td>-0.1946</td>
<td>0.4407</td>
<td>-0.0076</td>
<td>-0.0083</td>
<td>0.0342</td>
</tr>
<tr>
<td>M13_N2</td>
<td>0.4795</td>
<td>-0.2084</td>
<td>0.6879</td>
<td>0.0252</td>
<td>0.0199</td>
<td>0.0335</td>
</tr>
<tr>
<td>C17_O3N2</td>
<td>0.4921</td>
<td>-0.3747</td>
<td>0.8668</td>
<td>-0.0063</td>
<td>-0.0081</td>
<td>0.0383</td>
</tr>
<tr>
<td>C17_N2</td>
<td>0.5910</td>
<td>-0.2383</td>
<td>0.8293</td>
<td>0.0372</td>
<td>0.0336</td>
<td>0.0437</td>
</tr>
<tr>
<td>Average</td>
<td>0.5721</td>
<td>-0.3977</td>
<td>0.9698</td>
<td>-0.0006</td>
<td>0.0019</td>
<td>0.0693</td>
</tr>
</tbody>
</table>

Table 1: All metallicity calibrations with their associated discrepancies. The average of the non-border metallicities is subtracted from the average of the border metallicities from each radial bin to produce a large set of differences by radial bin. The maximum and minimum results of this subtraction are displayed alongside the overall ranges of values, as well as median, mean, and standard deviation in the metallicity difference array.
7 Conclusions & Future Work

I have introduced and described my process for identifying and characterizing the effects of AGN on the spectra of nearby star-forming regions. I located metallicity-marked spaxels which lie physically close to AGN-flagged regions in the galaxies of the SDSS Data Release 15. I then compared the metallicities calculated for the star-forming spaxels which are on the border of an AGN-flagged region to the metallicities of spaxels which do not lie adjacent to AGN-flagged spaxels. I found that the current method for masking AGN-impacted regions is sufficient for large-scale data analysis, but single-spaxel investigations would benefit from a more robust masking procedure at that small scale. The largest median metallicity deviation between border spaxels and non-border spaxels is 0.0467 dex in the PP04_N2 calibration, which implies that, in a large dataset, not much deviation is observed. However, the presence of metallicity differences as large as the calculated 0.9350 dex in the D16 calibration is concerning for spaxel-by-spaxel analyses. The largest mean offset also occurs in the PP04_N2 calibration, which displays a mean difference of 0.0522 dex with a standard deviation from the mean of 0.0590 dex.

To begin to identify potential improvements to our current masking process, we must conduct further investigation into the characteristics of the spaxels that provided the largest metallicity differences. Any commonalities between the spaxels with a large metallicity difference could more directly point to the flaws in our masking approach thus far. In particular, the local behavior of their host galaxies could prove to be a significant clue in identifying areas for improvement. It is likely that the heavily-affected border spaxels should have been included in an AGN mask, but ended up being skipped by the algorithm. In these cases, having a larger masked area could benefit future spatially-resolved spectroscopic surveys.

On the other hand, the very low median of the calculated metallicity difference implies that our current masking procedure works very well for large sample sizes of spaxels. One
potential avenue for future investigation is the maximization of that sample size. If the 
median metallicity difference as calculated above can be characterized as a function of the 
“strictness” of our masking procedure, we could create a mask which maximizes the available 
spectral data with minimal AGN interference, effectively shrinking our mask to its minimum 
size wherever possible while preserving data integrity. This kind of investigation could result 
in thousands of additional spaxels’ worth of spectroscopic data in the future.

Further still, we have seen the publication of Data Release 17 (DR17) of SDSS’ fourth 
generation in December 2021, officially concluding the MaNGA Survey and doubling the 
overall galactic sample size from 4,621 to 9,269 galaxies (Abdurro’uf et al., 2022; Law et al., 
2021). This large sample size only necessitates further inquiry into the reliability of our 
masking and analysis techniques. In the near future, an identical investigation to that which 
I have outlined in this thesis could be simply applied to the calculated metallicities for. 
DR17’s much larger set of spatially-resolved star-forming regions.
Appendices

A All Results (Figures)
Figure 13: All 11 histograms describing the count of radius bins versus the metallicity difference between border spaxels and non-border spaxels at that radius, depending on metallicity calibration. The legend in the top right of each image lists the median, mean, and standard deviation of the metallicity difference, rounded to the fourth decimal place.
Figure 15: All 11 histograms describing the count of galaxies versus the number of border spaxels found per galaxy. There are no galaxies with anywhere between 320 and 380 border spaxels, and only a few galaxies contain between 380 and 420. The C17_N2 calibration has a comparatively small number of border spaxels per galaxy. This calibration is unique in this regard, and the cause of its deviation from the average likely due to a difference in range of validity when compared with the other calibrations (Scudder et al., 2021).
Figure 17: All 11 histograms describing the count of galaxies versus the proportion of border spaxels to all metallicity-marked spaxels per galaxy, with the plots indexed based on the metallicity calibration considered. The D16 calibration has fewer galaxies with a high proportion of border spaxels. This could be caused by the large scatter when converting between the D16 calibration and others (Scudder et al., 2021).
B Python Code

B.1 border_spaxels.py - KE08

The code to find and compile all border spaxels for the KE08 metallicity calibration. This code can be easily modified to consider each of the other 10 metallicity calibrations by replacing all instances of “KE08” in each SQL query to instances of the desired metallicity calibration.

```python
import numpy as np
import array
import math
import scipy as sci
import pymysql as sql
import datetime
import matplotlib.pyplot as plt

# Query the database to extract all spaxels with associated non-null metallicities (will not include K01_SF or AGN), along with the attached Object ID and metallicity value (KE08 calibration)

print('MetalQueryStart',datetime.datetime.now().time())
print("***** LOADING METALLICITY SPAXELS *****")

db = sql.connect(host="xxxxxxxxxx",user="xxxxxxxxxx", password="xxxxxxxxxx", database="xxxxxxxxxx")
c = db.cursor()
c.execute(''ATEST
spaxels.objid, metal.spaxID, metal.KE08_smc from
    dr15_metallicities metal,
    dr15_spaxels_uber spaxels where
    metal.spaxID=spaxels.spaxID and metal.ke08_smc is not null and spaxels.agn_flag_smc is not null;''
)
print('MetalQueryDoneDateTime:',datetime.datetime.now().time())
rowsMetal = np.asarray(c.fetchall())
db.close()

# The result is a 2-dimensional array containing each desired column (SpaxID, ObjID, KE08)

column_names = [description[0] for description in c.description]
```

# Breaking the SQL query array into its component columns
MetalObjID = rowsMetal[:,0]
MetalSpaxid = rowsMetal[:,1]
ke08 = rowsMetal[:,2].astype(float)

# Split the Spaxel ID array of strings into three arrays of ObjID, X-Coordinate, and Y-Coordinate
MetalSplitArray = np.asarray(np.char.split(MetalSpaxid, sep='_' ))
MetalSplitArray = np.stack(MetalSplitArray.ravel())

print(np.shape(MetalSplitArray))
ke08_ID = MetalSplitArray[:,0]
ke08_X = MetalSplitArray[:,1].astype(int)
ke08_Y = MetalSplitArray[:,2].astype(int)

# Query the Database to extract the Object ID, Spax ID, and AGN flag of all spaxels flagged as AGN or K01_SF. These will be used to check for bordering metallicities

print('AGNQueryStartDateTime:',datetime.datetime.now().time())
print('******** LOADING AGN FLAGGED SPAXELS *******')
db = sql.connect(host="xxxxxxxxxxx",user="xxxxxxxxxxx", password="xxxxxxxxxxx", database="xxxxxxxxxxx")
c = db.cursor()
c.execute("'select objid, spaxID, agn_flag_smc from dr15_spaxels_uber where agn_flag_smc is not null and agn_flag_smc != "S06_SF" and agn_flag_smc != "K03_SF";"')
rowsAGN = np.asarray(c.fetchall())
db.close()

# Similarly, the output is a 2-dimensional array containing columns for each desired property
print('AGNQueryDoneDateTime:',datetime.datetime.now().time())
column_names = [description[0] for description in c.description]
# print('Column Names:',column_names)

# Splitting up the 2-D AGN array
AGNObjID = rowsAGN[:,0]
AGNSpaxid = rowsAGN[:,1]
AGNFlag = rowsAGN[:,2]

# Splitting up the spaxID string into objid, x coordinate, and y coordinate
AGNSplitArray = np.asarray(np.char.split(AGNSpaxid, sep='_' ))
AGNSplitArray = np.stack(AGNSplitArray.ravel())

print(np.shape(AGNSplitArray))
AGN_ID = AGNSplitArray[:,0]
AGN_X = AGNSplitArray[:,1].astype(int)
AGN_Y = AGNSplitArray[:,2].astype(int)

# Finding all unique Object IDs which have at least one metallicity-marked spaxel and one AGN-flagged spaxel. This will serve as the array of strings through which I will iterate to check each galaxy for border spaxels

print("******* GATHERING GALAXY IDs *******")
uniqueAGN_ID = np.unique(AGNObjID)
print('uniqudAGNshape = ',np.shape(uniqueAGN_ID))
uniqueke08_ID = np.unique(MetalObjID)
print('uniqudke08shape = ',np.shape(uniqueke08_ID))

common galaxy = np.where(np.isin(uniqueAGN_ID, uniqueke08_ID))
GalaxyIDs = uniqueAGN_ID[common galaxy]
print('galaxyID shape = ', np.shape(GalaxyIDs))

# Defining all of the arrays for use in the below loop
RatioArray = np.asarray([]) # Array to record the ratio of border spaxels to total number of metallicity-marked spaxels
RatioBorderCount = np.asarray([]) # Array to record the total number of bordering spaxels in each galaxy
RatioID = np.asarray([]) # Recording the ID of the galaxy

print('LOOP START')
print('DateTime: ',datetime.datetime.now().time())
print()
print('Finding border spaxels for {num} galaxies.'.format(num = len(GalaxyIDs)))
print()
print('Spaxel count: ', len(ke08)+len(AGNFlag))
print()

BIGBorderX = np.asarray([]) # Arrays with the word BIG in front are not delineated by galaxy but rather contain all border spaxels across all galaxies
BIGBorderY = np.asarray([])
BIGObjID = np.asarray([])
BIGMetallicities = np.asarray([])
BIGSpaxID = np.asarray([])

# Begin the loop through each galaxy to check for border spaxels
for ID in GalaxyIDs:
BorderSpaxels_X = np.asarray([])  # For recording the spatial coordinates of the bordering spaxels
BorderSpaxels_Y = np.asarray([])
BorderObjID = np.asarray([])
ke08NearValue = np.asarray([])  # For recording the metallicities of the bordering spaxels
BorderX = np.asarray([])
BorderY = np.asarray([])
BorderID = np.asarray([])
Metallicities = np.asarray([])
BorderSpaxID = np.asarray([])

# Creating and applying masks to only address the current galaxy in the loop

AGNGalaxy = np.where(ID == AGN_ID)
ke08Galaxy = np.where(ID == ke08_ID)

AGNX = AGN_X[AGNGalaxy]
AGNY = AGN_Y[AGNGalaxy]

ke08X = ke08_X[ke08Galaxy]
RatioX = ke08X
ke08Y = ke08_Y[ke08Galaxy]
ke08ID = ke08_ID[ke08Galaxy]
SpaxID = MetalSpaxid[ke08Galaxy]
ke08Value = ke08[ke08Galaxy]

# Creating a mask to find only those AGN-flagged spaxels which are relatively near metallicity-marked spaxels. This is only finding those which are within 1 spaxel in EITHER the X or Y direction.

NearX1 = np.where(((np.isin(AGNX-1, ke08X)) | (np.isin(AGNX+1, ke08X)) | (np.isin(AGNX, ke08X))) & ((np.isin(AGNY-1, ke08Y)) | (np.isin(AGNY+1, ke08Y)) | (np.isin(AGNY, ke08Y))))

MaskedAGNX = AGNX[NearX1]
MaskedAGNY = AGNY[NearX1]

# This loop runs through all relatively nearby AGN spaxels and find metallicity-marked spaxels which are within 1 X AND Y-coordinate of an AGN-flagged spaxel.
# Adds these spaxels to arrays to save as border spaxels.

for i in range(len(MaskedAGNX)):
    NearCheck = np.where(((ke08X+1 == MaskedAGNX[i]) | (ke08X-1 == MaskedAGNX[i])) & ((ke08Y+1 == MaskedAGNY[i]) | (ke08Y-1 == MaskedAGNY[i])))
BorderSpaxels_X = np.append(BorderSpaxels_X, ke08X[NearCheck])
ke08X = np.delete(ke08X, NearCheck[0])
BorderSpaxels_Y = np.append(BorderSpaxels_Y, ke08Y[NearCheck])
ke08Y = np.delete(ke08Y, NearCheck[0])
BorderObjID = np.append(BorderObjID, ke08ID[NearCheck])
BorderSpaxID = np.append(BorderSpaxID, SpaxID[NearCheck])
SpaxID = np.delete(SpaxID, NearCheck[0])
ke08NearValue = np.append(ke08NearValue, ke08Value[NearCheck])
ke08Value = np.delete(ke08Value, NearCheck[0])

# Appending to arrays to save all border spaxels
BorderX = np.append(BorderX, BorderSpaxels_X)
BorderY = np.append(BorderY, BorderSpaxels_Y)
BorderID = np.append(BorderID, BorderObjID)
Metallicities = np.append(Metallicities, ke08NearValue)

# Calculating the ratio of border spaxels to all metallicity spaxels
if len(RatioX) == 0:
    BorderRatio = 0
else:
    BorderRatio = (len(BorderSpaxels_X)/len(RatioX))
RatioArray = np.append(RatioArray, BorderRatio)
RatioArray = np.trunc(RatioArray*1000)/1000
RatioID = np.append(RatioID, ID)
RatioBorderCount = np.append(RatioBorderCount, len(BorderSpaxels_X))

# Appending to the BIG list of border spaxels and exporting as a text file for use in radius_compare.py
if BorderRatio != 0:
    BIGObjID = np.append(BIGObjID, BorderObjID)
    BIGBorderX = np.append(BIGBorderX, BorderX)
    BIGBorderY = np.append(BIGBorderY, BorderY)
    BIGMetallicities = np.append(BIGMetallicities, Metallicities)
    BIGSpaxID = np.append(BIGSpaxID, BorderSpaxID)
    np.savetxt('{ObjID}_BorderSpaxels_Count={Count}.txt'.format(ObjID=ID, Count=len(BorderSpaxels_X)), np.c_[BorderID, BorderX, BorderY, Metallicities], header='GalaxyID, X, Y, Metallicity', fmt='%s', delimiter=' | ')

# Exporting all results as text files.
np.savetxt('MetallicityRatios.txt', np.c_[RatioID, RatioArray, RatioBorderCount], header='GalaxyID, Border#/Total#, Count_BorderSpaxels', fmt='%s', delimiter=' | ')
np.savetxt('GalaxyIDs.txt', GalaxyIDs, header='GalaxyIDs', fmt='%s')
np.savetxt('BIGList.txt', np.c_[BIGSpaxID, BIGObjID, BIGBorderX,}
B.2 radius_compare.py - KE08

The code to bin all spaxels by radius and calculate the average metallicities of those which lie on the border versus those which do not (per radius-dependent bin). This code adjusts for the general radial-dependence on galactic metallicities.

```python
import numpy as np
import array
import math
import scipy as sci
import pymysql as sql
import datetime
import matplotlib.pyplot as plt

# Query the database to extract the Object ID, Spaxel ID, and Radius from the center of the galaxy (r.kpc). The Spaxel ID will be used to match this array with the spaxels in the "BIG" borderspaxel list.

db = sql.connect(host="xxxxxxxxxx",user="xxxxxxxxxx", password="xxxxxxxxxx", database="xxxxxxxxxx")
c = db.cursor()
c.execute('''select spaxels.objid, metal.spaxID, metal.ke08_smc, spaxels.r_kpc from dr15_metallicities metal, dr15_spaxels_uber spaxels
where metal.spaxID=spaxels.spaxID and metal.ke08_smc is not null and spaxels.agn_flag_smc is not null''')
rowsRad = np.asarray(c.fetchall())
db.close()

# Output is another 2-D array with columns for each of the properties we care about
# Splitting the array up into columns so I can easily compare between these columns and the columns ion the BIG metallicity list

ObjID = rowsRad[:,0]
SpaxID = rowsRad[:,1]
Metallicity = rowsRad[:,2].astype(float)
```
Radius = rowsRad[:,3].astype(float)

SplitSpaxID = np.asarray(np.char.split(SpaxID, sep=' '))
SplitSpaxID = np.stack(SplitSpaxID.ravel())

# Loading in all unique Object IDs (Now called Galaxy IDs). These were
defined in border_spaxels.py.
Consists of all galaxies with at
least one metallicity-marked spaxel
and at least one AGN-flagged spaxel

GalaxyIDs = np.loadtxt('GalaxyIDs.txt', dtype=str, usecols=0)
print(type(GalaxyIDs[0]))
print((GalaxyIDs[0]))
print(np.shape(GalaxyIDs))

# Loading in the Spaxel IDs of all border spaxels and their corresponding
metallicity values

BigSpaxID = np.loadtxt('BIGList.txt', dtype=str, usecols=0).astype(str)

# Loading Spaxel IDs for all border
spaxels

BigMetal = np.loadtxt('BIGList.txt', dtype=str, usecols=8).astype(float)

# Loading all metallicity values for
border spaxels

# Creating arrays for binning and comparing metallicity values

NoBorderAvgList = np.asarray([])
RadBinList = np.asarray([])
BorderAvgList = np.asarray([])
ObjIDList = np.asarray([])

# Beginning of loop to create radius-dependent bins and compare within
them

print('LoopStart:', datetime.datetime.now().time())
for ID in GalaxyIDs[:,]:
    RightGalaxy = np.where(ID == ObjID)  # Masking to select spaxels only in
    the current galaxy in the loop
    SelectRadii = Radius[RightGalaxy]
    SelectMetal = Metallicity[RightGalaxy]
    SelectSpaxID = SpaxID[RightGalaxy]

    if len(SelectRadii) != 0:  # Removing galaxies with no radius measurements

        ## CREATING BINS ##

        # NOTE: Bins are created by rounding the largest radius measurement UP to
        the nearest 0.5 kpc. Bins are defined
        as 0.5 kpc intervals from 0 kpc to
        the maximum value.

        RadiusBins = np.arange(0, np.ceil(np.max(SelectRadii))*0.5, 0.5)

        # Looping through all bins to compare metallicities within
for RadBin in RadiusBins[1:]:
    InBin = np.where((SelectRadii >= RadBin-0.5) & (SelectRadii < RadBin)) # Masking to create bins of radius, metallicity, and spaxID

    RadiiInBin = SelectRadii[InBin]
    MetalInBin = SelectMetal[InBin]
    SpaxIDInBin = SelectSpaxID[InBin]

    SpaxIDMatch = np.where(np.isin(SpaxIDInBin, BigSpaxID)) # Matching the BIG list to spaxels in each bin by spaxID. This way, we determine which spaxels in each bin are found on the border of AGN-flagged regions

    OppSpaxIDMatch = np.where(np.isin(SpaxIDInBin, BigSpaxID, invert=True)) # Finding all Spaxels which DO NOT lie on the border of an AGN-flagged region.

    BorderSpaxIDInBin = SpaxIDInBin[SpaxIDMatch]
    BorderMetal = MetalInBin[SpaxIDMatch]
    NoBorderSpaxIDInBin = SpaxIDInBin[OppSpaxIDMatch]
    NoBorderMetal = MetalInBin[OppSpaxIDMatch]

    BorderAvg = np.average(BorderMetal) # Average the metallicities of spaxels on a border in each bin

    NoBorderAvg = np.average(NoBorderMetal) # Average the metallicities of spaxels NOT on a border in each bin

    RadBinList = np.append(RadBinList, RadBin)
    ObjIDList = np.append(ObjIDList, ID)
    BorderAvgList = np.append(BorderAvgList, BorderAvg)
    NoBorderAvgList = np.append(NoBorderAvgList, NoBorderAvg)

    # Result, a list of averages, an average of metallicities amongst border spaxels and one amongst non-border spaxels, PER BIN

    # Recording Object ID, Radius bin, average border metallicity, and average non-border metallicity for each bin

    np.savetxt('RadBins.txt', np.c_[ObjIDList, RadBinList, BorderAvgList, NoBorderAvgList], header='GalaxyID, Radius Bin, BorderAvgMetal, NoBorderAvgMetal', fmt='%s', delimiter=' | ')

    print('LoopEnd:', datetime.datetime.now().time())
References

Abdurro’uf et al., 2022, ApJS, 259, 35
Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, Astrophysics and Space Science, 361
Kewley L. J., Nicholls D. C., Sutherland R. S., 2019, Annual Review of Astronomy and Astrophysics, 57, 511


